Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

On the Usefulness or Lack Thereof of Optimality Criteria for Structural Change Tests

Pierre Perron
Yohei Yamamoto

November 2012

Hi-Stat Discussion Paper

On the Usefulness or Lack Thereof of Optimality Criteria for Structural Change Tests[*]

Pierre Perron[†]   **Yohei Yamamoto**[‡]

Boston University   **Hitotsubashi University**

May 28, 2008; Revised May 16, 2012.

Abstract

Elliott and Müller (2006) considered the problem of testing for general types of parameter variations, including infrequent breaks. They developed a framework that yields optimal tests, in the sense that they nearly attain some local Gaussian power envelop. The main ingredient in their setup is that the variance of the process generating the changes in the parameters must go to zero at a fast rate. They recommended the so-called $q\hat{L}L$ test, a partial sums type test based on the residuals obtained from the restricted model. We show that for breaks that are very small, its power is indeed higher than other tests, including the popular sup-Wald test. However, the differences are very minor. When the magnitude of change is moderate to large, the power of the test is very low in the context of a regression with lagged dependent variables or when a correction is applied to account for serial correlation in the errors. In many cases, the power goes to zero as the magnitude of change increases. The power of the sup-Wald test does not show this non-monotonicity and its power is far superior to the $q\hat{L}L$ test when the break is not very small. We claim that the optimality of the $q\hat{L}L$ test does not come from the properties of the test statistics but the criterion adopted, which is not useful to analyze structural change tests. Instead, we use fixed-break size asymptotic approximations to assess the relative efficiency or power of the two tests. When doing so, it is shown that the sup-Wald test indeed dominates the $q\hat{L}L$ test and, in many cases, the latter has zero relative asymptotic efficiency.

JEL Classification Number: C22.

Keywords: structural change, local asymptotics, Bahadur efficiency, hypothesis testing, parameter variations.

1 Introduction

The problem of testing for structural changes has been an active area of theoretical and applied research for over 50 years (see Perron, 2006). In the last fifteen years or so, substantial advances have been made to cover models at a level of generality that allows a host of interesting practical applications in the context of unknown change points. These include models with general stationary regressors and errors that can exhibit temporal dependence and heteroskedasticity. In this respect, Andrews (1993) considered the limit distribution of the so-called Sup-type tests, which are based on the maximal value of some statistic over possible break dates within a pre-specified set that excludes some proportion of the data near the beginning or the end of the sample. Bai and Perron (1998) generalized this approach to the case of multiple structural changes in the context of the linear model. Andrews and Ploberger (1994) developed an asymptotic analysis based on a local asymptotic framework, whereby the parameters under the alternative hypothesis are made local to the null value. These tests are optimal in the sense that they maximize a weighted average of the local asymptotic power envelop. More recently, Elliott and Müller (2006) (henceforth EM) considered the problem of testing for general types of parameter variations, including infrequent breaks. They developed a framework that yields optimal tests, in the sense that they (nearly) attain some local Gaussian power envelop. The main ingredient in their setup is that the variance of the process generating the changes in parameters must go to zero at a fast rate. This allows them to show that all small sample efficient tests for parameter instability are asymptotically equivalent. Hence, one does not need to specify the exact nature of the alternative hypothesis to have a test that is efficient and optimal against any such particular alternative hypothesis. The dramatic implication is that one can choose any such small sample efficient test and it will be so-called optimal when some conditions are satisfied, the most important being that the breaks are local to zero as the sample size increases. Accordingly, they recommend the $q\hat{L}L$ test, an extension of the most powerful invariant test in a Gaussian unobserved component model as analyzed by Franzini and Harvey (1983) and Shively (1988). They show the validity of their test under general conditions including dynamic models and models with serially correlated errors provided it is "robustified" to account for this correlation. In the terminology of Perron (2006), the $q\hat{L}L$ test is a partial sums type test, as initially proposed by Gardner (1969). It is based on the residuals obtained from the model restricted to satisfy the null hypothesis of no change.

This result appears quite impressive as one can make a very wide class of tests to be

"optimal". Moreover, the literature that uses the $q\hat{L}L$ indeed labels it as an "optimal" test for changes in parameters. Our claim in this paper is that the framework EM adopt to yield an optimality criterion is rather useless, in the sense that it labels as optimal a test with very bad finite sample properties when confronted with practically relevant types of breaks. The main problem is the adoption of a framework in which the breaks are local to zero. Such a setup used to derive optimality criteria has already been shown to be inadequate in other contexts. Deng and Perron (2008) considered the CUSUM and CUSUM of squares tests (Brown, Durbin and Evans, 1975) and showed that the local asymptotic theoretical framework can give a very unreliable guide to the finite power function, especially when dynamics are involved with the errors and a correction is applied to account for this. Kim and Perron (2009) also criticized the local asymptotic framework used by Andrews and Ploberger (1994). They compared the asymptotic relative efficiency of the Exp, Mean, and Sup functionals of the Wald, LM and LR tests for structural change using the criterion of the relative approximate Bahadur slopes of the tests. They showed that tests based on the Mean functional are inferior to those based on the Sup and Exp when the same base statistic is used. Also, for a given functional, the Wald-based test dominates the LR-based test, which dominates the LM-based test. They also compared tests based on the Wald and LM statistics modified with a HAC estimator. In this case, the inferiority of the LM-based tests is especially pronounced. They assessed the relevance of these theoretical results in finite samples via simulations. The results are in contrast to those of Andrews and Ploberger (1994), based on a local asymptotic framework, and revealed its potential weaknesses in the context of structural change problems.

In this paper, we analyze the power of the $q\hat{L}L$ test. We show that for very small breaks, its power is indeed higher than other tests, including the popular sup-Wald test. However, the differences are very minor. Conversely, when the magnitude of change is moderate to large, we show that the power of the test is very low in the context of a regression with lagged dependent variables or when a correction is applied to account for serial correlation in the errors. In many cases, the power goes to zero as the magnitude of change increases. The power of the sup-Wald test does not show this non-monotonicity and its power is far superior to the $q\hat{L}L$ test when the break is not very small. The same can be said of the UDmax test of Bai and Perron (1998). Overall, the results show that the $q\hat{L}L$ test permits very modest gains in power in a static regression with serially uncorrelated errors and small breaks but has serious deficiencies in dynamic models or when a correction for serial correlation is applied. According to the framework EM adopt to define the optimality criterion used, the

sup-Wald test is not optimal while the $q\hat{L}L$ test is. In this sense, the claims of optimality do not come from the properties of the test statistics but the criterion adopted. Hence, we argue that such a framework to analyze structural change test is inadequate [1]. As in Deng and Perron (2008), we use asymptotic approximations for fixed break sizes to assess the relative efficiency of two tests. We show that the sup-Wald test has higher power than the $q\hat{L}L$ test unless the break size is very small and, in many cases, the latter has zero relative asymptotic efficiency, in the sense that its power goes to zero when the break is large. An earlier version of this paper used the concept of relative Bahadur (1960) efficiency to compare the two tests, as in Kim and Perron (2009). To simplify the exposition we resort to fixed-break asymptotic approximations. The conclusions are identical since both tests have the same tail behavior and the Bahadur efficiency measure reduces to comparing the limit values (properly scaled) under the alternative hypothesis. We show that the fixed-break asymptotic framework delivers better predictions about the finite sample properties compared to a local asymptotic framework, and that according to this criterion the sup-Wald is more efficient (powerful) in most cases.

It is important to note that we make no claim about any optimality property for the sup-Wald. Andrews and Ploberger (1995) showed that the sup-Wald is optimal (in some sense) against distant alternatives. But this result is of little comfort since they also showed the sup-LM to enjoy the same properties. As shown in Kim and Perron (2009), the power properties of the sup-LM are dramatically inferior those of the sup-Wald in the context of a regression with lagged dependent variables or when a correction is applied to account for serial correlation in the errors. This discrepancy again occurs because Andrews and Ploberger (1995) also adopt a local asymptotic framework.

The paper is organized as follows. Section 2 presents the model and tests considered. Section 3 motivates the problems to be analyzed with some empirical examples and we show that the $q\hat{L}L$ test fails to reject while the sup-Wald test gives p-values close to zero. Section 4 presents simulation evidence on the finite sample properties of the tests for four types of parameters changes including those used by EM. Section 5 provides an asymptotic comparison of the $q\hat{L}L$ and sup-Wald tests using the fixed-break asymptotic approximations. Section 6 offers brief concluding remarks and an appendix contains technical derivations.

---

[1]For an early example of the potential pitfalls of local asymptotic analyses in the context of tests with non-monotonic power, see Nelson and Savin (1990).

2 The model and tests considered

The problem of interest is the following. We have a linear regression model given by:

$$y_t = X_t'\beta_t + e_t, \tag{1}$$

for $t = 1, ..., T$. Under the null hypothesis, the parameter vector β of dimension k is constant throughout the sample, i.e., $H_0 : \beta_t = \bar{\beta}$ for all t. Under the alternative hypothesis, there is at least one change in β occurring within the sample, i.e., $H_1 : \beta_t \neq \bar{\beta}$ for some t.

We now describe the test statistics to be analyzed. The $q\hat{L}L$ test statistic of EM is constructed as follows:

Step 1. Compute the OLS residuals $\{\hat{e}_t\}$ from regressing $\{y_t\}$ on $\{X_t\}$.

Step 2. Construct a consistent estimator $\hat{V}_X$ of the $k \times k$ long-run covariance matrix of $\{X_t e_t\}$. If e_t can be assumed serially uncorrelated, a natural choice is a heteroskedasticity robust estimator. For the more general case of possibly autocorrelated errors, many such estimators have been suggested; see, e.g., Newey and West (1987) and Andrews (1991).

Step 3. For each $t = 1, ..., T$, compute $\{\hat{u}_t\} = \{\hat{V}_X^{-1/2} X_t \hat{e}_t\}$ and denote the k elements of $\{\hat{u}_t\}$ by $\{\hat{u}_{t,i}\}$, $i = 1, ..., k$.

Step 4. For each series $\{\hat{u}_{t,i}\}$, compute a new series $\{\hat{\omega}_{t,i}\}$, where $\hat{\omega}_{t,i} = \bar{r}\hat{\omega}_{t-1,i} + \Delta\hat{u}_{t,i}$ and $\hat{\omega}_{1,i} = \hat{u}_{1,i}$, with $\bar{r} = 1 - c/T$ and $c = 10$.

Step 5. Compute the sum of squared residuals from OLS regressions of $\{\hat{\omega}_{t,i}\}$ on $\{\bar{r}^t\}$ for each i and sum these over $i = 1, ..., k$.

Step 6. Multiply this sum of sum of squared residuals by $\bar{r}$, and subtract $\sum_{i=1}^{k}\sum_{t=1}^{T}\hat{u}_{t,i}^2$ to obtain the test statistic $q\hat{L}L$.

The sup-Wald test statistic is the maximal value of the Wald test statistics over the permissible break dates, i.e., $SW = \sup_{\lambda \in \Lambda_\epsilon} W_T(\lambda)$, where $\Lambda_\epsilon = [\epsilon, 1 - \epsilon]$ is the set of the possible break fractions. For the linear regression model with i.i.d. errors, the Wald test statistic is given by

$$W_T(\lambda) = T\left(\frac{SSR_{T\tau} - SSR_T(\lambda)}{SSR_T(\lambda)}\right)$$

where SSR_{Tr} and $SSR_T(\lambda)$ are the restricted and the unrestricted sum of squared residuals, respectively. The latter is obtained from the regression

$$y_t = X_1(\lambda)'_t \beta_1 + X_2(\lambda)'_t \beta_2 + e_t(\lambda), \tag{2}$$

where $X_1(\lambda)'_t = X_t$ if $t < [T\lambda]$ and 0 otherwise, while $X_2(\lambda)'_t = X_t$ if $t \geq [T\lambda]$ and 0 otherwise. When correcting for possible serial correlation in the errors,

$$W_T(\lambda) = \frac{1}{T}\left(\frac{T-2k}{k}\right)\hat{\beta}'H'(H\hat{V}(\hat{\beta})H')^{-1}H\hat{\beta},$$

where $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ is the OLS estimate from (2), H is the conventional matrix such that $(H\hat{\beta})' = (\hat{\beta}'_1 - \hat{\beta}'_2)$, and $\hat{V}(\hat{\beta})$ is an estimate of the variance covariance matrix of $\hat{\beta}$ that is robust to serial correlation and heteroskedasticity, i.e.,

$$\hat{V}(\hat{\beta}) = T(\bar{X}'\bar{X})^{-1}\hat{V}_X(\bar{X}'\bar{X})^{-1} \tag{3}$$

where $\bar{X} = [X_1(\lambda), X_2(\lambda)]$. Whenever a correction for serial correlation is needed, we shall use Andrews' (1991) method, so that

$$\hat{V}_X = \hat{\Gamma}_0 + \sum_{j=1}^{T-1}\kappa(j,m)[\hat{\Gamma}_j + \hat{\Gamma}'_j]$$

where $\hat{\Gamma}_j = T^{-1}\sum_{t=j}^{T}\hat{v}_t\hat{v}_{t-j}$ and $\kappa(j,m)$ is a kernel function with bandwidth m, $\hat{v}_t = X_t\hat{e}_t$ for the $q\hat{L}L$ test and for the sup-Wald test, $\hat{v}_t = X_t\hat{e}(\lambda)_t$, with $\hat{e}(\lambda)_t$ the OLS residuals from the unconstrained regression (2). Below, we use the Bartlett window and choose the bandwidth using the method of Andrews (1991) with an AR(1) approximation. In the scalar case, the method implies that $m = 1.1447(\alpha T)^{1/3}$, where $\alpha = 4\hat{\rho}/(1-\hat{\rho}^2)^2$ and $\hat{\rho}$ is the OLS estimate from a first-order autoregression applied to $\hat{v}_t$ (the results are qualitatively similar using other windows). If only a correction for heteroskedasticity is applied, then $\hat{V}_X = \hat{\Gamma}_0$.

Throughout, we use tests with a 5% nominal size so that with one coefficient allowed to change, the decision rule for the $q\hat{L}L$ is to reject when the statistic is less than 8.36 (see EM) and that for the Wald test, using a trimming $\epsilon = 0.15$, is to reject when the statistic is greater than 8.85 (see Andrews, 1993). The choice of the trimming is dictated by the desire to have tests with roughly the same critical values so that the relative power properties can be properly evaluated. If anything, it favors the $q\hat{L}L$ since its critical value is less in absolute terms.

3 Motivation

It is useful to start with empirical examples to illustrate the main issues involved. We first consider the US (ex-post) real interest rate series over the period 1961:1-1986:3 (the three-month treasury bill rate, deflated by the CPI inflation rate taken from the Citibase data bank). It is the same series analyzed by Garcia and Perron (1996) and Bai and Perron (2003). We reproduce the series in Figure 1. Garcia and Perron (1996) argued that the series was basically a white noise process around a mean that exhibited two changes, one in 1972:3 and one in 1980:3. Using more sophisticated procedures, Bai and Perron (2003) argued for the presence of a third small break in 1966:4. As can be seen from the figure, the change in 1980:3 is very large (7.44%), the change in 1972:3 is moderate (2.67%) and the change in 1966:4 is small (0.95%). This series therefore fits the framework of EM quite well, in the sense that one can view the changes in mean occurring with some small probability and random magnitude.

This is a simple problem of changes in mean so that the relevant regressor is a constant. It is important first to check if the residuals are correlated under the null hypothesis since this dictates the appropriate versions of the tests to use. To do so here, one must simply assess whether the demeaned series shows evidence of autocorrelation. The Lagrange Multiplier test statistic for serial correlation up to order four is highly significant with a value of 54.42 (the first-order serial correlation of the estimated residuals is 0.628). We therefore need to correct any statistic used for the presence of serial correlation in the errors. We apply non-parametric corrections, the so-called robust versions of the tests as described in the previous section. The value of the $q\hat{L}L$ test is -6.51, which is insignificant, even at the 10% level. On the other hand, the value of the sup-Wald test is 74.50, which is highly significant with a p-value [2] less than 0.001 [3].

Another example of similar results can be found in Giacomini and Rossi (2006). They considered testing for structural change in a regression of the growth rate of industrial production on the difference between long-term and short-term interest rates. Their model implies serially correlated errors so that a correction is needed. Even though they fix the bandwidth a priori, the results show overwhelming rejections with Wald-type tests and non-rejections with the $q\hat{L}L$ and LM-type tests.

---

[2]Throughout, p-value refers to "asymptotic p-value".

[3]The value of the sup-Wald test for two changes considered in Garcia and Perron (1996) and Bai and Perron (1998) is 48.26, also highly significant. The UDmax test of Bai and Perron (1998) is also highly significant with a value of 56.53.

Consider now an example involving a regression with a lagged dependent variable so that if the model is well specified, the errors are uncorrelated and there is no need for a correction. Musso, Stracca and van Dick (2009) study instability and nonlinearity in the Euro Area Phillips curve. A specification they consider is the following simple model:

$$\pi_t = \mu + \alpha \pi_{t-1} + \beta x_t + e_t \tag{4}$$

where π_t is the inflation rate (here the GDP Deflator) and x_t is a measure of the output gap [4] (the data are quarterly for the period 1970:1 and 2005:4, see Musso et al., 2009, for more details [5]). Lags of $\Delta \pi_t$ and Δx_t can be added but the results reported below are qualitatively similar. Using a time varying parameter model, Musso et al. (2009) argue that a large change in the mean of inflation μ and a small change in the slope parameter β have occurred, and that once these changes are accounted for, there is no evidence of changes in the persistence parameter α. We applied the $q\hat{L}L$ and sup-Wald tests to this regression to test for a change in μ. The $q\hat{L}L$ test is unable to reject the null hypothesis of no change (the value is -5.255, so that the p-value is well above 10%). On the contrary, and in accordance with the results of Musso et al. (2009), the sup-Wald test overwhelmingly rejects the null hypothesis (the value is 129.31, so that the p-value is well below 0.001). With the specification (4), neither of the tests are able to reject the null hypothesis of no change in β. However, if three lags of $\Delta \pi_t$ are added to the regression, both reject at the 5% level despite the fact that the change is modest (the values are -14.192 for the $q\hat{L}L$ and 10.57 for sup-Wald).

The goal of this paper is to explain why in all three examples, the $q\hat{L}L$ test has no power against large parameter changes while the sup-Wald has high power and why both have similar power against small changes.

4 Some simulation results

We start with simulation experiments to illustrate the problems to be addressed. The simulations are based on a simple regression with one regressor ($k = 1$). The data generating process is

$$y_t = \mu + x_t \beta_t + e_t, \tag{5}$$

---

[4]We used the first principle component of the following six indicators: the three versions of the output gap estimates based on the production function approach using three varieties of unobserved components models and the three types of filtered real GDP estimates using three different filters. See Musso et al. (2007) for more details, who also use other measures. We also tried, as they suggest, a simple average of these six indicators with no change in qualitative results.

[5]We are grateful to Dick van Dijk for supplying us with the data.

with $e_t \sim N(0, \sigma_e)$, and where the regressor x_t is a simple AR(1) process of the form

$$x_t = z_t + \kappa,$$

$$z_t = \rho z_{t-1} + v_t,$$

with $v_t \sim i.i.d.\ N(0, \sigma_v^2)$. We set $\kappa = 5$, $\mu = 0$, $\rho = 0.5$ and $\sigma_e = \sigma_v = 1$ in all cases. The results are robust to variations in the DGP but it is important that the mean of the regressor κ be non-zero. We consider the power of the $q\hat{L}L$ test, the sup-Wald test, the UDmax test of Bai and Perron (1998), as well as Nyblom's (1989) test Ny defined by

$$Ny = \frac{1}{T^2 \hat{\sigma}^2} tr[Q^{-1} \sum_{t=1}^{T} S_t S_t'],$$

where $S_t = \sum_{j=t}^{T} X_t \hat{e}_t$, $Q = T^{-1} \sum_{t=1}^{T} X_t X_t'$, and $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{e}_t^2$. We consider four types of parameter variations for the coefficient β_t, given by the following models:

Model A: $\beta_t = \beta_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, $\beta_0 = 0$,

Model B: $\beta_t = \theta I_{t > [T\lambda_c]}$,

Model C: $\beta_t = \eta I_{t > [T\lambda_c^r]}$, $\eta \sim i.i.d.\ N(0, \sigma^2)$,

Model D: $\beta_t = \sum_{i=2}^{[T/20]} \eta_i I_{[20(i-1) < t \leq 20i]}$, $\eta_i \sim i.i.d.\ N(0, \sigma^2)$,

where $\lambda_c = 0.5$, λ_c^r is uniformly distributed on the unit interval, and $t_i = 1, ..., [T/20]$. Model A specifies a random walk behavior for the coefficient; Model B, a one time break of fixed magnitude θ at a fixed date $[T\lambda_c]$; Model C, a one time break of some random magnitude at some random date; and Model D, a break of random magnitude every 20 observations. Note that Models A, C, and D can be viewed as satisfying the conditions stated in EM for the $q\hat{L}L$ test to be viewed as optimal. Here, the parameter that indexes the magnitude of the break is σ^2, the variance of the shocks affecting β_t. Model B does not satisfy the conditions for the $q\hat{L}L$ to be optimal given that the break is non-random, but we include it since the sup-Wald test is especially designed for this case. Also, Models C and D are those used by EM to document the finite sample performance of their test. Unless otherwise indicated, $T = 100$ and throughout, the results are obtained from 3,000 replications. We computed both rejection frequencies and size-adjusted power. *But given that the size distortions are very small (as will be shown in the figures), the conclusions are the same using either measure of power and we report only raw power.*

8

We compute the power function of four tests: the $q\hat{L}L$, the sup-Wald, Nyblom's test (5% critical value xxx), and the UDmax test of Bai and Perron (1998) with the maximum number of breaks set to 5 (with a 15% trimming so that the 5% critical value is xxx). We consider three versions for each test: a) the "static case" for which the investigator assumes a priori that the errors are serially uncorrelated and homoskedastic and hence, the regression is (5) and the non-robust versions of the tests are used [6]; b) the "robust case" for which no such assumption about the nature of serial correlation in the errors is made and, hence, the robust versions of the tests are used, as described in Section 2; and c) the "dynamic case" where a lagged dependent variable is included, either as part of the model or to account for serial correlation in a more parametric fashion and the non-robust versions of the tests are applied to the regression

$$y_t = \alpha y_{t-1} + x_t \beta_t + e_t. \tag{6}$$

The fact that we set $\alpha = 0$ and, hence, that the lagged dependent variable is an irrelevant regressor is completely inconsequential. With a non-zero value, the power loss of the $q\hat{L}L$ reported below would simply occur earlier and be exacerbated.

The results for Models A and B are presented in Figure 2, while those for Models C and D are presented in Figure 3. We plot the power of the tests as a function of σ for Models A, C, and D, and as a function of θ for Model B. Consider first the static case presented in the top panels. For Models A, C, and D with very small breaks, the test $q\hat{L}L$ indeed has the highest power. For larger breaks, the power of all tests is basically the same. For Model B, the sup-Wald is slightly more powerful, as expected. But again the differences are minor and vanish as the magnitude of the break increases. The results for Models C and D are consistent with those reported in EM. It should be noted that the differences between the $q\hat{L}L$ and the sup-Wald tests are very minor so one could draw the conclusion that, though not a member of the family of optimal tests as considered by EM, the sup-Wald test is nevertheless nearly as efficient.

Things are very different when considering the robust versions of the tests, which correct for potential serial correlation. These results are presented in the second panels. In all cases, the power of the $q\hat{L}L$ test initially increases with the magnitude of the breaks but quickly decreases (sometimes to zero) subsequently. This is also the case for Nyblom's test. But the power of the sup-Wald test does not show this non-monotonicity and its power is far superior to the $q\hat{L}L$ test when the break is not very small. The same can be said of the UDmax test,

---

[6] We also considered assuming a priori no serial correlation but applying a correction for heteroskedasticity. The results are similar to those of the "static case" and, hence, are not reported.

which has the best overall performance. Similar results are obtained when considering the "dynamic case" in which the regression has a lagged dependent variable, given in the third panel.

Overall, the results show that the $q\hat{L}L$ test permits modest gains in power in the "static case" but has serious deficiencies in dynamic models or when correcting for serial correlation. Some may argue that the relevant case is that of the "static regression". But we argue this is not so. Indeed, a careful investigator would rarely use a static regression with no correction for serial correlation because if breaks are present, the estimated residuals will exhibit serial correlation. Take the simple case of a single change, random or fixed. Then the residuals will exhibit a change in mean. As shown in Perron (1990), this change will cause an increase in the sum of the autoregressive coefficients in an autoregression fitted to the series (here the fitted residuals). Hence, a test for serial correlation applied to the estimated residuals will be significant unless the break is very small. A similar logic applies to multiple breaks, including a random walk process. To document this, Figures 4 and 5 present the power of the LM test for serial correlation up to order 4 in the residuals for each of the four models when estimating the static regression. In all cases, the power increases as the magnitude of the breaks increases. We also present the mean of the estimates of the autoregressive coefficient in an AR(1) model applied to the fitted residuals. The results show that it increases rapidly and flattens out at a level below one (see Perron, 1990, for an explanation in the single break case). Thus, an investigator who has no prior knowledge about the presence or absence of serial correlation in the residuals would, upon performing a test for serial correlation, conclude that the robust version is needed when the breaks are not very small. The top panels of Figures 4 and 5 present the power of the hybrid version of the $q\hat{L}L$ test. By "hybrid", we mean that if a test for serial correlation in the residuals does not reject, the non-robust version is used, otherwise the robust version is used. As expected, the power initially increases but quickly drops as the magnitude of change increases. So the good finite sample performance of the $q\hat{L}L$ test as reported in EM crucially depends on prior knowledge that the errors are not serially correlated, knowledge that is unavailable in practice.

These simulation results go a long way in explaining the empirical findings reported in the previous section. For the real interest rate series, there is at least one very large change in mean. This break contaminates the estimated residuals when fitting a restricted model that allows only for an intercept, as required to construct the $q\hat{L}L$ test. Accordingly, it biases the estimate of the long-run variance to correct for serial correlation in a way that reduces power to a very low level. The simulation results also explain the lack of power when

the regressors include a lagged dependent variable and the break is large, as in the example discussed in Section 3. The loss of power in the case of a model with a lagged dependent variable is especially interesting. It shows that the problem of the $q\hat{L}L$ does not occur solely because it does not get the right scaling pertaining to get the correction for serial correlation as in the robust case. The sup-Wald does not get it right either (except for a one time break) and yet does not suffer from such problems in both the robust and dynamic cases.

In the next section, we provide a theoretical analysis of the properties of the $q\hat{L}L$ and sup-Wald tests based on fixed-break asymptotic approximations. These will show that indeed, in all cases, the $q\hat{L}L$ test has zero asymptotic efficiency compared to the sup-Wald test when a correction for serial correlation in the errors is applied or when a dynamic regression is involved. With a static regression and the prior knowledge of no serial correlation, the two tests have similar properties and in some cases, the sup-Wald actually dominates.

5   Asymptotic comparisons using fixed-break asymptotic approximations

We carry the theoretical analysis for the random walk Model A, Model B involving a one-time change of non-random magnitude at some fixed date, as well as for a random level shift model for the parameters, which is suitable for Models C and D. Throughout, we shall maintain (1) as the data-generating process and impose the following high-level assumptions on the regressors and errors as in Kim and Perron (2009), where as a matter of notation, $\|X\|$ denotes the L_2-norm of a random matrix X, i.e., $\|X\| = (\sum_i \sum_j E |X_{ij}|^2)^{1/2}$.

Assumption 1 *Let $v_t = e_t X_t$, $E(v_t v'_{t-j}) = \Gamma_j$, and $\Sigma = \Gamma_0 + \sum_{j=1}^{\infty}(\Gamma_j + \Gamma'_j)$. Then $E(v_t) = 0$, $\sup_{j\in\{0,...,L_T\}} \sup_{r\in[0,1]} \|T^{-1}\sum_{t=1}^{[Tr]} v_t v'_{t-j} - r\Gamma_j\| = o_p(1)$ for $L_T = o(T)$ ($\Gamma_{-j} = \Gamma'_j$), and the partial sums of v_t satisfy the functional central limit theorem $T^{-1/2}\sum_{t=1}^{[Tr]} v_t \Rightarrow \Sigma^{1/2}W_k(r)$, where $W_k(r)$ is a k-dimensional Wiener process defined on $[0,1]$. Note that if the errors v_t are uncorrelated, then $\Sigma = \Gamma_0$.*

Assumption 2 *The regressors are such that*

$$\sup_{j\in\{0,...,L_T\}} \sup_{r\in[0,1]} \|T^{-1}\sum_{t=j+1}^{[Tr]} X_t X'_{t-j} - rQ_j\| = o_p(1), \tag{7}$$

where $L_T = o(T)$, $Q_{-j} = Q'_j$, and Q_j is some non-singular fixed matrix bounded uniformly in j, i.e., $\sup_{|j|\leq L_T} \|Q_j\| \leq c < \infty$.

Assumption 3 $T^{-1}\sum_{t=1}^{T}\|X_tX_t'\|^2 = O_p(1)$, and with $L_T = o(T)$,

$$\sup_{r\in[0,1]} \left\|T^{-1}\sum_{t=j+1}^{[Tr]} X_tX_t'X_tX_t' - rMM'\right\| = o_p(1),$$

$$\sup_{j\in\{0,...,L_T\}}\sup_{r\in[0,1]} \left\|T^{-1}\sum_{t=j+1}^{[Tr]} X_tX_t'X_tX_{t-j}' - rM_jM_j'\right\| = o_p(1),$$

$$\sup_{j\in\{0,...,L_T\}}\sup_{r\in[0,1]} \left\|T^{-1}\sum_{t=j+1}^{[Tr]} X_tX_t'X_{t-j}X_{t-j}' - rM_{jj}M_{jj}'\right\| = o_p(1),$$

where M, M_j, and M_{jj} are some fixed matrices bounded uniformly in j.

5.0.1 The random walk model

In this first case, we suppose that β_t is generated by a random walk process, i.e.,

$$\beta_t = \beta_{t-1} + \varepsilon_t, \tag{8}$$

with ε_t is a k-dimensional i.i.d. vector process with $E(\varepsilon_t) = 0$, $E(\varepsilon_t\varepsilon_t') = \Lambda$ and $E(\varepsilon_t e_{t-j}) = 0$ for all j. Also, ε_t is uniform mixing with mixing coefficient ϕ of size $-r/(2r-2)$ or strong mixing with mixing coefficient α of size $-r/(r-2)$, $r > 2$. We also assume that ε_t and X_t are uncorrelated so that the following functional central limit theorem holds:

$$T^{-3/2}\sum_{t=1}^{[Tr]} X_tX_t'\beta_t \Rightarrow Q\Lambda \int_0^r W(s)ds,$$

where $W(s)$ is the k-dimensional standard Wiener process and " $\Rightarrow$ " denotes weak convergence under the Skorohod topology. We also define the Ornstein-Uhlenbeck process $W_c(s) \equiv W(s) - c\int_0^s e^{-c(s-\lambda)}W(\lambda)d\lambda$. We start with the case of the static regression for which the investigator has prior knowledge that the errors are serially uncorrelated. Note that here and throughout the paper, the symbol $O_p(1)$ stands for a variable that is stochastically bounded but is not $o_p(1)$.

Theorem 1 *Assume that the $q\hat{L}L$ and sup-Wald (SW) tests are constructed from the static regression (5) with no correction for serial correlation and that the data are generated by (1) with β_t specified by the random walk (8), then:*

$$T^{-1}q\hat{L}L \Rightarrow tr[\bar{V}_X^{-1/2}Q\Lambda\Xi\Lambda'Q'\bar{V}_X'^{-1/2}] = O_p(1)$$

uniformly in $\|\Lambda\|$, *where*

$$\bar{V}_X \equiv M\Lambda[\int_0^1 WW' - \int_0^1 W \int_0^1 W']\Lambda'M'$$

and

$$\Xi \equiv \int_0^1 W_c W_c' - \int_0^1 WW' + \int_0^1 W \int_0^1 W' - \frac{2c}{1-e^{-2c}} \int_0^1 e^{-cr}W \int_0^1 e^{-cr}W'.$$

Also,

$$T^{-1}SW \Rightarrow \frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})} = O_p(1),$$

uniformly in $\|\Lambda\|$, *where*

$$
\begin{aligned}
SSR^{r*} &\equiv \int_0^1 W'\Lambda'Q\Lambda W - \int_0^1 W'\Lambda'Q\Lambda \int_0^1 W, \\
SSR^*(\lambda) &\equiv \int_0^1 W'\Lambda'Q\Lambda W - \frac{1}{\lambda}\int_0^\lambda W'\Lambda'Q\Lambda \int_0^\lambda W - \frac{1}{1-\lambda}\int_\lambda^1 W'\Lambda'Q\Lambda \int_\lambda^1 W,
\end{aligned}
$$

and $\bar{\lambda}$ *is the value of* λ *that minimizes* $SSR^*(\lambda)$ *over the set* $[\epsilon, 1-\epsilon]$.

Consider the case with a single regressor that is a constant. The tests then amount to distinguishing between an *i.i.d.* process and a $IMA(1,1)$ process. The expressions in Theorem 1 reduce to

$$T^{-1}q\hat{L}L \Rightarrow \left[\int_0^1 W^2 dr - (\int_0^1 W dr)^2\right]^{-1} \Xi \tag{9}$$

and

$$T^{-1}SW \Rightarrow \frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})}, \tag{10}$$

where

$$
\begin{aligned}
SSR^*(\lambda) &= \int_0^1 W^2 dr - \frac{1}{\lambda}(\int_0^1 W dr)^2 + \frac{2}{\lambda}\int_0^1 W dr \int_\lambda^1 W dr - \frac{1}{\lambda(1-\lambda)}(\int_\lambda^1 W dr)^2, \\
SSR^{r*} &= \int_0^1 W^2 dr - (\int_0^1 W dr)^2,
\end{aligned}
$$

and again $\bar{\lambda}$ is the value of λ that minimizes $SSR^*(\lambda)$ over the set $[\epsilon, 1-\epsilon]$.

Note that, appropriately scaled, the limits of the $q\hat{L}L$ and sup-Wald tests are random in this case. Also, in the case of a single regressor, they do not depend on any nuisance parameters. This is because in the limit, the random walk component dominates. Since the power depends on the relative variances of the shocks e_t and those affecting β_t, we should not expect a good finite sample approximation for small breaks. Nevertheless, it is instructive to compare the mean of the limit random variables in both cases. To that effect, we simulated the random variables in (9) and (10) using 1,000 steps to approximate the Wiener process

and 1,000 replications. The negative of the mean of the random variable in (9) is 1.13. The mean of the random variable in (10) depends on the trimming used to search for the break dates that minimizes the sum of squared residuals. For $\epsilon = 0.15$, the value is 2.01. We performed a similar analysis of the DGP used in the simulations of Section 4. In this case, the mean of the limit distribution of the sup-Wald test is essentially the same, while the mean of the limit distribution of the $q\hat{L}L$ test is somewhat smaller at 0.76. These values are large enough to imply that both tests reject when $T = 100$ at a size of 5%. The cdf of the limit distribution of the sup-Wald test almost uniformly dominates that of the $q\hat{L}L$ test even when $\Lambda = 1$, a small value (the limiting cdf of (minus) the $q\hat{L}L$ test is to the right only for the extreme left tail). But the numerical results predict that both tests will have good (and similar) finite sample power as documented in the simulations [xxx Get the probabilities that each test rejects when $T = 100$ for various values of Λ].

We now turn to the case where one accounts for possible serial correlation in the errors using a non-parametric correction. In the case of the $q\hat{L}L$ test, this amounts to replacing the variance estimate $\hat{V}_X$ by $\hat{h}(0)$, i.e., an estimate of the spectral density function at frequency zero. Typically, it is obtained using a weighted sum of the autocovariances of $\{\hat{v}_t = X_t\hat{e}_t\}$ so that

$$h\hat{(0)} = \hat{\Gamma}_0 + \sum_{j=1}^{T-1} \kappa(j,m)[\hat{\Gamma}_j + \hat{\Gamma}'_j] \tag{11}$$

where $\hat{\Gamma}_j = T^{-1}\sum_{t=j}^{T}\hat{v}_t\hat{v}'_{t-j}$ and $\kappa(\cdot)$ is a kernel function with bandwidth m. Suppose we use the Bartlett kernel for $\kappa(\cdot)$ and Andrews' (1991) data dependent method for selecting the bandwidth based on an AR(1) approximation then the following results hold.

Theorem 2 *Assume that the $q\hat{L}L$ and sup-Wald (SW) tests are constructed from the static regression (5) and that the process is generated by (1) with β_t specified by the random walk (8). Suppose the serial correlation in the errors is accounted for using an estimate of the form (11) with the Bartlett kernel and the bandwidth chosen by Andrew's (1991) data dependent method based on an AR(1) approximation. a) If $\{M - M_{11}\}_{ii} = 0$ for all i, where $\{Z\}_{ii}$ is the ith diagonal element of the matrix Z, then*

$$q\hat{L}L \Rightarrow tr[\bar{h}(0)^{-1/2}Q\Lambda\Xi\Lambda'Q'\bar{h}(0)'^{-1/2}] = O_p(1)$$

uniformly in $\|\Lambda\|$, where $\bar{h}(0) \equiv \lim_{T\to\infty} T^{-1}\sum_{j=-T}^{T}\kappa(j,m_T)\bar{\Gamma}_j$, with $m_T \equiv 1.1447(\bar{\alpha}T)^{1/3}$, $\bar{\alpha} \equiv \phi^{-2}$, $\bar{\Gamma}_j \equiv M_{jj}\Lambda[\int_0^1 WW' - \int_0^1 W \int_0^1 W']\Lambda'M'_{jj}$, Ξ as defined in Theorem 1, and

$$\phi \equiv \frac{-W(1)^2 - (\int_0^1 W)^2 + W(1)\int_0^1 W}{\int_0^1 W^2 - (\int_0^1 W)^2}.$$

14

Also,

$$SW \Rightarrow (\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} W - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda^*}^1 W)'Q'[\bar{h}_1(0) + \bar{h}_2(0)]^{-1}Q$$
$$\times (\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} W - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda*}^1 W) = O_p(1)$$

uniformly in $\|\Lambda\|$, *where* λ^* *is the value of* λ *that maximizes the limit of the Wald test over the set* $[\epsilon, 1-\epsilon]$, $\bar{h}_i(0) \equiv \lim_{T\to\infty} T^{-1}\sum_{j=-T}^{T}\kappa(j, m_{i,T})\bar{\Gamma}_{i,j}$ $(i = 1, 2)$, *with* $m_{i,T} \equiv 1.1447(\bar{\alpha}_i T)^{1/3}$, $\bar{\alpha}_i \equiv \phi_i^{-2}$,

$$\bar{\Gamma}_{1,j} \equiv M_{jj}\Lambda[\int_0^{\lambda^*} WW' - \int_0^{\lambda^*} W \int_0^{\lambda^*} W']\Lambda'M_{jj}',$$
$$\bar{\Gamma}_{2,j} \equiv M_{jj}\Lambda[\int_{\lambda^*}^1 WW' - \int_{\lambda^*}^1 W \int_{\lambda^*}^1 W']\Lambda'M_{jj}',$$
$$\phi_1 \equiv \frac{-W(\lambda^*)^2 - (\int_0^{\lambda^*} W)^2 + W(\lambda^*)\int_0^{\lambda^*} W}{(\int_0^{\lambda*} W^2) - (\int_0^{\lambda^*} W)^2},$$

and

$$\phi_2 \equiv \frac{-(W(1) - W(\lambda^*))^2 - (\int_{\lambda^*}^1 W)^2 + (W(1) - W(\lambda^*))\int_{\lambda^*}^1 W}{\int_{\lambda^*}^1 W^2 - (\int_{\lambda^*}^1 W)^2}.$$

b) If $\{M - M_{11}\}_{ii} \neq 0$ *for some i, then*

$$T^{-2/3}q\hat{L}L \Rightarrow tr[\bar{h}(0)^{-1/2}Q\Lambda\Xi\Lambda'Q'\bar{h}(0)'^{-1/2}] = O_p(1)$$

uniformly in $\|\Lambda\|$, *where* $\bar{h}(0) \equiv \lim_{T\to\infty} T^{-1/3}\sum_{j=-T}^{T}\kappa(j, m_T)\bar{\Gamma}_j$, *with* $\bar{\Gamma}_j$, Ξ *and* m_T *as define in part (a)*, $\bar{\alpha} \equiv 4\bar{\rho}^2/[(1-\bar{\rho}^2)^2]$, *and* $\bar{\rho} \equiv M_{11}^2/M^2$. *Also,*

$$T^{-2/3}SW \Rightarrow (\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} W - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda^*}^1 W)'Q' [\bar{h}_1(0) + \bar{h}_2(0)]^{-1}Q$$
$$\times (\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} W - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda*}^1 W) = O_p(1)$$

uniformly in $\|\Lambda\|$, *where* λ^* *is the value of* λ *that maximizes the limit of the Wald test over the set* $[\epsilon, 1-\epsilon]$, $\bar{h}_i(0) \equiv \lim_{T\to\infty} T^{-1/3}\sum_{j=-T}^{T}\kappa(j, m_{i,T})\bar{\Gamma}_{i,j}$ $(i = 1, 2)$, *with* $\bar{\Gamma}_{i,j}$, $m_{i,T}$, $\bar{\alpha}$ *and* $\bar{\rho}$ *as defined directly above for the limit of the* $q\hat{L}L$ *test.*

Though the results are quite complex, we can simulate these limit distributions. Note that we approximate $\bar{h}(0)$ with a scaled average where $T = 100$. Similarly m_T is evaluated at $T = 100$. We ran simulations for the DGP considered in the simulations of Section 4, as specified by (5). This is a case where part (b) of Theorem 2 applies since $M \neq M_{11}$. The means, as a function of Λ, are presented in Figure 6. The figure shows that the limit is

invariant to Λ, as in the static case with *i.i.d.* errors. But, we can use these limit results to see what the theory predicts about the likelihood of a rejection. For a given sample size T, a rejection will occur if the realization of the limit random variable is above the critical values divided by $T^{2/3}$ (using $-q\hat{L}L$). As in the simulations, we considered 5% tests. The relevant critical values are plotted as the dotted line in each panel. The results predict that on average the $q\hat{L}L$ test will not reject, while the sup-Wald will. To get a better description in terms of the probability of rejecting, the second panel in Figure 6 presents the cdfs of the limit random variables. The results clearly show that the $q\hat{L}L$ test has a very small probability of rejecting when $\Lambda = 1$ and nearly none if $\Lambda = 15$, while the sup-Wald test has large rejection probabilities in both cases. These theoretical predictions are in accordance with the simulations. Though, these results are DGP specific, we found that the same features hold for other specifications.

Suppose now that we use a dynamic regression of the form (6) with a lagged dependent variable as a regressor. The results are stated in the next theorem.

Theorem 3 *Assume that the $q\hat{L}L$ and sup-Wald (SW) tests are constructed from a dynamic regression of the form (6), with a lagged dependent variable as a regressor, and the process is generated by (1), with β_t specified by the random walk (8). Then,*

$$T^{-1}q\hat{L}L \Rightarrow tr[\bar{V}_X^{-1/2}[Q - \bar{\alpha}Q_1]\Lambda\Xi\Lambda'[Q - \bar{\alpha}Q_1]'\bar{V}_X'^{-1/2}] = O_p(1)$$

uniformly in $||\Lambda||$, where Ξ is as defined in Theorem 2, and

$$\bar{\alpha} \equiv \{\textstyle\int_0^1 W'\Lambda'Q\Lambda W - \int_0^1 W'\Lambda'Q_1'Q^{-1}Q_1\Lambda \int_0^1 W\}^{-1}\{\int_0^1 W'\Lambda'Q_1\Lambda W - \int_0^1 W'\Lambda'Q_1\Lambda \int_0^1 W\},$$

and

$$\begin{aligned}
\bar{V}_X &= M\Lambda[\textstyle\int_0^1 WW' - \int_0^1 W \int_0^1 W']\Lambda'M' + \bar{\alpha}^2 MQ^{-1}Q_1\Lambda \int_0^1 W \int_0^1 W'\Lambda'Q_1'Q'^{-1}M' \\
&\quad + \bar{\alpha}^2 M_{11}\Lambda \int_0^1 WW'\Lambda'M_{11}' - 2\bar{\alpha}M_1\Lambda[\textstyle\int_0^1 WW' - \int_0^1 W \int_0^1 W']\Lambda'M_1' \\
&\quad - \bar{\alpha}^2 M_1 Q^{-1}Q_1\Lambda \int_0^1 W \int_0^1 W'\Lambda'M_1' - \bar{\alpha}^2 M_1\Lambda \int_0^1 W \int_0^1 W'\Lambda'Q_1'Q'^{-1}M_1'.
\end{aligned}$$

Also,

$$T^{-1}SW \Rightarrow \frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})} = O_p(1)$$

uniformly in $||\Lambda||$, where $\bar{\lambda}$ is the value of λ that minimizes $SSR^*(\lambda)$ over the set $[\epsilon, 1 - \epsilon]$,

$$\begin{aligned}
SSR^{r*} &\equiv (1 + \bar{\alpha}^2)\textstyle\int_0^1 W'\Lambda'Q\Lambda W - 2\bar{\alpha}\int_0^1 W'\Lambda'Q_1\Lambda W \\
&\quad - \textstyle\int_0^1 W'\Lambda'[Q - \bar{\alpha}Q_1]'Q^{-1}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 W,
\end{aligned}$$

16

and

$$
\begin{aligned}
SSR^*(\lambda) \;\equiv\; & (1+\bar{\alpha}_u^2)\int_0^1 W'\Lambda'Q\Lambda W - 2\bar{\alpha}_u\int_0^1 W'\Lambda'Q_1\Lambda W \\
& -\frac{1}{\lambda}\int_0^\lambda W'\Lambda'[Q-\bar{\alpha}_u Q_1]'Q^{-1}[Q-\bar{\alpha}_u Q_1]\Lambda\int_0^\lambda W \\
& -\frac{1}{1-\lambda}\int_\lambda^1 W'\Lambda'[Q-\bar{\alpha}_u Q_1]'Q^{-1}[Q-\bar{\alpha}_u Q_1]\Lambda\int_\lambda^1 W,
\end{aligned}
$$

with

$$
\begin{aligned}
\bar{\alpha}_u \;\equiv\; & \{\int_0^1 W'\Lambda'Q\Lambda W - \frac{1}{\lambda}\int_0^\lambda W'\Lambda Q_1'Q^{-1}Q_1\Lambda\int_0^\lambda W - \frac{1}{1-\lambda}\int_\lambda^1 W'\Lambda Q_1'Q^{-1}Q_1\Lambda\int_\lambda^1 W\}^{-1} \\
& \times\{\int_0^1 W'\Lambda'Q_1\Lambda W - \frac{1}{\lambda}\int_0^\lambda W'\Lambda'Q_1\Lambda\int_0^\lambda W - \frac{1}{1-\lambda}\int_\lambda^1 W'\Lambda'Q_1\Lambda\int_\lambda^1 W\}.
\end{aligned}
$$

and $\bar{\alpha}$ as defined above.

The above expressions are quite complex. Consider the mean shift model with $X_t = \{1\}$ for all t. Then,

$$
T^{-1}q\hat{L}L \Rightarrow \frac{c^2\Lambda^2\int_0^1 W_c^2 - c\Lambda^2 - 2c\sigma_e^2 - (2c\Lambda^2/(1-e^{-2c}))[W_c(1)-c\int_0^1 e^{-cr}W_c]^2}{\Lambda^2+2\sigma_e^2} = O_p(1)
$$

and

$$
\begin{aligned}
T^{-1}SW \;\Rightarrow\; & \{-2(\bar{\alpha}-\bar{\alpha}_u)[\frac{1}{2}\Lambda^2(W(1)-1)-\sigma_e^2] + \Lambda^2 W(1)^2 - \Lambda^2 W(\lambda)^2 - \Lambda^2 W(1-\lambda)^2 \\
& -\bar{\alpha}^2\Lambda^2\int_0^1 W^2 + \bar{\alpha}_u^2\frac{1}{\lambda}\Lambda^2(\int_0^\lambda W)^2 + \bar{\alpha}_u^2\frac{1}{1-\lambda}\Lambda^2(\int_\lambda^1 W)^2\}/(\Lambda^2+2\sigma_e^2) = O_p(1),
\end{aligned}
$$

where

$$
\bar{\alpha} \equiv [(1/2)(W(1)^2-1)-\sigma_e^2/\Lambda^2-W(1)\int_0^1 W]/[\int_0^1 W^2-(\int_0^1 W)^2]
$$

and

$$
\bar{\alpha}_u \equiv \frac{(1/2)(W(1)^2-1)-\sigma_e^2/\Lambda^2-W(\lambda)(1/\lambda)\int_0^\lambda W - W(1-\lambda)\int_\lambda^1 W/(1-\lambda)}{\int_0^1 W^2-(1/\lambda)(\int_0^\lambda W)^2-(\int_\lambda^1 W)^2/(1-\lambda)}
$$

We again simulated these limit distributions for the DGP specified by (5). The means, as a function of Λ, are presented in the first panel of Figure 7 along with the critical values pertaining to a 5% test and $T = 100$. The results predict that on average the $q\hat{L}L$ test will not reject while the sup-Wald will. The second and third panels present the cdfs of the limit random variables for $\Lambda = 1$ and $\Lambda = 15$. The results clearly show that the $q\hat{L}L$ test has again a near-zero probability of rejecting and that this probability is not affected by the value of Λ, while the sup-Wald has a large probability of rejecting for both values. These theoretical predictions are in accordance with the simulations discussed above.

17

5.1 Infrequent Breaks Model

We now consider a model that specifies relatively rare occurrences of shifts in the parameters that is suitable for Models C and D. The specification is as follows:

$$\beta_t = \sum_{j=1}^{t} \delta_{T,j}, \; \delta_{T,t} = \pi_{T,t}\eta_t \tag{12}$$

Here, $\eta_t \sim i.i.d. \, (0, \Lambda)$ and $\pi_{T,t}$ is a binomial variable that takes value 1 with probability p/T and value 0 with probability $(1 - p/T)$, i.e. $\pi_{T,t} \sim i.i.d. \, B(p/T, 1)$. We also assume that the components $\pi_{T,t}$ and η_t are mutually independent. Note that the parameter p is independent of the sample size T. Hence, as T increases, the level shifts become relatively rare. This is an important ingredient that will allow us to derive interesting results. Intuitively, we need this specification to model structural changes, i.e., relatively infrequent events that affect the time series properties of the parameters in a permanent fashion. If p/T converges to some value in $(0, 1)$, the model is best construed as depicting a standard unit root process. This model is the most interesting to analyze since it bridges the gap between the two extremes of random walk parameter variations and a single change at some known date. Also, it does not suffer from the discontinuity problem present in the random walk model. Finally, it is more in line with the type of processes for which the $q\hat{L}L$ is especially designed.

A crucial ingredient used here is a Functional Central Limit Theorem for the cumulative level shifts process $u_{T,t}$. This has been considered by Georgiev (2002) and Leipus and Viano (2003). The results relevant to our analysis are stated in the following lemma.

Lemma 1 *(Georgiev, 2002; Leipus and Viano, 2003) Consider β_t specified by (12) with $0 < p < \infty$ and define $u_T(s) = \sum_{t=1}^{[Ts]} \delta_{T,t}$, then $u_T(s) \Rightarrow J(s)$, where $J(s)$ is a compound Poisson process defined by $J(s) = \sum_{j=0}^{N(s)} \eta_j$ with $N(s)$ being a Poisson process with jump intensity p, independent of η_j for all j.*

We have the following theorem for the three scenarios considered.

Theorem 4 *a) Assume that the $q\hat{L}L$ and sup-Wald (SW) test statistics are constructed from the static regression (5) with no correction for serial correlation and that the process is generated by (1) with β_t specified by the random level shift model (12), then*

$$T^{-1}q\hat{L}L \Rightarrow tr[\bar{V}_X^{-1/2} Q \Lambda \Psi \Lambda' Q' \bar{V}_X'^{-1/2}] = O_p(1)$$

uniformly in $\|\Lambda\|$, *where* Ψ *is a function of* $J(r)$ *defined by (A.14) in the appendix and*

$$\bar{V}_X \equiv \sigma_e^2 Q + M\Lambda[\int_0^1 JJ' - \int_0^1 J \int_0^1 J']\Lambda' M'.$$

Also,

$$T^{-1}SW \Rightarrow \frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})} = O_p(1),$$

where

$$SSR^{r*} \equiv \sigma_e^2 + \int_0^1 J'\Lambda'Q\Lambda J - \int_0^1 J'\Lambda'Q\Lambda \int_0^1 J,$$

$$SSR^*(\lambda) \equiv \sigma_e^2 + \int_0^1 J'\Lambda'Q\Lambda J - \frac{1}{\lambda}\int_0^\lambda J'\Lambda'Q\Lambda \int_0^\lambda J - \frac{1}{1-\lambda}\int_\lambda^1 J'\Lambda'Q\Lambda \int_\lambda^1 J,$$

and $\bar{\lambda}$ *is the value of* λ *that minimizes* $SSR^*(\lambda)$ *over the set* $[\epsilon, 1-\epsilon]$.

We again considered simulating the limit random variables to assess the relative efficiency of the two tests for the DGP used in Section 4. The results are presented in Figure 8, which reports the means and cdfs of the limit random variables pertaining to the $q\hat{L}L$ and sup-Wald tests for $p/T = 0.05$, 0.10, and 0.50 (along with the critical values corresponding to a 5% size test with $T = 100$). Looking at the cdfs of the limit distributions, the sup-Wald and $q\hat{L}L$ are nearly as powerful for small alternatives ($\Lambda = 1$). On the other hand, the sup-Wald dominates the $q\hat{L}L$ for large alternatives [xxGet some numbers on the power of the two test for $T = 100$, $\Lambda = 1, 15$ and $p/T = 0.05$, 0.10, and 0.50]. What transpires from these results is the fact that both tests will perform well in the static case, with good power that is monotonic as documented in the simulations.

Consider now the case where a correction for serial correlation is applied. We then have the following results.

Theorem 5 *Assume that the* $q\hat{L}L$ *and sup-Wald (SW) tests are constructed from the static regression (5) and that the process is generated by (1) with* β_t *specified by the random level shift model (12). If the serial correlation in the errors is accounted for using an estimate of the form (11) with the Bartlett kernel and the bandwidth chosen by Andrew's (1991) data dependent method based on an AR(1) approximation. Then, a) If* $\{M - M_{11}\}_{ii} = 0$ *for all* i, *where* $\{Z\}_{ii}$ *is the ith diagonal element of the matrix* Z, *then*

$$T^{-2/3}q\hat{L}L \Rightarrow tr[\bar{h}(0)^{-1/2}Q\Lambda\Xi\Lambda'Q'\bar{h}(0)'^{-1/2}] = O_p(\|\Lambda\|^{-4/3})$$

19

uniformly in $\|\Lambda\|$, *where* $\bar{h}(0) \equiv \lim_{T\to\infty} T^{-1/3} \sum_{j=-T}^{T} \kappa(j, m_T)\bar{\Gamma}_j$ *with* $m_T \equiv 1.1447(\bar{\alpha}T)^{1/3}$,
$\bar{\alpha} \equiv 4\bar{\rho}^2/(1 - \bar{\rho}^2)^2$, $\bar{\Gamma}_j \equiv M_{jj}\Lambda[\int_0^1 JJ' - \int_0^1 J \int_0^1 J']\Lambda' M'_{jj}$, *and*

$$\bar{\rho} \equiv \frac{M^2\Lambda^2[\int_0^1 J^2 - (\int_0^1 J)^2]}{\sigma_e^2 Q + M^2\Lambda^2[\int_0^1 J^2 - (\int_0^1 J)^2]}.$$

Also,

$$T^{-2/3}SW \quad \Rightarrow \quad (\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} J - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda^*}^1 J)'Q'[\bar{h}_1(0) + \bar{h}_2(0)]^{-1}Q$$
$$\times(\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} J - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda*}^1 J) = O_p(\|\Lambda\|^{-4/3})$$

uniformly in $\|\Lambda\|$, *where* λ^* *is the value of* λ *that maximizes the limit of the Wald test over the set* $[\epsilon, 1-\epsilon]$, $\bar{h}_i(0) \equiv \lim_{T\to\infty} T^{-1/3} \sum_{j=-T}^{T} \kappa(j, m_{i,T})\bar{\Gamma}_{i,j}$ *(i = 1,2) with* $m_{i,T} \equiv$
$1.1447(\bar{\alpha}_i T)^{1/3}$, $\bar{\alpha}_i \equiv 4\bar{\rho}_i^2/(1 - \bar{\rho}_i^2)^2$,

$$\begin{aligned}
\bar{\Gamma}_{1,j} &\equiv M_{jj}\Lambda[\int_0^{\lambda^*} JJ' - \int_0^{\lambda^*} J \int_0^{\lambda^*} J']\Lambda' M'_{jj}, \\
\bar{\Gamma}_{2,j} &\equiv M_{jj}\Lambda[\int_{\lambda^*}^1 JJ' - \int_{\lambda^*}^1 J \int_{\lambda^*}^1 J']\Lambda' M'_{jj}, \\
\bar{\rho}_1 &\equiv \frac{M^2\Lambda^2 \left[[\int_0^{\lambda^*} J^2 - (\int_0^{\lambda^*} J)^2]\right]}{\sigma_e^2 Q + M^2\Lambda^2 \left[[\int_0^{\lambda^*} J^2 - (\int_0^{\lambda^*} J)^2]\right]},
\end{aligned}$$

and

$$\bar{\rho}_2 \equiv \frac{M^2\Lambda^2[\int_{\lambda^*}^1 J^2 - (\int_{\lambda^*}^1 J)^2]}{\sigma_e^2 Q + M^2\Lambda^2[\int_{\lambda^*}^1 J^2 - (\int_{\lambda^*}^1 J)^2]}.$$

b) If $\{M - M_{11}\}_{ii} \neq 0$ *for some i, then*

$$T^{-2/3}q\hat{L}L \Rightarrow tr[\bar{h}(0)^{-1/2}Q\Lambda\Xi\Lambda'Q'\bar{h}(0)'^{-1/2}] = O_p(1)$$

uniformly in $\|\Lambda\|$, *where* $\bar{h}(0)$ *and* Ξ *are as defined in part (a) with*

$$\bar{\rho} \equiv \frac{M_{11}^2\Lambda^2[\int_0^1 J^2 - (\int_0^1 J)^2]}{\sigma_e^2 Q + M^2\Lambda^2[\int_0^1 J^2 - (\int_0^1 J)^2]}.$$

Also,

$$T^{-2/3}SW \quad \Rightarrow \quad (\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} J - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda^*}^1 J)'Q'[\bar{h}_1(0) + \bar{h}_2(0)]^{-1}Q$$
$$\times(\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} J - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda*}^1 J) = O_p(1)$$

uniformly in $\|\Lambda\|$, *where* λ^* *is the value of* λ *that maximizes the limit of the Wald test over the set* $[\epsilon, 1 - \epsilon]$ *and* $\bar{h}_i(0)$ *is as defined in part (a) with*

$$\bar{\rho}_1 \equiv \frac{M_{11}^2 \Lambda^2 [\int_0^{\lambda^*} J^2 - (\int_0^{\lambda^*} J)^2]}{\sigma_e^2 Q + M^2 \Lambda^2 [\int_0^{\lambda^*} J^2 - (\int_0^{\lambda^*} J)^2]},$$

and

$$\bar{\rho}_2 \equiv \frac{M_{11}^2 \Lambda^2 \left[[\int_{\lambda^*}^1 J^2 - (\int_{\lambda^*}^1 J)^2] \right]}{\sigma_e^2 Q + M^2 \Lambda^2 [\int_{\lambda^*}^1 J^2 - (\int_{\lambda^*}^1 J)^2]}.$$

The results show that when $\{M - M_{11}\}_{ii} = 0$ for all i, as in the mean shift case, both tests will eventually exhibit zero power as the magnitude of the variance of the shocks increases since the limit values tend to zero (though for a fixed $\|\Lambda\|$ both tests are consistent). Things are more interesting when $\{M - M_{11}\}_{ii} \neq 0$ for some i, as for the DGP used in the simulations presented in Section 4. Here, both tests are again consistent but their scaled limit values are increasing in $\|\Lambda\|$, but they are $O_p(1)$ uniformly in $\|\Lambda\|$. The question of interest for assessing relative efficiency is how the limit distributions compare as a function of $\|\Lambda\|$. This is too complex to do for the general case so we again present results based on simulating the limit distributions for the DGP considered in Section 4 (the method described in the previous section is used to approximate $\bar{h}(0)$). The results are presented in Figure 9, which report the means and cdfs of the limit random variables pertaining to the $q\hat{L}L$ and sup-Wald tests for $p/T = 0.05$, 0.10, and 0.50. Consider first the results for the $q\hat{L}L$ test. The mean of the limit distribution initially increases (in absolute value) as Λ increases but quickly drops down to a value below the critical value associated with a 5% test using a sample with 100 observations. This decrease is quicker and more pronounced as the intensity of the level shifts p/T increases. Hence, the theory predicts that on average the $q\hat{L}L$ test will reject for small values of Λ but will not for larger values especially as the level shift intensity increases. Things are similar for the sup-Wald test but there is one important difference. As Λ increases, the mean of the limit random variable stabilizes to a value implying a rejection for all values of Λ that are not very small. We plotted the cdf for a small ($\Lambda = 1$) and large ($\Lambda = 15$) value of the variance of the shocks. The figures make clear that the probability of rejecting the null hypothesis with the $q\hat{L}L$ test is small and shrinks as Λ increases or as the intensity of the level shifts increases. On the other hand, the probability of rejecting with the sup-Wald test is nearly one in all cases.

The next theorem pertains to the case with a lagged dependent variable as a regressor.

21

Theorem 6 *Assume that the $q\hat{L}L$ and sup-Wald (SW) tests are constructed from a dynamic regression of the form (6) with a lagged dependent variable as a regressor and the process is generated by (1) with β_t specified by the random level shift model (12), then*

$$T^{-1}q\hat{L}L \Rightarrow tr[\bar{V}_X^{-1/2}[Q - \bar{\alpha}Q_1]\Lambda\Psi\Lambda'[Q - \bar{\alpha}Q_1]'\bar{V}_X'^{-1/2}] = O_p(1)$$

uniformly in $\|\Lambda\|$, where Ψ is defined by (A.14) in the appendix,

$$\bar{\alpha} \equiv \{\sigma_e^2 + [\int_0^1 J'\Lambda'Q\Lambda J] - \int_0^1 J'\Lambda'Q_1'Q^{-1}Q_1\Lambda \int_0^1 J\}^{-1}\{\int_0^1 J'\Lambda'Q_1\Lambda J - \int_0^1 J'\Lambda'Q_1\Lambda \int_0^1 J\},$$

and

$$\begin{aligned}
\bar{V}_X &= (1 + \bar{\alpha}^2)\sigma_e^2 + M\Lambda[\int_0^1 JJ' - \int_0^1 J \int_0^1 J']\Lambda'M' + \bar{\alpha}^2 MQ^{-1}Q_1\Lambda \int_0^1 J \int_0^1 J'\Lambda'Q_1'Q'^{-1}M' \\
&\quad + \bar{\alpha}^2 M_{11}\Lambda \int_0^1 JJ'\Lambda'M_{11}' - 2\bar{\alpha}M_1\Lambda[\int_0^1 JJ' - \int_0^1 J \int_0^1 J']\Lambda'M_1' \\
&\quad - \bar{\alpha}^2 M_1 Q^{-1}Q_1\Lambda \int_0^1 J \int_0^1 J'\Lambda'M_1' - \bar{\alpha}^2 M_1\Lambda \int_0^1 J \int_0^1 J'\Lambda'Q_1'Q'^{-1}M_1'.
\end{aligned}$$

Also,

$$T^{-1}SW \Rightarrow \frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})} = O_p(1),$$

where $\bar{\lambda}$ is the value of λ that minimizes $SSR^*(\lambda)$ over the set $[\epsilon, 1-\epsilon]$,

$$\begin{aligned}
SSR^{r*} &\equiv (1 + \bar{\alpha}^2)\sigma_e^2 + (1 + \bar{\alpha}^2)\int_0^1 J'\Lambda'Q\Lambda J - 2\bar{\alpha}\int_0^1 J'\Lambda'Q_1\Lambda J \\
&\quad - \int_0^1 J'\Lambda'[Q - \bar{\alpha}Q_1]'Q^{-1}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 J,
\end{aligned}$$

and

$$\begin{aligned}
SSR^*(\lambda) &\equiv (1 + \bar{\alpha}_u^2)\sigma_e^2 + (1 + \bar{\alpha}_u^2)\int_0^1 J'\Lambda'Q\Lambda J - 2\bar{\alpha}_u \int_0^1 J'\Lambda'Q_1\Lambda J \\
&\quad - \frac{1}{\lambda}\int_0^\lambda J'\Lambda'[Q - \bar{\alpha}_u Q_1]'Q^{-1}[Q - \bar{\alpha}_u Q_1]\Lambda \int_0^\lambda J \\
&\quad - \frac{1}{1-\lambda}\int_\lambda^1 J'\Lambda'[Q - \bar{\alpha}_u Q_1]'Q^{-1}[Q - \bar{\alpha}_u Q_1]\Lambda \int_\lambda^1 J
\end{aligned}$$

with

$$\begin{aligned}
\bar{\alpha}_u &= \{\sigma_e^2 + \int_0^1 J'\Lambda'Q\Lambda J - \frac{1}{\lambda}\int_0^\lambda J'\Lambda Q_1'Q^{-1}Q_1\Lambda \int_0^\lambda J - \frac{1}{1-\lambda}\int_\lambda^1 J'\Lambda Q_1'Q^{-1}Q_1\Lambda \int_\lambda^1 J\}^{-1} \\
&\quad \times \{\int_0^1 J'\Lambda'Q_1\Lambda J - \frac{1}{\lambda}\int_0^\lambda J'\Lambda'Q_1\Lambda \int_0^\lambda J - \frac{1}{1-\lambda}\int_\lambda^1 J'\Lambda'Q_1\Lambda \int_\lambda^1 J\}.
\end{aligned}$$

Again, the theoretical expressions are quite complex and do not lend themselves easily to allow a comparison of the tests under general conditions. So we present simulations of the

means and cdfs of the limit random variables for the DGP used in the simulations of Section 4. The results are presented in Figure 10, again for $p/T = 0.05$, 0.10 and 0.50. The results are qualitatively similar to the case with no lagged dependent variable and a correction for serial correlation. The mean of the limit random variable of the $q\hat{L}L$ test statistic initially increases (in absolute value) with Λ but quickly starts decreasing as Λ increases and at a lower value of Λ than for the case with a correction for serial correlation. The decrease in the case of $p/T = 0.50$ happens so quickly that on average, the $q\hat{L}L$ test does not reject for any value of Λ. On the other hand, the mean of the limit value of the sup-Wald test does not decrease as Λ increases but stabilizes to a level that implies a clear rejection. This becomes more pronounced as the intensity of the level shifts p/T decreases. The cdfs of the limit random variables are presented for $\Lambda = 1$ and $\Lambda = 15$. When $\Lambda = 1$ and p/T is small, the values imply a lower probability of rejection for the $q\hat{L}L$ than for the sup-Wald. When $\Lambda = 1$ and $p/T = 0.5$ or when $\Lambda = 15$, the cdf corresponding to the sup-Wald test is much further to the right of that of (minus) the $q\hat{L}L$ test so that the sup-Wald test has much higher power, in accordance with the simulations.

5.2 Single break model

We now consider the case where the parameters exhibit a single break of fixed magnitude at some fixed date:

$$\beta_t = \theta I_{t>[T\lambda_c]}, \tag{13}$$

where θ is a vector of constants representing the magnitudes of the break, I is the indicator function and $\lambda_c \in [0,1]$ represents the break fraction. The probability limits of the scaled statistics are stated in the following theorem for the three scenarios considered. We first consider the case where one imposes a priori the assumption that the errors are serially uncorrelated.

Theorem 7 *Assume that the $q\hat{L}L$ and sup-Wald (SW) tests are constructed from the static regression (5) with no correction for serial correlation and that the process is generated by (1) with β_t specified by (13), then*

$$p\lim_{T\to\infty} T^{-1}q\hat{L}L = tr[\xi\bar{V}_X^{-1/2}Q\theta\theta'Q'\bar{V}_X'^{-1/2}] = O_p(1)$$

uniformly in $\|\theta\|$, where

$$\bar{V}_X \equiv \sigma_e^2 Q + \lambda_c(1-\lambda_c)M\theta\theta'M',$$

23

and

$$\begin{aligned}
\xi &= \lambda_c(\lambda_c - 1) + \frac{1}{c}(1 - \lambda_c + \frac{1}{2}\lambda_c^2) - \frac{1}{c}(1 - \lambda_c)^2 e^{-c} \\
&\quad - \frac{1}{2c}(1 - \lambda_c)^2 e^{-2c} - \frac{1}{c}(1 - \lambda_c)e^{-c\lambda_c} + \frac{1}{c}(1 - \lambda_c)e^{c(\lambda_c - 2)} - \frac{1}{2c}e^{2c(\lambda_c - 1)} \\
&\quad - \frac{1}{2c(1 - e^{-2c})}\left[\lambda_c - 1 + (1 - \lambda_c)e^{-2c} + e^{-c\lambda_c} - e^{c(\lambda_c - 2)}\right]^2.
\end{aligned}$$

Also,

$$p\lim_{T \to \infty} T^{-1}SW = \frac{\lambda_c(1 - \lambda_c)}{\sigma_e^2}\theta'Q\theta = O_p(\|\theta\|^2)$$

so that SW dominates $q\hat{L}L$ for sufficiently large $\|\theta\|$.

The results in this case are qualitatively different from those of the other models in that the limit values are non-stochastic functions of $\|\theta\|$ and other parameters of the DGP. The results reveal that the sup-Wald test dominates the $q\hat{L}L$ test for large enough values of the magnitude of the shifts $\|\theta\|$, irrespective of the other features of the DGP (within the class considered). This is so because the limit value of the sup-Wald test is an increasing function of $\|\theta\|$ while the limit value of the $q\hat{L}L$ test is uniformly bounded in $\|\theta\|$. For most DGPs however, the limit values of both test statistics imply a rejection of the null hypothesis when $\|\theta\|$ is not too small. Nonetheless, it is interesting to compare the limit values as a function of the magnitude of change. We do so again for the DGP considered in Section 4. The results are presented in Figure 11 for θ ranging between 0 and 5. It turns out that the sup-Wald test is far more efficient than the $q\hat{L}L$ test unless the magnitude of change is very small (in which case, the probability of rejection is small). In particular, the limit value of the $q\hat{L}L$ test flattens out as θ increases while that of the sup-Wald test increases exponentially. The implication for the finite sample power of the two tests is that if a 5% test is used both tests will have good and monotonic power (as documented in the simulations), but if a (unrealistic) very small size was used the $q\hat{L}L$ would have lower power than the sup-Wald. The next theorem covers the case where a non-parametric correction for serial correlation in the errors is applied.

Theorem 8 *Assume that the $q\hat{L}L$ and sup-Wald (SW) tests are constructed from the static regression (5) and that the process is generated by (1) with β_t specified by (13). If potential serial correlation in the errors is accounted for using an estimate of the form (11) with the Bartlett kernel and the bandwidth chosen by Andrew's (1991) data dependent method based*

24

on an AR(1) approximation. a) If $\theta'(M_{11} - M)\theta = 0$ as $\|\theta\| \to \infty$, then

$$p \lim_{T \to \infty} T^{-1} q\hat{L}L = tr[\xi \bar{h}(0)^{-1/2} Q\theta\theta' Q' \bar{h}(0)'^{-1/2}] = O_p(\|\theta\|^{-4/3})$$

uniformly in $\|\theta\|$, where $\bar{h}(0) \equiv \lim_{T \to \infty} T^{-1/3} \sum_{j=-T}^{T} \kappa(j, m_T) \bar{\Gamma}_j$ with $m_T \equiv 1.1447(\bar{\alpha}T)^{1/3}$, $\bar{\alpha} \equiv 4\bar{\rho}^2/(1 - \bar{\rho}^2)^2$, $\bar{\Gamma}_j \equiv \lambda_c(1 - \lambda_c)M_{jj}\theta\theta' M'_{jj}$, and

$$\bar{\rho} \equiv \frac{\lambda_c(1 - \lambda_c)M^2\theta^2}{\sigma_e^2 Q + \lambda_c(1 - \lambda_c)M^2\theta^2}.$$

b) If $\theta'(M_{11} - M)\theta \neq 0$, as $\|\theta\| \to \infty$, then

$$p \lim_{T \to \infty} T^{-1} q\hat{L}L = tr[\xi \bar{h}(0)^{-1/2} Q\theta\theta' Q' \bar{h}(0)'^{-1/2}] = O_p(1)$$

uniformly in $\|\theta\|$, where $\bar{h}(0)$ is defined as above and

$$\bar{\rho} \equiv \frac{\lambda_c(1 - \lambda_c)M_{11}^2\theta^2}{\sigma_e^2 Q + \lambda_c(1 - \lambda_c)M^2\theta^2}.$$

For the sup-Wald test, since $\hat{\Gamma}_j \xrightarrow{p} 0$ for $j \neq 0$ because the test uses unrestricted residuals in the construction of $\hat{h}(0)$, the results stated in Theorem 8 continue to apply so that in both cases, SW dominates $q\hat{L}L$ for sufficiently large $\|\theta\|$.

When $\theta'(M_{11} - M)\theta = 0$, as is the case in the mean shift model, the limit of the sup-Wald test is an increasing function of $\|\theta\|$ while that of the $q\hat{L}L$ test is decreasing in $\|\theta\|$. Hence for $\|\theta\|$ large enough, the sup-Wald test dominates the $q\hat{L}L$ test. Things are different when $\theta'(M_{11} - M)\theta \neq 0$, as is the case in the DGP used in Section 4 with an AR(1) process as the regressor. Here, the limit of the sup-Wald test increases as $\|\theta\|$ increases while the limit of the $q\hat{L}L$ test is uniformly bounded in $\|\theta\|$. To get an idea of the implications of this, Figure 12 presents the limit values of the $q\hat{L}L$ and sup-Wald tests as functions of $\|\theta\|$ for the DGP of Section 4. It is clear that the limit value of the sup-Wald test increases rapidly with $\|\theta\|$. However, the limit value of the $q\hat{L}L$ test initially increases with $\|\theta\|$ but rapidly decreases to stabilize at a low value, implying non-rejection. Hence, the $q\hat{L}L$ test will only reject for small values of $\|\theta\|$ while the sup-Wald test will reject for any value except, of course, very small ones.

The last theorem covers the case with a lagged dependent variable as a regressor.

Theorem 9 *Assume that the $q\hat{L}L$ and sup-Wald (SW) tests are constructed from a dynamic regression of the form (6) with a lagged dependent variable as a regressor and the process is*

generated by (1) with β_t specified by (13). Then,

$$p \lim_{T \to \infty} T^{-1} q\hat{L}L = tr[\xi \bar{V}_X^{-1/2}[Q - \bar{\alpha}Q_1]\theta\theta'[Q - \bar{\alpha}Q_1]'\bar{V}_X'^{-1/2}] = O_p(1)$$

uniformly in $\|\theta\|$, where ξ is as defined in Theorem 7,

$$\bar{\alpha} \equiv \left[\sigma_e^2 + (1 - \lambda_c)\theta'Q\theta - (1 - \lambda_c)^2\theta'Q_1Q^{-1}Q_1\theta\right]^{-1}[\lambda_c(1 - \lambda_c)\theta'Q_1\theta],$$

and

$$\begin{aligned}
\bar{V}_X &\equiv (1 + \bar{\alpha}^2)\sigma_e^2 Q + \bar{\alpha}^2(1 - \lambda_c)M_{11}\theta\theta'M_{11}' \\
&\quad + (1 - \lambda_c)M[\lambda_c\theta\theta' + \bar{\alpha}^2(1 - \lambda_c)Q^{-1}Q_1\theta\theta'Q_1'Q^{-1'}]M' \\
&\quad - \bar{\alpha}(1 - \lambda_c)M_1[2\lambda_c\theta\theta' + \bar{\alpha}(1 - \lambda_c)[Q^{-1}Q_1\theta\theta' + \theta\theta'Q_1'Q^{-1'}]]M_1'.
\end{aligned}$$

Also,

$$p \lim_{T \to \infty} T^{-1}SW = \left(\frac{SSR^{r^*} - \sigma_e^2}{\sigma_e^2}\right) = O_p(\|\theta\|^2),$$

where

$$\begin{aligned}
SSR^{r^*} &\equiv (1 + \bar{\alpha}^2)\sigma_e^2 + (1 - \lambda_c)(1 + \bar{\alpha}^2)\theta'Q\theta - 2(1 - \lambda_c)\bar{\alpha}\theta'Q_1\theta \\
&\quad - (1 - \lambda_c)^2\theta'[Q - \bar{\alpha}Q_1]'Q^{-1}[Q - \bar{\alpha}Q_1]\theta
\end{aligned}$$

so that SW dominates $q\hat{L}L$ for sufficiently large $\|\theta\|$.

Here the results are qualitatively similar to the case with no lagged dependent variable and a correction for serial correlation when $\theta'(M_{11} - M)\theta \neq 0$. Both statistics diverge at rate T but the limit value of the $q\hat{L}L$ is uniformly bounded in $\|\theta\|$ while that of the sup-Wald test is an increasing function of $\|\theta\|$. Hence, for large enough $\|\theta\|$, the sup-Wald test dominates. Though the theoretical expressions are complex, one can easily obtain numerical values. This is done again for the case of the DGP of Section 4. The limit values as functions of $\|\theta\|$ are plotted in Figure 13. Again, this function is rapidly increasing as $\|\theta\|$ increases for the sup-Wald test so that it rejects for all but very small values of $\|\theta\|$. On the other hand, while the limit value of the $q\hat{L}L$ initially increases with $\|\theta\|$, it quickly reverts back to stabilize at a small value implying non-rejection. Hence, the sup-Wald dominates the $q\hat{L}L$, even for relatively small values of $\|\theta\|$.

26

6 Conclusion

The usual methodology used to analyze the power of test statistics and suggest tests with "optimal" properties relies on a local asymptotic framework where the magnitude of change shrinks towards zero as the sample size increases. This approach was adopted by Andrews and Ploberger (1994) who considered optimal tests based on maximizing some weighted average of the local asymptotic power function. It was also the approach adopted by EM to devise so-called optimal tests for general parameter variations. Our simulations have shown that the $q\hat{L}L$ test, labelled as optimal under this criterion, has very poor finite sample properties. It offers gains over the sup-Wald test, which is not an optimal test under this criterion, that are trivial when the break is small and there is no evidence of serial correlation in the residuals of the regression. This minor gain is achieved at the expense of very large power losses when considering models with serially correlated errors or dynamic models including a lagged dependent variable. Our simulation results are corroborated by the use of an alternative asymptotic framwork whereby the break sizes are fixed. Under this criterion, the sup-Wald test dominates the $q\hat{L}L$ test, which in some cases has zero relative asymptotic efficiency (i.e., power goes to zero as the break size increases).

Our research, in conjunction with other work, has several implications. First, it implies that one should abandon the use of a local asymptotic framework where the breaks are local to zero in delivering optimality criteria. These type of frameworks do not yield useful predictions about the finite sample properties of tests. Second, as argued in Perron (2006), one should also abandon partial-sums type tests for which only a model restricted to satisfy the null hypothesis of no change is used. These include the CUSUM, LM and $q\hat{L}L$ tests, among many others. These tests are plagued by the problem of a non-monotonic power function such that the power of the test can go to zero as the magnitude of change increases. There is no available alternative other than attempting to model, as best as possible, the nature of the changes present in the data. The sup-Wald test does this to a first approximation by taking the largest break into account. The UDmax test of Bai and Perron (1998) improves upon the sup-Wald test by taking more breaks into account. It may be the case that these tests are prone to size-distortions under the null hypothesis. Nevertheless, this is not a reason to abandon them in favor of tests with poor power properties. On the contrary, our work and others show that research should be directed at alleviating such size distortion problems. For progress in this direction see, e.g., Diebold and Chen (1996), Hansen (2000), Kejriwal (2009) and Prodan (2008).

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817-858.

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

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

Andrews, D.W.K., Ploberger,W., 1995. Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. Annals of Statistics 23, 1609-1629.

Bahadur, R.R., 1960. Stochastic comparison of tests. Annals of Mathematical Statistics 31, 276-295.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47-78.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1-22.

Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society B 37, 149-163.

Deng,A., Perron, P., 2008. A non-local perspective on the power properties of the CUSUM and CUSUM of square tests for structural change. Journal of Econometrics 142, 212-240.

Diebold, F.X., Chen, C., 1996. Testing structural stability with endogenous breakpoints: a size comparison of analytic and bootstrap procedures. Journal of Econometrics 70, 221-241.

Elliott,G., Müller, U.K., 2006. Efficient tests for general persistent time variation in regression coefficients. Review of Economic Studies 73, 907-940.

Franzini, L., Harvey A., 1983. Testing for deterministic trend and seasonal components in time series models. Biometrika 70, 673-682.

Garcia, R., Perron, P., 1996. An analysis of the real interest rate under regime shifts. Review of Economics and Statistics 78, 111-125.

Gardner, L.A., 1969. On detecting changes in the mean of normal variates. The Annals of Mathematical Statistics 40, 116-126.

Georgiev, I., 2002. Functional weak limit theory for rare outlying events. Unpublished manuscript, European University Institute.

Giacomini, R., Rossi, B., 2006. How stable is the forecasting performance of the yield curve for output growth? Oxford Bulletin of Economics and Statistics 68, 783-795.

Hansen, B.E., 2000. Testing for structural change in conditional models. Journal of Econometrics 97, 93-115.

Kim, D., Perron, P., 2009. Assessing the relative power of structural break tests using a framework based on the approximate Bahadur slope. *Journal of Econometrics* 149 26-51.

Kejriwal, M., 2009. Tests for a shift in mean with good size and monotonic power. Economics Letters 102, 78-82.

Leipus. R., Viano, M.C., 2003. Long memory and stochastic trends. Statistics and Probability Letters 61, 177-190.

Musso, A., Stracca, L., van Dijk, D., 2009. Instability and nonlinearity in the Euro area Phillips curve. International Journal of Central Banking 5, 181-212.

Nelson, F.D., Savin, N.E., 1990. The danger of extrapolating asymptotic local power. Econometrica 58, 977-981.

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

Nyblom, J., 1989. Testing for constancy of parameters over time. Journal of the American Statistical Association 84, 223-230.

Perron, P., 1990. Testing for a unit root in a time series with a changing mean. Journal of Business and Economic Statistics 8, 153-162.

Perron, P., 2006. Dealing with Structural Breaks. In Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory, K. Patterson and T.C. Mills (eds.), Palgrave Macmillan, 2006, 278-352.

Phillips, P.C.B., 1988. Regression theory for near-integrated time series. Econometrica 56, 1021-1044.

Prodan, R., 2008. Potential pitfalls in determining multiple structural changes with an application to purchasing power parity. Journal of Business and Economic Statistics 26, 50-65.

Shively, T., 1988. An exact test for a stochastic coefficient in a time series regression model. Journal of Time Series Analysis 9, 81-88.

Appendix

Preliminaries for Theorems 1-9: Let $\hat{u}_t = \hat{V}_X^{-1/2} X_t \hat{e}_t$ and $\hat{\eta}_t$ be a $k \times 1$ vector of residuals whose ith component is the residual from a regression of $\{\hat{\omega}_{t,i}\}$ on $\{\bar{r}^t\}$, as defined in Step 5. Then from the definition in Step 6 we have $q\hat{L}L = tr(\hat{Q})$, where

$$
\begin{aligned}
\hat{Q} &= \bar{r} \sum_{t=1}^{T} \hat{\eta}_t \hat{\eta}_t' - \sum_{t=1}^{T} \hat{u}_t \hat{u}_t' \\
&= \bar{r} \sum_{t=1}^{T} [\hat{\omega}_t - (\sum_{t=1}^{T} \bar{r}^{2t})^{-1} (\sum_{t=1}^{T} \bar{r}^t \hat{\omega}_t) \bar{r}^t][\hat{\omega}_t - (\sum_{t=1}^{T} \bar{r}^{2t})^{-1} (\sum_{t=1}^{T} \bar{r}^t \hat{\omega}_t) \bar{r}^t]' - \sum_{t=1}^{T} \hat{u}_t \hat{u}_t' \\
&= \sum_{t=1}^{T} \hat{\omega}_t \hat{\omega}_t' - \sum_{t=1}^{T} \hat{u}_t \hat{u}_t' - \frac{c}{T} \sum_{t=1}^{T} \hat{\omega}_t \hat{\omega}_t' - (\sum_{t=1}^{T} \bar{r}^{2t})^{-1} (\sum_{t=1}^{T} \bar{r}^t \hat{\omega}_t)(\sum_{t=1}^{T} \bar{r}^t \hat{\omega}_t)', \quad (A.1)
\end{aligned}
$$

with $\bar{r} = 1 - c/T$. Using the fact that $\hat{\omega}_s = \bar{r}\hat{\omega}_{s-1} + \hat{u}_s - \hat{u}_{s-1}$, we have

$$
\hat{\omega}_s = \hat{u}_s - \frac{c}{T} \sum_{j=1}^{s-1} \bar{r}^{j-1} \hat{u}_{s-j} \quad (A.2)
$$

for $s = 2, ..., T$ and $\hat{\omega}_1 = \hat{u}_1$. Summing from 1 through $t - 1$ gives

$$
\sum_{s=1}^{t-1} \hat{\omega}_s = \sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{u}_{t-s}. \quad (A.3)
$$

We can then show that the first two terms of (A.1) are given by

$$
\begin{aligned}
\sum_{t=1}^{T} \hat{\omega}_t \hat{\omega}_t' - \sum_{t=1}^{T} \hat{u}_t \hat{u}_t' &= \frac{c^2}{T^2} \sum_{t=1}^{T} [\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{u}_{t-s}][\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{u}_{t-s}]' \\
&\quad - \frac{c}{T} \sum_{t=1}^{T} \hat{u}_t [\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{u}_{t-s}]' - \frac{c}{T} \sum_{t=1}^{T} [\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{u}_{t-s}] \hat{u}_t'. \quad (A.4)
\end{aligned}
$$

For the fourth term of (A.1), we have

$$
\sum_{t=1}^{T} \bar{r}^t \hat{\omega}_t = \sum_{t=1}^{T} \bar{r}^t \hat{u}_t - \frac{c}{T} \sum_{t=1}^{T} \bar{r}^t \sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{u}_{t-s}. \quad (A.5)
$$

Since $\hat{u}_t = \hat{V}_X^{-1/2} X_t \hat{e}_t$, it is useful to derive an expression in terms of $\hat{v}_t = X_t \hat{e}_t$, separating $\hat{V}_X$ from $\hat{u}_t$. By substituting (A.4) and (A.5) into (A.1) we obtain

$$
\begin{aligned}
\hat{Q} = \hat{V}_X^{-1/2} &\left[c^2 A_{1,T} - c[A_{2,T} + A_{2,T}'] - cA_{3,T} \right. \\
&\left. - A_{6,T}^{-1} \times [A_{4,T} - cA_{5,T}][A_{4,T} - cA_{5,T}]' \right] \hat{V}_X'^{-1/2} + \tau_t, \quad (A.6)
\end{aligned}
$$

where τ_t is of lower order than the other terms uniformly in $||\Lambda||$, $\hat{V}_X = T^{-1} \sum_{t=1}^{T} \hat{v}_t \hat{v}_t'$, $A_{1,T} = T^{-2} \sum_{t=1}^{T} [\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{v}_{t-s}][\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{v}_{t-s}]'$, $A_{2,T} = T^{-1} \sum_{t=1}^{T} \hat{v}_t [\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{v}_{t-s}]'$, $A_{3,T} =$

$T^{-1} \sum_{t=1}^{T} \hat{v}_t \hat{v}_t'$, $A_{4,T} = T^{-1/2} \sum_{t=1}^{T} \bar{r}^t \hat{v}_t$, $A_{5,T} = T^{-3/2} \sum_{t=1}^{T} \bar{r}^t \sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{v}_{t-s}$, and $A_{6,T} = T^{-1} \sum_{t=1}^{T} \bar{r}^{2t} \overset{p}{\to} (1/2c)[1 - e^{-2c}]$. Our task is then to derive the limits of $\hat{V}_X$ and $A_{i,T}$, $i = 1, ..., 5$.

For the proofs of Theorems 2-4 we shall use the following Lemma, whose results are by now standard (e.g., Phillips, 1988).

Lemma A.1 *Let $y_t = \sum_{i=1}^{t} (1 - (c/T))^{t-i} \varepsilon_i$ and assume that ε_t is uniform mixing with mixing coefficient ϕ of size $-r/(2r - 2)$ or strong mixing with mixing coefficient α of size $-r/(r - 2)$, $r > 2$. Suppose further that $E|\varepsilon| < \Delta < \infty$ and $E(\varepsilon_t) = 0$ for $t = 1, 2....$ If $\{\varepsilon_t\}$ is globally covariance stationary with nonsingular global covariance matrix $\Lambda = \lim_{T \to \infty} var(T^{-1/2} \sum_{t=1}^{T} \varepsilon_t)$, then a) $T^{-1/2} y_{[Tr]} \Rightarrow \Lambda W_c(r)$; b) $T^{-3/2} \sum_{t=1}^{T} y_t \Rightarrow \Lambda \int_0^1 W_c(r) dr$; c) $T^{-2} \sum_{t=1}^{T} y_t y_t' \Rightarrow \Lambda \int_0^1 W_c(r) W_c(r)' dr \Lambda$; and d) $T^{-1} \sum_{t=1}^{[Tr]} (1 - (c/T))^s \sum_{t=1}^{[Tr]-s} \varepsilon_j \Rightarrow (1/c)\Lambda[W(r) - W_c(r)]$, where $W_c(r) = W(r) - c \int_0^r e^{-c\varsigma} W(r - \varsigma) d\varsigma$.*

Proof of Theorem 1 (random walk parameters, i.i.d. errors assumed): Under H_0, the OLS estimate of β is $\hat{\beta} = [T^{-1} \sum_{t=1}^{T} X_t X_t']^{-1}[T^{-1} \sum_{t=1}^{T} X_t X_t' \beta_t + T^{-1} \sum_{t=1}^{T} X_t e_t]$, so that $T^{-1/2} \hat{\beta} \Rightarrow \Lambda \int_0^1 W(r) dr$. The OLS residuals are given by

$$\hat{e}_t = e_t + X_t' \beta_t - X_t' \hat{\beta}, \tag{A.7}$$

which implies that

$$\hat{v}_t = X_t e_t + X_t X_t' \beta_t - X_t X_t' \hat{\beta}. \tag{A.8}$$

Using these representations, we have the following limit results for the components of $\hat{Q}$:

$$T^{-1} \hat{V}_X$$
$$= T^{-2} \sum_{t=1}^{T} \hat{v}_t \hat{v}_t'$$
$$= T^{-2} \sum_{t=1}^{T} \{X_t X_t' \beta_t \beta_t' X_t X_t' - X_t X_t' \beta_t \hat{\beta}' X_t X_t' - X_t X_t' \hat{\beta} \beta_t' X_t X_t' + X_t X_t' \hat{\beta} \hat{\beta}' X_t X_t'\} + o_p(1)$$
$$\Rightarrow M\Lambda[\int_0^1 WW' - \int_0^1 W \int_0^1 W']\Lambda' M' \equiv \bar{V}_X = O_p(\|\Lambda\|^2),$$

$$T^{-2} A_{1,T}$$
$$= T^{-4} \sum_{t=1}^{T} [\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{v}_{t-s}][\sum_{s=1}^{t-1} \bar{r}^{s-1} \hat{v}_{t-s}]'$$
$$= T^{-4} \sum_{t=1}^{T} \{[\sum_{s=1}^{t-1} \bar{r}^{s-1} X_{t-s} X_{t-s}' \beta_{t-s}][\sum_{s=1}^{t-1} \bar{r}^{s-1} X_{t-s} X_{t-s}' \beta_{t-s}]'$$
$$- [\sum_{s=1}^{t-1} \bar{r}^{s-1} X_{t-s} X_{t-s}' \beta_{t-s}][\sum_{s=1}^{t-1} \bar{r}^{s-1} X_{t-s} X_{t-s}' \hat{\beta}]' - [\sum_{s=1}^{t-1} \bar{r}^{s-1} X_{t-s} X_{t-s}' \hat{\beta}][\sum_{s=1}^{t-1} \bar{r}^{s-1} X_{t-s} X_{t-s}' \beta_{t-s}]'$$
$$+ [\sum_{s=1}^{t-1} \bar{r}^{s-1} X_{t-s} X_{t-s}' \hat{\beta}][\sum_{s=1}^{t-1} \bar{r}^{s-1} X_{t-s} X_{t-s}' \hat{\beta}]'\} + o_p(1)$$

31

$$\Rightarrow \quad \frac{1}{c^2}Q\Lambda[\int_0^1(W-W_c)(W-W_c)']\Lambda'Q' - \frac{1}{c}Q\Lambda\int_0^1\left(\int_0^r e^{-c\varsigma}d\varsigma\right)(W-W_c)dr\int_0^1 W'\Lambda'Q'$$

$$-\frac{1}{c}Q\Lambda\int_0^1 W[\int_0^1\left(\int_0^r e^{-c\varsigma}d\varsigma\right)(W-W_c)dr]'\Lambda'Q' + Q\Lambda\int_0^1\left(\int_0^r e^{-c\varsigma}d\varsigma\right)^2 dr\int_0^1 W\int_0^1 W'\Lambda'Q'$$

$$= \quad O_p(\|\Lambda\|^2),$$

$$T^{-2}A_{2,T} = T^{-3}\sum_{t=1}^{T}\hat{v}_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}]'$$

$$= \quad T^{-3}\sum_{t=1}^{T}\{X_tX_t'\beta_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]' - X_tX_t'\beta_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\hat{\beta}]'$$

$$-X_tX_t'\hat{\beta}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]' + X_tX_t'\hat{\beta}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\hat{\beta}]'\} + o_p(1)$$

$$\Rightarrow \quad \frac{1}{c}Q\Lambda[\int_0^1(WW' - WW_c')]\Lambda'Q' - Q\Lambda\int_0^1\left(\int_0^r e^{-c\varsigma}d\varsigma\right)Wdr\int_0^1 W'\Lambda'Q'$$

$$-\frac{1}{c}Q\Lambda\int_0^1 W[\int_0^1(W-W_c)]'\Lambda'Q' + Q\Lambda\int_0^1\int_0^r e^{-c\varsigma}d\varsigma dr\int_0^1 W\int_0^1 W'\Lambda'Q' = O_p(\|\Lambda\|^2),$$

$$T^{-2}A_{3,T} \to 0_{k\times k},$$

$$T^{-1}A_{4,T} = \quad T^{-3/2}\sum_{t=1}^{T}\bar{r}^t\hat{v}_t = T^{-3/2}\sum_{t=1}^{T}\bar{r}^t X_tX_t'\beta_t - T^{-3/2}\sum_{t=1}^{T}\bar{r}^t X_tX_t'\hat{\beta} + o_p(1)$$

$$\Rightarrow \quad Q\Lambda\int_0^1 e^{-cr}W - Q\Lambda\int_0^1 e^{-cr}\int_0^1 W = O_p(\|\Lambda\|),$$

$$T^{-1}A_{5,T} = \quad T^{-5/2}\sum_{t=1}^{T}\bar{r}^t\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}$$

$$= \quad T^{-5/2}\sum_{t=1}^{T}\bar{r}^t\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s} - T^{-5/2}\sum_{t=1}^{T}\bar{r}^t\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\hat{\beta} + o_p(1)$$

$$\Rightarrow \quad \frac{1}{c}Q\Lambda[\int_0^1 e^{-cr}(W-W_c)] - Q\Lambda\int_0^1 e^{-cr}\int_0^r e^{-c\varsigma}d\varsigma dr\int_0^1 W = O_p(\|\Lambda\|).$$

Collecting terms, by (A.6) we have, $T^{-1}q\hat{L}L \Rightarrow tr\{\bar{V}_X^{-1/2}Q\Lambda\Xi\Lambda'Q'\bar{V}'^{-1/2}\}$, where

$$\Xi = \int_0^1 W_cW_c' - \int_0^1 WW' + \int_0^1 W\int_0^1 W' - \frac{2c}{1-e^{-2c}}\int_0^1 e^{-cr}W\int_0^1 e^{-cr}W'.$$

We now consider the limit of the sup-Wald test. The unrestricted OLS residuals allowing for a single break at some $\lambda \in [0,1]$ are given by

$$\hat{e}(\lambda)_t = e_t + X_t'\beta_t - X_1(\lambda)_t'\hat{\beta}_1 - X_2(\lambda)_t'\hat{\beta}_2, \tag{A.9}$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are OLS estimates of β_1 and β_2. It is straightforward to show that $T^{-1/2}\hat{\beta}_1 \Rightarrow (1/\lambda)\Lambda \int_0^\lambda W$ and $T^{-1/2}\hat{\beta}_2 \Rightarrow (1/(1-\lambda))\Lambda \int_\lambda^1 W$. Using (A.9), the unrestricted sum of squared residuals is such that

$$
\begin{aligned}
T^{-2}SSR_T(\lambda) &= T^{-2}\sum_{t=1}^{T}\left[e_t + X_t'\beta_t - X_1(\lambda)_t'\hat{\beta}_1 - X_2(\lambda)_t'\hat{\beta}_2\right]^2 \\
&\Rightarrow [\int_0^1 W'\Lambda'Q\Lambda W] - \frac{1}{\lambda}\int_0^\lambda W'\Lambda'Q\Lambda \int_0^\lambda W - \frac{1}{1-\lambda}\int_\lambda^1 W'\Lambda'Q\Lambda \int_\lambda^1 W \\
&\equiv SSR^*(\lambda) = O_p(\|\Lambda\|^2),
\end{aligned}
$$

while using (A.7), the restricted sum of squared residuals is such that

$$
\begin{aligned}
T^{-2}SSR_T^r &= T^{-2}\sum_{t=1}^{T}\left[e_t + X_t'\beta_t - X_t'\hat{\beta}\right]^2 \\
&\Rightarrow \int_0^1 W'\Lambda'Q\Lambda W - \int_0^1 W'\Lambda'Q\Lambda \int_0^1 W \equiv SSR^{r*} = O_p(\|\Lambda\|^2).
\end{aligned}
$$

By the continuous mapping theorem, the limit of the sup-Wald test is given by

$$
T^{-1}SW = \left(\frac{T^{-2}SSR_T^r - T^{-2}SSR_T(\bar{\lambda}_T)}{T^{-2}SSR_T(\bar{\lambda}_T)}\right) \Rightarrow \left(\frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})}\right) = O_p(1),
$$

where $\bar{\lambda}_T$ is the value of λ that minimizes $SSR_T(\lambda)$ and $\bar{\lambda}$ minimizes $SSR^*(\lambda)$.

Proof of Theorem 2 (random walk parameters, correction for serial correlation):
If we take into account serial correlation in the errors, we replace $\hat{V}_X$ with the heteroskedasticity and autocorrelation robust estimate given by

$$
\hat{h}(0) = \hat{V}_X + 2\sum_{j=1}^{T-1}\kappa(j,m)\hat{\Gamma}_j,
$$

where $\hat{\Gamma}_j = T^{-1}\sum_{t=j+1}^{T}\hat{v}_t\hat{v}_{t-j}'$. For the $q\hat{L}L$ test, this is constructed by using $\hat{v}_t$ given in (A.8). We first consider the limit of $\hat{\Gamma}_j$. We have,

$$
\begin{aligned}
T^{-1}\hat{\Gamma}_j &= T^{-2}\sum_{t=j+1}^{T}\{X_tX_t'\beta_t\beta_{t-j}'X_{t-j}X_{t-j}' - X_tX_t'\beta_t\hat{\beta}'X_{t-j}X_{t-j}' \\
&\quad - X_tX_t'\hat{\beta}\beta_{t-j}'X_{t-j}X_{t-j}' + X_tX_t'\hat{\beta}\hat{\beta}'X_{t-j}X_{t-j}'\} + o_p(1) \\
&\Rightarrow M_{jj}\Lambda[\int_0^1 WW' - \int_0^1 W\int_0^1 W']\Lambda'M_{jj}' \equiv \bar{\Gamma}_j = O_p(\|\Lambda\|^2)
\end{aligned}
$$

uniformly in $j = o(T)$. In the next step, we consider the case of a single regressor, for simplicity. Andrews's (1991) method is based on choosing the bandwidth such that, for the Bartlett kernel $m = 1.1447(\alpha T)^{1/3}$, where $\alpha = 4\hat{\rho}^2/(1-\hat{\rho}^2)^2$ with $\hat{\rho}$ the OLS estimate in an AR(1) regression applied to $\hat{v}_t$. We have

$$
\hat{\rho} = \hat{\Gamma}_1/\hat{\Gamma}_0 \Rightarrow \frac{M_{11}^2\Lambda^2[\int_0^1 W^2 - (\int_0^1 W)^2]}{M^2\Lambda^2[\int_0^1 W^2 - (\int_0^1 W)^2]} = \frac{M_{11}^2}{M^2}.
$$

If $M_{11} = M$, the limit value is one and

$$T(\hat{\rho} - 1) \Rightarrow \frac{-W(1)^2 - (\int_0^1 W)^2 + W(1)\int_0^1 W}{\int_0^1 W^2 - (\int_0^1 W)^2} \equiv \phi_r = O_p(1).$$

Hence, $\hat{\rho} = 1 - \phi_r/T + o_p(1)$ so that $\alpha = O_p(T^2)$ and $m = O_p(T^{2/3}T^{1/3}) = O_p(T)$. Therefore, $\hat{h}(0) = O_p(\|\Lambda\|^2 T^2)$ and substituting its limit value instead of the limit value of $\hat{V}_X$ gives the stated result. If $\{M - M_{11}\}_{ii} > 0$ for some i, then $\hat{\rho}$ converges to a random variable with mean less than one. Hence, $\alpha = O_p(1)$ and $m = O_p(T^{1/3})$. Therefore $\hat{h}(0) = O_p(\|\Lambda\|^2 T^{4/3})$ giving the result in part(b) for the $q\hat{L}L$ test. For the sup-Wald test, we use the expression

$$Wald = \frac{1}{T}\left(\frac{T - 2k}{k}\right)\hat{\beta}' H' (H\hat{V}H')^{-1} H\hat{\beta},$$

where $\hat{V} \equiv T(\bar{X}'\bar{X})^{-1}\hat{V}_X(\bar{X}'\bar{X})^{-1}$. First, we have,

$$T^{-1/2}\hat{\beta}' H' = T^{-1/2}\hat{\beta}_1 - T^{-1/2}\hat{\beta}_2 \Rightarrow \frac{1}{\lambda}\Lambda \int_0^\lambda W - \frac{1}{1-\lambda}\Lambda \int_\lambda^1 W = O_p(1).$$

Second, $T^{-1/3}\hat{V} = diag(V1, V2)$, where

$$V1 \equiv (T^{-1}\sum_{t=1}^{[T\lambda]} XX)^{-1}(T^{-4/3}\hat{h}_1(0))(T^{-1}\sum_{t=1}^{[T\lambda]} XX)^{-1},$$

$$V2 \equiv (T^{-1}\sum_{t=[T\lambda]+1}^{T} XX)^{-1}(T^{-4/3}\hat{h}_2(0))(T^{-1}\sum_{t=[T\lambda]+1}^{T} XX)^{-1},$$

with $\hat{h}_1(0)$ the HAC robust variance estimator using

$$\hat{v}_t^1 = \begin{cases} X_t e_t + X_t X_t'\beta_t - X_t X_t'\hat{\beta}_1, & \text{for } t < [T\lambda], \\ 0, & \text{otherwise} \end{cases}$$

and $\hat{h}_2(0)$ using

$$\hat{v}_t^2 = \begin{cases} 0, & \text{for } t < [T\lambda], \\ X_t e_t + X_t X_t'\beta_t - X_t X_t'\hat{\beta}_2, & \text{otherwise.} \end{cases}$$

As in the proof for $q\hat{L}L$, we can show that $\hat{h}_1(0)$ and $\hat{h}_2(0)$ are both $O_p(\|\Lambda\|^2 T^2)$ if $M_{11} = M$. In this case,

$$SW \Rightarrow (\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} W - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda^*}^1 W)'Q'$$
$$\times [\bar{h}_1(0) + \bar{h}_2(0)]^{-1}Q(\frac{1}{\lambda^*}\Lambda \int_0^{\lambda^*} W - \frac{1}{1-\lambda^*}\Lambda \int_{\lambda*}^1 W) = O_p(1),$$

where λ^* is the value of λ that maximizes the limit of the Wald test over the all permissible λ. If $\{M_{11} - M\}_{ii} \neq 0$ for some i, the condition in the theorem does not hold $\hat{h}_1(0)$ and $\hat{h}_2(0)$ are both $O_p(\|\Lambda\|^2 T^{4/3})$ and $T^{-2/3}SW = O_p(1)$.

34

Proof of Theorem 3 (random walk parameters, lagged dependent variable): We now consider the dynamic regression. Under the null hypothesis, the OLS residuals are

$$\hat{e}_t = e_t - \hat{\alpha}e_{t-1} + X_t'\beta_t - \hat{\alpha}X_{t-1}'\beta_{t-1} - X_t'\hat{\beta}, \tag{A.10}$$

and

$$\hat{v}_t = X_t e_t - \hat{\alpha}X_t e_{t-1} + X_t X_t'\beta_t - \hat{\alpha}X_t X_{t-1}'\beta_{t-1} - X_t X_t'\hat{\beta}, \tag{A.11}$$

where the OLS estimates $\hat{\alpha}$ and $\hat{\beta}$ have the following limits:

$$\hat{\alpha} \Rightarrow \{\textstyle\int_0^1 W'\Lambda'Q\Lambda W - \int_0^1 W'\Lambda'Q_1'Q^{-1}Q_1\Lambda \int_0^1 W\}^{-1}$$
$$\times \{\textstyle\int_0^1 W'\Lambda'Q_1\Lambda W - \int_0^1 W'\Lambda'Q_1\Lambda \int_0^1 W\} \equiv \bar{\alpha} \tag{A.12}$$

and $T^{-1/2}\hat{\beta} \Rightarrow Q^{-1}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 W \equiv \bar{\beta}$. Consider first the $q\hat{L}L$ test. We take the limits of the components of (A.6). We have, $T^{-2}A_{3,T} \rightarrow 0_{k\times k}$,

$$T^{-1}\hat{V}_X = T^{-2}\sum_{t=1}^{T}\{X_t X_t'\beta_t\beta_t' X_t X_t' - \hat{\alpha}X_t X_t'\beta_t\beta_{t-1}' X_{t-1}X_t' - X_t X_t'\beta_t\hat{\beta}' X_t X_t'$$

$$-\hat{\alpha}X_t X_{t-1}'\beta_{t-1}\beta_t' X_t X_t' + \hat{\alpha}^2 X_t X_{t-1}'\beta_{t-1}\beta_{t-1}' X_{t-1}X_t' + \hat{\alpha}X_t X_{t-1}'\beta_{t-1}\hat{\beta}' X_t X_t'$$

$$-X_t X_t'\hat{\beta}\beta_t' X_t X_t' + \hat{\alpha}X_t X_t'\hat{\beta}\beta_{t-1}' X_{t-1}X_t' + X_t X_t'\hat{\beta}\hat{\beta}' X_t X_t'\} + o_p(1)$$

$$\Rightarrow\quad M\Lambda\textstyle\int_0^1 WW'\Lambda'M' - \bar{\alpha}M_1\Lambda\int_0^1 WW'\Lambda'M_1' - M\Lambda\int_0^1 W\bar{\beta}'M' - \bar{\alpha}M_1\Lambda\int_0^1 WW'\Lambda'M_1'$$

$$+\bar{\alpha}^2 M_{11}\Lambda\textstyle\int_0^1 WW'\Lambda'M_{11}' + \bar{\alpha}M_1\Lambda\int_0^1 W\bar{\beta}'M_1' - M\bar{\beta}\int_0^1 W'\Lambda'M'$$

$$+\bar{\alpha}M_1\bar{\beta}\textstyle\int_0^1 W'\Lambda'M_1' + M\bar{\beta}\bar{\beta}'M'$$

$$=\quad M\Lambda[\textstyle\int_0^1 WW' - \int_0^1 W \int_0^1 W']\Lambda'M' + \bar{\alpha}^2 MQ^{-1}Q_1\Lambda\int_0^1 W \int_0^1 W'\Lambda'Q_1'Q'^{-1}M'$$

$$+\bar{\alpha}^2 M_{11}\Lambda\textstyle\int_0^1 WW'\Lambda'M_{11}' - 2\bar{\alpha}M_1\Lambda[\int_0^1 WW' - \int_0^1 W \int_0^1 W']\Lambda'M_1'$$

$$-\bar{\alpha}^2 M_1 Q^{-1}Q_1\Lambda\textstyle\int_0^1 W \int_0^1 W'\Lambda'M_1' - \bar{\alpha}^2 M_1\Lambda\int_0^1 W \int_0^1 W'\Lambda'Q_1'Q'^{-1}M_1'$$

$$\equiv\quad \bar{V}_X = O_p(\|\Lambda\|^2),$$

$$T^{-2}A_{1,T}$$

$$=\quad T^{-4}\sum_{t=1}^{T}\{[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]([\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]' - \hat{\alpha}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s-1}'\beta_{t-s-1}]')$$

$$-[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]\hat{\beta}'[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}']'$$

$$-\hat{\alpha}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s-1}'\beta_{t-s-1}]([\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]' - \hat{\alpha}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s-1}'\beta_{t-s-1}]')$$

$$+\hat{\alpha}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s-1}'\beta_{t-s-1}]\hat{\beta}'[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}']'$$

$$-[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}']\hat{\beta}([\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]' - \hat{\alpha}\hat{\beta}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s-1}'\beta_{t-s-1}]')$$

$$+[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}']\hat{\beta}\hat{\beta}'[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}']'\} + o_p(1)$$

35

$$\Rightarrow \quad \frac{1}{c^2}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 (W - W_c)(W - W_c)'\Lambda'[Q - \bar{\alpha}Q_1]'$$
$$+ [Q - \bar{\alpha}Q_1]\Lambda \int_0^1 \left(\int_0^r e^{-c\varsigma}d\varsigma\right)^2 dr \int_0^1 W \int_0^1 W'\Lambda'[Q - \bar{\alpha}Q_1]'$$
$$- \frac{1}{c}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 \left(\int_0^r e^{-c\varsigma}d\varsigma\right)(W - W_c)dr \int_0^1 W'\Lambda'[Q - \bar{\alpha}Q_1]'$$
$$- \frac{1}{c}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 W \int_0^1 \left(\int_0^r e^{-c\varsigma}d\varsigma\right)(W - W_c)dr'\Lambda'[Q - \bar{\alpha}Q_1]' = O_p(\|\Lambda\|^2),$$

$$T^{-2}A_{2,T}$$
$$= T^{-3}\sum_{t=1}^T X_t X_t'\{\beta_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]' - \hat{\alpha}\beta_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s-1}'\beta_{t-s-1}]'$$
$$- \beta_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\hat{\beta}]' - \hat{\beta}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]' + \hat{\alpha}\hat{\beta}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s-1}'\beta_{t-s-1}]'$$
$$+ \hat{\beta}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\hat{\beta}]'\} - \hat{\alpha}T^{-3}\sum_{t=1}^T X_t X_{t-1}'\{\beta_{t-1}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\beta_{t-s}]'$$
$$- \hat{\alpha}\beta_{t-1}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s-1}'\beta_{t-s-1}]' - \beta_{t-1}[\sum_{s=1}^{t-1}\bar{r}^{s-1}X_{t-s}X_{t-s}'\hat{\beta}]'\} + o_p(1)$$

$$\Rightarrow \quad \frac{1}{c}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 (WW' - WW_c')\Lambda'[Q - \bar{\alpha}Q_1]'$$
$$- [Q - \bar{\alpha}Q_1]\Lambda \int_0^1 \left(\int_0^r e^{-c\varsigma}d\varsigma\right)W dr \int_0^1 W'\Lambda'[Q - \bar{\alpha}Q_1]'$$
$$- \frac{1}{c}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 W \int_0^1 (W - W_c)'\Lambda'[Q - \bar{\alpha}Q_1]'$$
$$+ [Q - \bar{\alpha}Q_1]\Lambda \int_0^1 \int_0^r e^{-c\varsigma}d\varsigma dr \int_0^1 W \int_0^1 W'\Lambda'[Q - \bar{\alpha}Q_1]' = O_p(\|\Lambda\|^2),$$

$$T^{-1}A_{4,T} = T^{-3/2}\sum_{t=1}^T \bar{r}^t X_t X_t'\beta_t - \hat{\alpha}T^{-3/2}\sum_{t=1}^T \bar{r}^t X_t X_{t-1}'\beta_{t-1} - T^{-3/2}\sum_{t=1}^T \bar{r}^t X_t X_t'\hat{\beta} + o_p(1)$$
$$\Rightarrow \quad [Q - \bar{\alpha}Q_1]\Lambda \int_0^1 e^{-cr}W - [Q - \bar{\alpha}Q_1]\Lambda \int_0^1 e^{-cr}\int_0^1 W = O_p(\|\Lambda\|)$$

$$T^{-1}A_{5,T} = T^{-5/2}\sum_{t=1}^T \bar{r}^t \sum_{s=1}^{t-1}\bar{r}^{s-1}\{X_{t-s}X_{t-s}'\beta_{t-s} - \hat{\alpha}X_{t-s}X_{t-s-1}'\beta_{t-s-1} - X_{t-s}X_{t-s}'\hat{\beta}\} + o_p(1)$$
$$\Rightarrow \quad \frac{1}{c}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 e^{-cr}(W - W_c) - [Q - \bar{\alpha}Q_1]\Lambda \int_0^1 e^{-cr}\int_0^r e^{-c\varsigma}d\varsigma dr \int_0^1 W = O_p(\|\Lambda\|).$$

Collecting terms according to (A.6) yields the stated result. We now consider the sup-Wald test. The regression is

$$y_t = \alpha_u y_{t-1} + X_1(\lambda)_t'\beta_1 + X_2(\lambda)_t'\beta_2 + e_t,$$

and the OLS estimates are

$$
\begin{bmatrix}
\hat{\alpha}_u \\
T^{-1/2}[\hat{\beta}_1, \hat{\beta}_2]'
\end{bmatrix}
=
\begin{bmatrix}
T^{-2}\sum_{t=2}^{T} y_{t-1}^2 & T^{-3/2}\sum_{t=2}^{T} y_{t-1}\bar{X}(\lambda)_t' \\
T^{-3/2}\sum_{t=2}^{T} \bar{X}(\lambda)_t y_{t-1} & T^{-1}\sum_{t=2}^{T} \bar{X}(\lambda)_t \bar{X}(\lambda)_t'
\end{bmatrix}^{-1}
\begin{bmatrix}
T^{-2}\sum_{t=2}^{T} y_{t-1}y_t \\
T^{-3/2}\sum_{t=2}^{T} \bar{X}(\lambda)_t y_t
\end{bmatrix}
$$

where $\bar{X}(\lambda)_t = [X_1(\lambda)_t, X_2(\lambda)_t]$. After some algebra, we obtain

$$
\hat{\alpha}_u \;\Rightarrow\; \{\textstyle\int_0^1 W'\Lambda'Q\Lambda W - \frac{1}{\lambda}\int_0^\lambda W'\Lambda Q_1' Q^{-1} Q_1 \Lambda \int_0^\lambda W - \frac{1}{1-\lambda}\int_\lambda^1 W'\Lambda Q_1' Q^{-1} Q_1 \Lambda \int_\lambda^1 W\}^{-1}
$$

$$
\times \{\textstyle\int_0^1 W'\Lambda'Q_1\Lambda W - \frac{1}{\lambda}\int_0^\lambda W'\Lambda'Q_1\Lambda \int_0^\lambda W - \frac{1}{1-\lambda}\int_\lambda^1 W'\Lambda'Q_1\Lambda \int_\lambda^1 W\} \equiv \bar{\alpha}_u, \text{(A.13)}
$$

and

$$
T^{-1/2}\hat{\beta}_1 \;\Rightarrow\; \frac{1}{\lambda}\Lambda Q^{-1}[Q - \bar{\alpha}_u Q_1]\int_0^\lambda W,
$$

$$
T^{-1/2}\hat{\beta}_2 \;\Rightarrow\; \frac{1}{1-\lambda}\Lambda Q^{-1}[Q - \bar{\alpha}_u Q_1]\int_\lambda^1 W.
$$

Using these results, the limit of $SSR(\lambda)_T$ is given by:

$$
\begin{aligned}
T^{-2}SSR(\lambda)_T \;=\;& T^{-2}\sum_{t=1}^{T}[X_t'\beta_t - \hat{\alpha}X_{t-1}'\beta_{t-1} - X_1(\lambda)_t'\hat{\beta}_1 - X_2(\lambda)_t'\hat{\beta}_2]^2 + o_p(1)\\
\Rightarrow\;& (1+\bar{\alpha}_u^2)\textstyle\int_0^1 W'\Lambda'Q\Lambda W - 2\bar{\alpha}_u \int_0^1 W'\Lambda'Q_1\Lambda W\\
& -\frac{1}{\lambda}\int_0^\lambda W'\Lambda'[Q-\bar{\alpha}_u Q_1]'Q^{-1}[Q-\bar{\alpha}_u Q_1]\Lambda\int_0^\lambda W\\
& -\frac{1}{1-\lambda}\int_\lambda^1 W'\Lambda'[Q-\bar{\alpha}_u Q_1]'Q^{-1}[Q-\bar{\alpha}_u Q_1]\Lambda\int_\lambda^1 W \equiv SSR(\lambda)^* = O_p(\|\Lambda\|^2),
\end{aligned}
$$

where $\bar{\alpha}_u$ is defined by (A.13). The limit of the restricted sum of squared residuals is

$$
\begin{aligned}
& T^{-2}SSR_T^r \\
=\;& T^{-2}\sum_{t=1}^{T}[X_t'\beta_t - \hat{\alpha}X_{t-1}'\beta_{t-1} - X_t\hat{\beta}]^2 + o_p(1)\\
\Rightarrow\;& (1+\bar{\alpha}^2)\textstyle\int_0^1 W'\Lambda'Q\Lambda W - 2\bar{\alpha}\int_0^1 W'\Lambda'Q_1\Lambda W - \int_0^1 W'\Lambda'[Q-\bar{\alpha}Q_1]'Q^{-1}[Q-\bar{\alpha}Q_1]\Lambda\int_0^1 W\\
\equiv\;& SSR^{r*} = O_p(\|\Lambda\|^2),
\end{aligned}
$$

where $\bar{\alpha}$ is defined by (A.12). By the continuous mapping theorem, the limit of the sup-Wald test is given by

$$
T^{-1}SW = \left(\frac{T^{-2}SSR_T^r - T^{-2}SSR_T(\bar{\lambda}_T)}{T^{-2}SSR_T(\bar{\lambda}_T)}\right) \Rightarrow \left(\frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})}\right) = O_p(1),
$$

with $\bar{\lambda}_T$ the value that minimizes $SSR_T(\lambda)$ and $\bar{\lambda}$ minimizes $SSR^*(\lambda)$. We next consider the special case of a mean shift model when $x_t = 1$ for all t with the dynamic regression. For the restricted model, the regression equation is

$$
y_t = \alpha y_{t-1} + \beta + e_t
$$

37

and the OLS estimates are such that

$$T(\hat{\alpha} - 1) = \frac{T^{-1}\sum_{t=2}^{T} y_{t-1}(\varepsilon_t + e_t - e_{t-1}) - [T^{-3/2}\sum_{t=2}^{T} y_{t-1}][T^{-1/2}\sum_{t=2}^{T}(\varepsilon_t + e_t - e_{t-1})]}{T^{-2}\sum_{t=2}^{T} y_{t-1}^2 - [T^{-3/2}\sum_{t=2}^{T} y_{t-1}]^2}$$

$$\Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1] - \sigma_e^2/\Lambda^2 - W(1)\int_0^1 W}{\int_0^1 W^2 - (\int_0^1 W)^2} \equiv \bar{\alpha} = O_p(1),$$

$$T^{1/2}\hat{\beta} \Rightarrow -\bar{\alpha}\Lambda \int_0^1 W + \Lambda W(1) \equiv \bar{\beta}.$$

For the unrestricted model, the regression equation is

$$y_t = \alpha_u y_{t-1} + \beta_1 I_t^1 + \beta_2 I_t^2 + e_t,$$

where $I_t^1 = 1$ if $t < [T\lambda]$ and 0 otherwise and $I_t^2 = 0$ if $t < [T\lambda]$ and 1 otherwise. Let $I_t = [I_t^1, I_t^2]$. The OLS estimates are such that

$$T(\hat{\alpha}_u - 1)$$

$$= \frac{T^{-1}\sum_{t=2}^{T} y_{t-1}(\varepsilon_t + e_t - e_{t-1}) - [T^{-3/2}\sum_{t=2}^{T} y_{t-1}I_t'][T^{-1}\sum_{t=2}^{T} I_t I_t'][T^{-1/2}\sum_{t=2}^{T} I_t(\varepsilon_t + e_t - e_{t-1})]}{T^{-2}\sum_{t=2}^{T} y_{t-1}^2 - [T^{-3/2}\sum_{t=2}^{T} y_{t-1}]^2}$$

$$\Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1] - \sigma_e^2/\Lambda^2 - \frac{1}{\lambda}\int_0^\lambda WW(\lambda) - \frac{1}{1-\lambda}\int_\lambda^1 WW(1-\lambda)}{\int_0^1 W^2 - \frac{1}{\lambda}(\int_0^\lambda W)^2 - \frac{1}{1-\lambda}(\int_\lambda^1 W)^2} \equiv \bar{\alpha}_u = O_p(1),$$

$$T^{1/2}\hat{\beta}_1 \Rightarrow -\bar{\alpha}_u \frac{1}{\lambda}\Lambda \int_0^\lambda W + \frac{1}{\lambda}\Lambda W(\lambda) \equiv \bar{\beta}_1,$$

$$T^{1/2}\hat{\beta}_2 \Rightarrow -\bar{\alpha}_u \frac{1}{1-\lambda}\Lambda \int_\lambda^1 W + \frac{1}{1-\lambda}\Lambda W(1-\lambda) \equiv \bar{\beta}_2.$$

Therefore, we have $\hat{\alpha}, \hat{\alpha}_u \xrightarrow{p} 1$ and $\hat{\beta}, \hat{\beta}_1, \hat{\beta}_2 \xrightarrow{p} 0$. Also, $T(\hat{\alpha} - \hat{\alpha}_u) \Rightarrow \bar{\alpha} - \bar{\alpha}_u$ and $T(\hat{\alpha}^2 - \hat{\alpha}_u^2) \Rightarrow 2(\bar{\alpha} - \bar{\alpha}_u)$. Consider first the $q\hat{L}L$ test. Since $x_t = 1$ for all t, $\hat{v}_t = \hat{e}_t$ and

$$\hat{v}_t = e_t - \hat{\alpha}e_{t-1} + \beta_t - \hat{\alpha}\beta_{t-1} - \hat{\beta}.$$

The components in (A.6) are such that (since e_t and ε_t are uncorrelated)

$$\hat{V}_X = T^{-1}\sum_{t=2}^{T} \hat{v}_t^2 \xrightarrow{p} \Lambda^2 + 2\sigma_e^2,$$

$$A_{1,T} = T^{-2}\sum_{t=2}^{T}[\sum_{s=1}^{t-1} \bar{r}^{s-1}(\varepsilon_{t-s} + e_{t-s} - e_{t-s-1})]^2 + o_p(1) \Rightarrow \Lambda^2 \int_0^1 W_c^2,$$

$$A_{2,T} = T^{-1}\sum_{t=2}^{T}(\varepsilon_t + e_t - e_{t-1})[\sum_{s=1}^{t-1} \bar{r}^{s-1}(\varepsilon_{t-s} + e_{t-s} - e_{t-s-1})] + o_p(1) \Rightarrow 0,$$

$$A_{3,T} = T^{-1}\sum_{t=2}^{T} \hat{v}_t^2 \xrightarrow{p} \Lambda^2 + 2\sigma_e^2,$$

38

$$A_{4,T} = T^{-1/2} \sum_{t=2}^{T} \bar{r}^t (\varepsilon_t + e_t - e_{t-1}) + o_p(1) \Rightarrow \Lambda W_c(1),$$

$$A_{5,T} = T^{-3/2} \sum_{t=2}^{T} \bar{r}^t \sum_{s=1}^{t-1} (\varepsilon_{t-s} + e_{t-s} - e_{t-s-1}) + o_p(1) \Rightarrow \Lambda \int_0^1 e^{-cr} W_c.$$

Therefore,

$$q\hat{L}L \Rightarrow \frac{c^2 \Lambda^2 \int_0^1 W_c^2 - c\Lambda^2 - 2c\sigma_e^2 - (2c\Lambda^2/1 - e^{-2c})[W_c(1) - c\int_0^1 e^{-cr} W_c]^2}{\Lambda^2 + 2\sigma_e^2}.$$

We next consider the sup-Wald test, $Wald(\lambda)_T = (SSR_T^r - SSR_T(\lambda))/T^{-1} SSR_T(\lambda)$. Using the expression for the restricted and the unrestricted residuals yields

$$SSR_T^r - SSR_T(\lambda)$$
$$= \sum_{t=2}^{T}[y_t - \hat{\alpha} y_{t-1} - \hat{\beta}]^2 - \sum_{t=2}^{T}[y_t - \hat{\alpha}_u y_{t-1} - \hat{\beta}_1 I_{t<[T\lambda]} - \hat{\beta}_2 I_{t\geq[T\lambda]}]^2 + o_p(1)$$
$$= (\hat{\alpha}^2 - \hat{\alpha}_u^2) \sum_{t=2}^{T} y_{t-1}^2 + T\hat{\beta}^2 - [T\lambda]\hat{\beta}_1^2 - (T - [T\lambda])\hat{\beta}_2^2$$
$$-2(\hat{\alpha} - \hat{\alpha}_u) \sum_{t=2}^{T} y_t y_{t-1} - 2(1 - \hat{\alpha})\hat{\beta} \sum_{t=2}^{T} y_{t-1} + 2(1 - \hat{\alpha}_u)[\hat{\beta}_1 \sum_{t=2}^{[T\lambda]} y_{t-1} + \hat{\beta}_2 \sum_{t=[T\lambda]+1}^{T} y_{t-1}] + o_p(1)$$
$$\Rightarrow -2(\bar{\alpha} - \bar{\alpha}_u)[(1/2)\Lambda^2 (W(1) - 1) - \sigma_e^2] + \Lambda^2 W(1)^2 - \Lambda^2 W(\lambda)^2 - \Lambda^2 W(1 - \lambda)^2$$
$$-\bar{\alpha}^2 \Lambda^2 (\int_0^1 W)^2 + \bar{\alpha}_u^2 \frac{1}{\lambda} \Lambda^2 (\int_0^\lambda W)^2 + \bar{\alpha}_u^2 \frac{1}{1-\lambda} \Lambda^2 (\int_\lambda^1 W)^2$$
$$\equiv N = O_p(\Lambda^2)$$

and

$$T^{-1} SSR_T(\lambda) = \sum_{t=2}^{T}[y_t - \hat{\alpha}_u y_{t-1} - \hat{\beta}_1 I_{t<[T\lambda]} - \hat{\beta}_2 I_{t\geq[T\lambda]}]^2$$
$$= T^{-1} y_T^2 - 2\hat{\alpha}_u T^{-1} \sum_{t=2}^{T} y_{t-1}(\varepsilon_t + e_t - e_{t-1})$$
$$\Rightarrow \Lambda^2 + 2\sigma_e^2 \equiv D = O_p(\Lambda^2).$$

Therefore by the continuous mapping theorem $SW \Rightarrow N/D$.

Proof of Theorem 4 (infrequent break parameter; *i.i.d.* errors assumed) : For the $q\hat{L}L$ test, the OLS estimate of β is given by

$$\hat{\beta} = (T^{-1} \sum_{t=1}^{T} X_t X_t')^{-1} (T^{-1} \sum_{t=1}^{T} X_t X_t' \beta_t + T^{-1} \sum_{t=1}^{T} X_t e_t) \Rightarrow \int_0^1 J(s) ds.$$

Using the expression (A.8) for $\hat{v}_t$, the limit of the variance is

$$\hat{V}_X = T^{-1} \sum_{t=1}^{T} \hat{v}_t \hat{v}_t' \Rightarrow \sigma_e^2 Q + M\Lambda[\int_0^1 JJ' - \int_0^1 J \int_0^1 J']\Lambda' M' \equiv \bar{V}_X = O_p(\|\Lambda\|^2).$$

The components of (A.6) are such that (again, $T^{-1}A_{3,T} \to 0_{k\times k}$)

$$T^{-1}A_{1,T}$$
$$= T^{-3}\sum_{t=1}^{T}[\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}][\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}]'$$
$$\Rightarrow Q\Lambda\int_0^1[\int_0^r e^{-c\varsigma}J(r-\varsigma)d\varsigma][\int_0^r e^{-c\varsigma}J(r-\varsigma)d\varsigma]'dr\Lambda'Q'$$
$$-Q\Lambda\int_0^1\int_0^r e^{-c\varsigma}d\varsigma\int_0^r e^{-c\varsigma}J(r-\varsigma)d\varsigma dr\int_0^1 J'\Lambda'Q'$$
$$-Q\Lambda\int_0^1 J\int_0^1\int_0^r e^{-c\varsigma}d\varsigma\int_0^r e^{-c\varsigma}J(r-\varsigma)'d\varsigma dr\Lambda'Q'+Q\Lambda\int_0^1\int_0^r e^{-c\varsigma}d\varsigma^2 dr\int_0^1 J\int_0^1 J'\Lambda'Q'$$
$$\equiv Q\Lambda\Psi_1\Lambda'Q' = O_p(\|\Lambda\|^2),$$

$$T^{-1}A_{2,T} = T^{-2}\sum_{t=1}^{T}\hat{v}_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}]'$$
$$\Rightarrow Q\Lambda\int_0^1 J(r)\int_0^r e^{-c\varsigma}J(r-\varsigma)d\varsigma dr\Lambda Q' - Q\Lambda\int_0^1\int_0^r e^{-c\varsigma}d\varsigma J(r)dr\int_0^1 J'\Lambda'Q'$$
$$-Q\Lambda\int_0^1 J\int_0^1\int_0^r e^{-c\varsigma}J(r-\varsigma)d\varsigma dr'\Lambda'Q'+Q\Lambda\int_0^1\int_0^r e^{-c\varsigma}d\varsigma dr\int_0^1 J\int_0^1 J'\Lambda'Q'$$
$$\equiv Q\Lambda\Psi_2\Lambda'Q' = O_p(\|\Lambda\|^2),$$

$$T^{-1/2}A_{4,T} = T^{-1}\sum_{t=1}^{T}\bar{r}^t\hat{v}_t \Rightarrow Q\Lambda\int_0^1 e^{-cr}J - Q\Lambda\int_0^1 e^{-cr}\int_0^1 J \equiv Q\Lambda\Psi_4 = O_p(\|\Lambda\|),$$

$$T^{-1/2}A_{5,T} = T^{-2}\sum_{t=1}^{T}\bar{r}^t\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}$$
$$\Rightarrow Q\Lambda\int_0^1 e^{-cr}\int_0^r e^{-c\varsigma}J(r-\varsigma)d\varsigma dr - Q\Lambda\int_0^1 e^{-cr}\int_0^r e^{-c\varsigma}d\varsigma dr\int_0^1 J$$
$$\equiv Q\Lambda\Psi_5 = O_p(\|\Lambda\|).$$

Solving the integrals and collecting terms according to (A.6) yields

$$T^{-1}q\hat{L}L \Rightarrow tr\{\bar{V}_X^{-1/2}Q\Lambda\Psi\Lambda'Q'\bar{V}'^{-1/2}\},$$

where

$$\Psi = c^2\Psi_1 - c(\Psi_2 + \Psi_2') - \frac{2c}{1-e^{-2c}}[\Psi_4 - c\Psi_5]'[\Psi_4 - c\Psi_5]. \qquad (A.14)$$

We next consider the sup-Wald test. The unrestricted residuals are given by (A.9) and we have $\hat{\beta}_1 \Rightarrow (1/\lambda)\Lambda\int_0^\lambda J$ and $\hat{\beta}_2 \Rightarrow (1/(1-\lambda))\Lambda\int_\lambda^1 J$. Hence, after some algebra, we get

$$T^{-1}SSR(\lambda)_T \Rightarrow \sigma_e^2 + \int_0^1 J'\Lambda'Q\Lambda J - \frac{1}{\lambda}\int_0^\lambda J'\Lambda'Q\Lambda\int_0^\lambda J - \frac{1}{1-\lambda}\int_\lambda^1 J'\Lambda'Q\Lambda\int_\lambda^1 J$$
$$\equiv SSR^*(\lambda) = O_p(\|\Lambda\|^2).$$

For the restricted regression,

$$T^{-1}SSR_T^r \Rightarrow \sigma_e^2 + \int_0^1 J'\Lambda'Q\Lambda J - \int_0^1 J'\Lambda'Q\Lambda\int_0^1 J \equiv SSR^{r*} = O_p(\|\Lambda\|^2).$$

Therefore by the continuous mapping theorem,

$$T^{-1}SW = \left(\frac{T^{-1}SSR_T^r - T^{-1}SSR_T(\bar{\lambda}_T)}{T^{-1}SSR_T(\bar{\lambda}_T)} \right) \Rightarrow \left(\frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})} \right),$$

where $\bar{\lambda}_T$ is the value of λ that minimizes $SSR_T(\lambda)$ and $\bar{\lambda}$ is the value that minimizes $SSR^*(\bar{\lambda})$.

Proof of Theorem 5 (infrequent break parameter; correction for serial correlation): We consider now the robust version of the test. For the $q\hat{L}L$ test, the limit of $\hat{\Gamma}_j$ is given by

$$\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^{T} X_t X_t' \beta_t \beta_{t-j}' X_{t-j} X_{t-j}' - T^{-1} \sum_{t=j+1}^{T} X_t X_t' \beta_t (\sum_{s=1}^{T} \beta_s') X_{t-j} X_{t-j}' + o_p(1)$$

$$\Rightarrow M_{jj}\Lambda \left[\int_0^1 JJ' - \int_0^1 J \int_0^1 J' \right] \Lambda' M_{jj}' \equiv \bar{\Gamma}_j = O_p(\|\Lambda\|^2)$$

uniformly in $j = o(T)$. To derive an expression for the bandwidth, we again consider for simplicity the scalar case. With variables defined in the proof for the random walk case, we have

$$\hat{\rho} \Rightarrow \frac{M_{11}^2 \Lambda^2 [\int_0^1 J^2 - (\int_0^1 J)^2]}{\sigma_e^2 Q + M^2 \Lambda^2 [\int_0^1 J^2 - (\int_0^1 J)^2]}.$$

The term $\sigma_e^2 Q$ is dominated as $\|\Lambda\|$ increases, and if $\{M - M_{11}\}_{ii} = 0$ for all i, $\hat{\rho}$ converges to 1. Therefore, $\alpha = O_p(\|\Lambda\|^4)$ and $m = O_p(\|\Lambda\|^{4/3} T^{1/3})$. This yields $\hat{h}(0) = O_p(\|\Lambda\|^{10/3} T^{1/3})$. Since $\hat{V}_X = O_p(\|\Lambda\|^2)$, the result follows. For the sup-Wald test, we construct $\hat{h}(0)$ using the unrestricted residuals and we have,

$$\hat{\rho} \Rightarrow \frac{M_{11}^2 \Lambda^2 [\int_0^1 J^2 - \frac{1}{\lambda}(\int_0^\lambda J)^2 - \frac{1}{1-\lambda}(\int_\lambda^1 J)^2]}{\sigma_e^2 Q + M^2 \Lambda^2 [\int_0^1 J^2 - \frac{1}{\lambda}(\int_0^\lambda J)^2 - \frac{1}{1-\lambda}(\int_\lambda^1 J)^2]}.$$

Using similar arguments, $\hat{h}(0) = O_p(\|\Lambda\|^{10/3} T^{1/3})$. If the $\{M - M_{11}\}_{ii} \neq 0$ for some i, then $\alpha = O_p(1)$ and $m = O_p(T^{1/3})$. This yields $\hat{h}(0) = O_p(\|\Lambda\|^2 T^{1/3})$ for both the $q\hat{L}L$ and sup-Wald tests. With these orders for $\hat{h}(0)$, the rest of the proof is the same as for the random walk case.

Proof of Theorem 6 (infrequent break parameter; lagged dependent variable): We now consider the dynamic regression. As in the static case, the OLS residuals $\hat{e}_t$ and $\hat{v}_t$ under the null are given by (A.7) and (A.8) and the limit of the OLS estimates $\hat{\alpha}$ and $\hat{\beta}$ are given by

$$\hat{\alpha} \Rightarrow \{\sigma_e^2 + \int_0^1 J'\Lambda'Q\Lambda J - \int_0^1 J'\Lambda'Q_1'Q^{-1}Q_1\Lambda \int_0^1 J\}^{-1}\{\int_0^1 J'\Lambda'Q_1\Lambda J - \int_0^1 J'\Lambda'Q_1\Lambda \int_0^1 J\}$$

$$\equiv \bar{\alpha} \tag{A.15}$$

and $\hat{\beta} \Rightarrow Q^{-1}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 J$. Consider first the $q\hat{L}L$ test. The components of (A.6) will have the following limits (again, $T^{-1}A_{3,T} \xrightarrow{p} 0_{k \times k}$):

$$
\begin{aligned}
\hat{V}_X &= T^{-1} \sum_{t=1}^{T} \hat{v}_t \hat{v}_t' \\
&\Rightarrow (1 + \bar{\alpha}^2)\sigma_e^2 Q + M\Lambda[\int_0^1 JJ' - \int_0^1 J \int_0^1 J']\Lambda' M' + \bar{\alpha}^2 M_{11}\Lambda[\int_0^1 JJ']\Lambda' M_{11}' \\
&\equiv \bar{V}_X = O_p(\|\Lambda\|^2),
\end{aligned}
$$

$$
\begin{aligned}
T^{-1}A_{1,T} &= T^{-3} \sum_{t=1}^{T} [\sum_{s=1}^{t-1} \bar{r}^{s-1}\hat{v}_{t-s}][\sum_{s=1}^{t-1} \bar{r}^{s-1}\hat{v}_{t-s}]' \\
&\Rightarrow [Q - \alpha Q_1]\Lambda \int_0^1 [\int_0^r e^{-c\varsigma} J(r - \varsigma)d\varsigma][\int_0^r e^{-c\varsigma} J(r - \varsigma)d\varsigma]'dr\Lambda'[Q - \alpha Q_1]' \\
&\quad +[Q - \alpha Q_1]\Lambda \int_0^1 \int_0^r e^{-c\varsigma}d\varsigma^2 dr \int_0^1 J \int_0^1 J'\Lambda'[Q - \alpha Q_1]' \\
&\quad -[Q - \alpha Q_1]\Lambda \int_0^1 \int_0^r e^{-c\varsigma}d\varsigma \int_0^r e^{-c\varsigma} J(r - \varsigma)d\varsigma dr \int_0^1 J'\Lambda'[Q - \alpha Q_1]' \\
&\quad -[Q - \alpha Q_1]\Lambda \int_0^1 J \int_0^1 \int_0^r e^{-c\varsigma}d\varsigma \int_0^r e^{-c\varsigma} J(r - \varsigma)d\varsigma dr'\Lambda'[Q - \alpha Q_1]' = O_p(\|\Lambda\|^2),
\end{aligned}
$$

$$
\begin{aligned}
T^{-1}A_{2,T} &= T^{-2} \sum_{t=1}^{T} \hat{v}_t [\sum_{s=1}^{t-1} \bar{r}^{s-1}\hat{v}_{t-s}]' \\
&\Rightarrow [Q - \alpha Q_1]\Lambda \int_0^1 J(r) \int_0^r e^{-c\varsigma} J(r - \varsigma)d\varsigma dr\Lambda[Q - \alpha Q_1]' \\
&\quad -[Q - \alpha Q_1]\Lambda \int_0^1 \int_0^r e^{-c\varsigma}d\varsigma J(r)dr \int_0^1 J'\Lambda'[Q - \alpha Q_1]' \\
&\quad -[Q - \alpha Q_1]\Lambda \int_0^1 J \int_0^1 \int_0^r e^{-c\varsigma} J(r - \varsigma)d\varsigma dr'\Lambda'[Q - \alpha Q_1]' \\
&\quad +[Q - \alpha Q_1]Q\Lambda \int_0^1 \int_0^r e^{-c\varsigma}d\varsigma dr \int_0^1 J \int_0^1 J'\Lambda'[Q - \alpha Q_1]' = O_p(\|\Lambda\|^2),
\end{aligned}
$$

$$
T^{-1/2}A_{4,T} = T^{-1} \sum_{t=1}^{T} \bar{r}^t \hat{v}_t \Rightarrow [Q - \alpha Q_1]\Lambda \int_0^1 e^{-cr} J - [Q - \alpha Q_1]\Lambda \int_0^1 e^{-cr} \int_0^1 J = O_p(\|\Lambda\|),
$$

$$
\begin{aligned}
T^{-1/2}A_{5,T} &= T^{-2} \sum_{t=1}^{T} \bar{r}^t \sum_{s=1}^{t-1} \bar{r}^{s-1}\hat{v}_{t-s} \\
&\Rightarrow [Q - \alpha Q_1]\Lambda \int_0^1 e^{-cr} \int_0^r e^{-c\varsigma} J(r - \varsigma)d\varsigma dr - [Q - \alpha Q_1]\Lambda \int_0^1 e^{-cr} \int_0^r e^{-c\varsigma}d\varsigma dr \int_0^1 J \\
&= O_p(\|\Lambda\|).
\end{aligned}
$$

Collecting terms yields the stated results for the $q\hat{L}L$ test. For the sup-Wald test, the restricted sum of squared residuals is such that

$$
\begin{aligned}
T^{-1}SSR_T^r &= T^{-1} \sum_{t=1}^{T} \left[e_t - \hat{\alpha}e_{t-1} + X_t'\beta_t - \hat{\alpha}X_{t-1}'\beta_{t-1} - X_t\hat{\beta}\right]^2 \\
&\Rightarrow (1 + \bar{\alpha}^2)\sigma_e^2 + (1 + \bar{\alpha}^2) \int_0^1 J'\Lambda'Q\Lambda J - 2\bar{\alpha} \int_0^1 J'\Lambda'Q_1\Lambda J \\
&\quad - \int_0^1 J'\Lambda'[Q - \bar{\alpha}Q_1]'Q^{-1}[Q - \bar{\alpha}Q_1]\Lambda \int_0^1 J \equiv SSR^{r*} = O_p(\|\Lambda\|^2),
\end{aligned}
$$

where $\bar{\alpha}$ is given by (A.15). For the unrestricted regression the residuals are

$$\hat{e}(\lambda)_t = e_t + X_t'\beta_t - \hat{\alpha}_u e_{t-1} - \hat{\alpha}_u X_{t-1}'\beta_{t-1} - X_1(\lambda)_t'\hat{\beta}_1 - X_2(\lambda)_t'\hat{\beta}_2,$$

where the estimates of the coefficients have the following limits:

$$\hat{\alpha}_u \Rightarrow \{\sigma_e^2 + \int_0^1 J'\Lambda'Q\Lambda J - \frac{1}{\lambda}\int_0^\lambda J'\Lambda Q_1'Q^{-1}Q_1\Lambda\int_0^\lambda J - \frac{1}{1-\lambda}\int_\lambda^1 J'\Lambda Q_1'Q^{-1}Q_1\Lambda\int_\lambda^1 J\}^{-1}$$

$$\times\{\int_0^1 J'\Lambda'Q_1\Lambda J - \frac{1}{\lambda}\int_0^\lambda J'\Lambda'Q_1\Lambda\int_0^\lambda J - \frac{1}{1-\lambda}\left[\int_\lambda^1 J\right]'\Lambda'Q_1\Lambda\left[\int_\lambda^1 J\right]\} \equiv \bar{\alpha}_u, \text{ (A.16)}$$

$T^{-1/2}\hat{\beta}_1 \Rightarrow (1/\lambda)\Lambda Q^{-1}[Q - \bar{\alpha}_u Q_1]\int_0^\lambda J$ and $T^{-1/2}\hat{\beta}_2 \Rightarrow (1/(1-\lambda))\Lambda Q^{-1}[Q - \bar{\alpha}_u Q_1]\int_\lambda^1 J$. Therefore, the limit of $SSR(\lambda)_T$ is given by

$$\begin{aligned}
T^{-1}SSR(\lambda)_T &= T^{-2}\sum_{t=1}^T\left[e_t + X_t'\beta_t - \hat{\alpha}_u e_{t-1} - \hat{\alpha}_u X_{t-1}'\beta_{t-1} - X_1(\lambda)_t'\hat{\beta}_1 - X_2(\lambda)_t'\hat{\beta}_2\right]^2 \\
&\Rightarrow (1+\bar{\alpha}_u^2)\sigma_e^2 + (1+\bar{\alpha}_u^2)\int_0^1 J'\Lambda'Q\Lambda J - 2\bar{\alpha}_u\int_0^1 J'\Lambda'Q_1\Lambda J \\
&\quad -\frac{1}{\lambda}\int_0^\lambda J'\Lambda'[Q - \bar{\alpha}_u Q_1]'Q^{-1}[Q - \bar{\alpha}_u Q_1]\Lambda\int_0^\lambda J \\
&\quad -\frac{1}{1-\lambda}\int_\lambda^1 J'\Lambda'[Q - \bar{\alpha}_u Q_1]'Q^{-1}[Q - \bar{\alpha}_u Q_1]\Lambda\int_\lambda^1 J \equiv SSR(\lambda)^* = O_p(\|\Lambda\|^2),
\end{aligned}$$

where $\bar{\alpha}_u$ is given by (A.16). By the continuous mapping theorem, the limit of the sup-Wald test is given by

$$T^{-1}SW = \left(\frac{T^{-1}SSR_T^r - T^{-1}SSR_T(\bar{\lambda}_T)}{T^{-1}SSR_T(\bar{\lambda}_T)}\right) \Rightarrow \left(\frac{SSR^{r*} - SSR^*(\bar{\lambda})}{SSR^*(\bar{\lambda})}\right) = O_p(1),$$

where $\bar{\lambda}_T$ is the value of λ that minimizes $SSR_T(\lambda)$ and $\bar{\lambda}$ minimizes $SSR^*(\lambda)$.

Proof of Theorem 7 (single break parameter; *i.i.d.* errors assumed): The OLS estimate of β now has the following limit:

$$\hat{\beta} = [T^{-1}\sum_{t=1}^T X_t X_t']^{-1}[T^{-1}\sum_{t=1}^T X_t X_t'\beta_t + T^{-1}\sum_{t=1}^T X_t e_t] \xrightarrow{p} (1-\lambda_c)\theta$$

and $\hat{e}_t$ and $\hat{v}_t$ are still given by (A.7) and (A.8). We now consider the limits of $\hat{V}_X$ and $A_{i,T}$, $i = 1,...,5$ (again, $T^{-1}A_{3,T} \xrightarrow{p} 0_{k\times k}$).

$$\hat{V}_X = T^{-1}\sum_{t=1}^T \hat{v}_t\hat{v}_t' \xrightarrow{p} \sigma_e^2 Q + \lambda_c(1-\lambda_c)M\theta\theta'M' \equiv \bar{V}_X = O_p(\|\theta\|^2),$$

$$\begin{aligned}
T^{-1}A_{1,T} &= T^{-3}\sum_{t=1}^T[\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}][\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}]' \\
&\xrightarrow{p} \int_{\lambda_c}^1\left(\int_0^{r-\lambda_c}e^{-c\varsigma}d\varsigma\right)^2 drQ\theta\theta'Q' + (1-\lambda_c)^2\int_0^1\left(\int_0^r e^{-c\varsigma}d\varsigma\right)^2 drQ\theta\theta'Q' \\
&\quad -2(1-\lambda_c)\int_{\lambda_c}^1\left(\int_0^{r-\lambda_c}e^{-c\varsigma}d\varsigma\right)\left(\int_0^r e^{-c\varsigma}d\varsigma\right)drQ\theta\theta'Q' = O_p(\|\theta\|^2),
\end{aligned}$$

43

$$T^{-1}A_{2,T} = T^{-2}\sum_{t=1}^{T}\hat{v}_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}]'$$

$$\xrightarrow{p} \int_{\lambda_c}^{1}\left(\int_0^{r-\lambda_c}e^{-c\varsigma}d\varsigma\right)drQ\theta\theta'Q' + (1-\lambda_c)^2\int_0^1\int_0^r e^{-c\varsigma}d\varsigma dr Q\theta\theta'Q'$$

$$-(1-\lambda_c)\int_{\lambda_c}^1\left(\int_0^r e^{-c\varsigma}d\varsigma\right)drQ\theta\theta'Q' - (1-\lambda_c)\int_{\lambda_c}^1\int_0^{r-\lambda_c}e^{-c\varsigma}d\varsigma dr Q\theta\theta'Q'$$

$$= O_p(\|\theta\|^2),$$

$$T^{-1/2}A_{4,T} = T^{-1}\sum_{t=1}^{T}\bar{r}^t\hat{v}_t \xrightarrow{p} \int_{\lambda_c}^1 e^{-cr}drQ\theta - (1-\lambda_c)\int_0^1 e^{-cr}drQ\theta = O_p(\|\theta\|),$$

$$T^{-1/2}A_{5,T} = T^{-2}\sum_{t=1}^{T}\bar{r}^t\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}$$

$$\xrightarrow{p} \int_{\lambda_c}^1 e^{-cr}\int_0^{r-\lambda_c}e^{-c\varsigma}d\varsigma dr Q\theta - (1-\lambda_c)\int_0^1 e^{-cr}\int_0^r e^{-c\varsigma}d\varsigma dr Q\theta = O_p(\|\theta\|).$$

Collecting terms and using the representation (A.6) yields the following limit for the $q\hat{L}L$ statistic:

$$T^{-1}q\hat{L}L \xrightarrow{p} \xi tr\{\bar{V}_X^{-1/2}Q\theta\theta'Q'\bar{V}'^{-1/2}\},$$

where

$$\xi = \lambda_c(\lambda_c-1) + \frac{1}{c}(1-\lambda_c+\frac{1}{2}\lambda_c^2) - \frac{1}{c}(1-\lambda_c)^2e^{-c} - \frac{1}{2c}(1-\lambda_c)^2e^{-2c} - \frac{1}{c}(1-\lambda_c)e^{-c\lambda_c}$$

$$+\frac{1}{c}(1-\lambda_c)e^{c(\lambda_c-2)} - \frac{1}{2c}e^{2c(\lambda_c-1)} - \frac{1}{2c(1-e^{-2c})}[\lambda_c-1+(1-\lambda_c)e^{-2c}+e^{-c\lambda_c}-e^{c(\lambda_c-2)}]^2.$$

For the sup-Wald test, the sum of squared restricted residuals is such that

$$T^{-1}SSR_T^r = T^{-1}\sum_{t=1}^{T}\left[e_t + X_t'\theta_t - X_t'\hat{\beta}\right]^2 \xrightarrow{p} \sigma_e^2 + \lambda_c(1-\lambda_c)\theta'Q\theta$$

and we can show that $T^{-1}SSR_T(\lambda) \xrightarrow{p} \sigma_e^2$. Hence, by the continuous mapping theorem

$$T^{-1}SW = \left(\frac{T^{-1}SSR_T^r - T^{-1}SSR_T(\bar{\lambda}_T)}{T^{-1}SSR_T(\bar{\lambda}_T)}\right) \xrightarrow{p} \frac{\lambda_c(1-\lambda_c)}{\sigma_e^2}\theta'Q\theta,$$

where $\bar{\lambda}_T$ minimizes $SSR_T(\lambda)$.

Proof of Theorem 8 (single break parameter; *i.i.d.* correction for serial correlation): Consider now the robust version of the tests. The $q\hat{L}L$ test uses the restricted residuals so that the autocovariances $\hat{\Gamma}_j$ are such that

$$\hat{\Gamma}_j = T^{-1}\sum_{t=j+1}^{T}X_te_te_{t-j}X_{t-j}' + T^{-1}\sum_{t=j+1}^{T}X_tX_t'\beta_t\beta_{t-j}'X_{t-j}'X_{t-j} + T^{-1}\sum_{t=j+1}^{T}X_tX_t'\hat{\beta}\hat{\beta}'X_{t-j}'X_{t-j}$$

$$-T^{-1}\sum_{t=j+1}^{T}X_tX_t'\beta_t\hat{\beta}'X_{t-j}'X_{t-j} - T^{-1}\sum_{t=j+1}^{T}X_tX_t'\hat{\beta}\beta_{t-j}'X_{t-j}'X_{t-j} + o_p(1)$$

$$\xrightarrow{p} \lambda_c(1-\lambda_c)M_{jj}\theta\theta'M_{jj}' = O_p(\|\theta\|^2).$$

Next we consider the bandwidth m, where again we consider the scalar case for simplicity. First,

$$\hat{\rho} \xrightarrow{p} \frac{\lambda_c(1-\lambda_c)M_1\theta^2}{\sigma_e^2 Q + \lambda_c(1-\lambda_c)M\theta^2}.$$

As θ becomes larger, the term $\sigma_e^2 Q$ is dominated and when $\theta'(M_{11}-M)\theta = 0$ as $\|\theta\| \to \infty$, $\hat{\rho}$ approaches 1. Hence, $\alpha = O_p(\|\theta\|^4)$ and $m = O_p(\|\theta\|^{4/3}T^{1/3})$. This yields $\hat{h}(0) = O_p(\|\theta\|^{10/3}T^{1/3})$. Since $\hat{V}_X = O_p(\|\theta\|^2)$, the result follows. The sup-Wald test uses unrestricted residuals in the construction of $\hat{h}(0)$. Now, it is straightforward to show that when $\lambda = \lambda_c$, $\hat{\Gamma}_j \xrightarrow{p} 0$ for $j \neq 0$, since $E(e_t e_{t-j}) = 0$ if $j \neq 0$. For the bandwidth, since $\hat{\rho} \xrightarrow{p} 0$, $\alpha = O_p(1)$, and $m = O_p(T^{1/3})$, so that $\hat{h}(0) = O_p(T^{1/3})$ and the result follows.

Proof of Theorem 9 (single break parameter; lagged dependent variable): For the model under the null hypothesis, the OLS residuals $\hat{e}_t$ and $\hat{v}_t$ are given by (A.10) and (A.11) and the OLS estimates $\hat{\alpha}$ and $\hat{\beta}$ have the following limits:

$$\hat{\alpha} \xrightarrow{p} [\sigma_e^2 + (1-\lambda_c)\theta'Q\theta - (1-\lambda_c)^2\theta'Q_1 Q^{-1}Q_1\theta]^{-1}[\lambda_c(1-\lambda_c)\theta'Q_1\theta] \equiv \bar{\alpha} \qquad (A.17)$$

and

$$\hat{\beta} \xrightarrow{p} (1-\lambda_c)\theta - \bar{\alpha}(1-\lambda_c)Q^{-1}Q_1\theta.$$

For the $q\hat{L}L$ test, the components of (A.6) have the following limits (again, $T^{-1}A_{3,T} \to 0_{k\times k}$):

$$
\begin{aligned}
\hat{V}_X &= T^{-1}\sum_{t=1}^{T}\hat{v}_t\hat{v}_t' \xrightarrow{p} (1+\bar{\alpha}^2)\sigma_e^2 Q + (1-\lambda_c)\lambda_c M\theta\theta'M' + \bar{\alpha}^2(1-\lambda_c)M_{11}\theta\theta'M_{11}' \\
&\quad -2\bar{\alpha}(1-\lambda_c)\lambda_c M_1\theta\theta'M_1' + \bar{\alpha}^2(1-\lambda_c)^2 MQ^{-1}Q_1\theta\theta'Q_1'Q^{-1'}M' \\
&\quad -\bar{\alpha}^2(1-\lambda_c)^2 M_1[Q^{-1}Q_1\theta\theta' + \theta\theta'Q_1'Q^{-1'}]M_1' \equiv \bar{V}_X = O_p(\|\Lambda\|^2),
\end{aligned}
$$

$$
\begin{aligned}
T^{-1}A_{1,T} &= T^{-3}\sum_{t=1}^{T}[\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}][\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}]' \xrightarrow{p} \int_{\lambda_c}^{1}(\int_0^{r-\lambda_c}e^{-cs}ds)^2 dr[Q-\bar{\alpha}Q_1]\theta\theta'[Q-\bar{\alpha}Q_1]' \\
&\quad +(1-\lambda_c)^2\int_0^1(\int_0^r e^{-cs}ds)^2 dr[Q-\bar{\alpha}Q_1]\theta\theta'[Q-\bar{\alpha}Q_1]' \\
&\quad -2(1-\lambda_c)\int_{\lambda_c}^1\int_0^{r-\lambda_c}e^{-cs}ds\int_0^r e^{-cs}dsdr[Q-\bar{\alpha}Q_1]\theta\theta'[Q-\bar{\alpha}Q_1]' = O_p(\|\theta\|^2),
\end{aligned}
$$

$$
\begin{aligned}
T^{-1}A_{2,T} &= T^{-2}\sum_{t=1}^{T}\hat{v}_t[\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s}]' \xrightarrow{p} \lambda_c\int_{\lambda_c}^1\int_0^{r-\lambda_c}e^{-cs}dsdr[Q-\bar{\alpha}Q_1]\theta\theta'[Q-\bar{\alpha}Q_1]' \\
&\quad -\lambda_c(1-\lambda_c)\int_{\lambda_c}^1\int_0^r e^{-cs}dsdr[Q-\bar{\alpha}Q_1]\theta\theta'[Q-\bar{\alpha}Q_1]' \\
&\quad -\lambda_c(1-\lambda_c)\int_{\lambda_c}^1\int_0^{r-\lambda_c}e^{-cs}dsdr[Q-\bar{\alpha}Q_1]\theta\theta'[Q-\bar{\alpha}Q_1]' = O_p(\|\theta\|^2),
\end{aligned}
$$

$$T^{-1/2}A_{4,T} = T^{-1}\sum_{t=1}^{T}\bar{r}^t\hat{v}_t \xrightarrow{p} \int_{\lambda_c}^1 e^{-cr}dr[Q-\bar{\alpha}Q_1]\theta - (1-\lambda_c)\int_0^1 e^{-cr}dr[Q-\bar{\alpha}Q_1]\theta = O_p(\|\theta\|),$$

45

$$T^{-1/2}A_{5,T} = T^{-2}\sum_{t=1}^{T}\bar{r}^t\sum_{s=1}^{t-1}\bar{r}^{s-1}\hat{v}_{t-s} \xrightarrow{p} \int_{\lambda_c}^{1}e^{-cr}\int_{0}^{r-\lambda_c}e^{-cs}dsdr[Q-\bar{\alpha}Q_1]\theta$$

$$-(1-\lambda_c)\int_{0}^{1}e^{-cr}\int_{0}^{r-\lambda_c}e^{-cs}dsdr[Q-\bar{\alpha}Q_1]\theta = O_p(\|\theta\|).$$

Then solving the integrals and collecting terms using (A.6) gives the result of the $q\hat{L}L$ test. For the sup-Wald test, the restricted sum of squared residuals is such that

$$T^{-1}SSR_T^r = T^{-1}\sum_{t=1}^{T}\hat{e}_t^2 \xrightarrow{p} (1+\bar{\alpha}^2)\sigma_e^2 + (1-\lambda_c)(1+\bar{\alpha}^2)\theta'Q\theta - 2(1-\lambda_c)\bar{\alpha}\theta'Q_1\theta$$

$$-(1-\lambda_c)^2\theta'[Q-\bar{\alpha}Q_1]'Q^{-1}[Q-\bar{\alpha}Q_1]\theta \equiv SSR^{r*},$$

where $\bar{\alpha}$ is given by (A.17). For the unrestricted regression

$$y_t = \alpha_u y_{t-1} + X_1(\lambda)_t'\beta_1 + X_2(\lambda)_t'\beta_2 + e_t$$

we can show that for $\lambda \leq \lambda_c$,

$$\hat{\alpha}_u \xrightarrow{p} [\sigma_e^2 + (1-\lambda_c)\theta'Q\theta - \frac{(1-\lambda_c)^2}{1-\lambda}\theta'Q_1'Q^{-1}Q_1\theta]^{-1}[1-\lambda_c)\theta'Q_1\theta - \frac{(1-\lambda_c)^2}{1-\lambda}\theta'Q_1'\theta] \equiv \bar{\alpha}_u,$$

$$\hat{\beta}_1 \xrightarrow{p} 0_{k\times 1},$$

$$\hat{\beta}_2 \xrightarrow{p} \frac{1-\lambda_c}{1-\lambda}Q^{-1}[Q-\bar{\alpha}_u Q_1]\theta.$$

For $\lambda > \lambda_c$,

$$\hat{\alpha}_u \xrightarrow{p} [\sigma_e^2 + (1-\lambda_c)\theta'Q\theta - \frac{(\lambda-\lambda_c)^2 + \lambda(1-\lambda)}{\lambda}\theta'Q_1'Q^{-1}Q_1\theta]^{-1}$$

$$\times[(1-\lambda_c)\theta'Q_1\theta - \frac{(\lambda-\lambda_c)^2 + \lambda(1-\lambda)}{\lambda}\theta'Q_1'\theta],$$

$$\hat{\beta}_1 \xrightarrow{p} \frac{\lambda-\lambda_c}{\lambda}Q^{-1}[Q-\bar{\alpha}_u Q_1]\theta,$$

$$\hat{\beta}_2 \xrightarrow{p} Q^{-1}[Q-\bar{\alpha}_u Q_1]\theta.$$

More importantly, for $\lambda = \lambda_c$, $\hat{\alpha}_u \xrightarrow{p} 0$, $\hat{\beta}_1 \xrightarrow{p} 0_{k\times 1}$, and $\hat{\beta}_2 \xrightarrow{p} \theta$. Now, $SSR(\lambda)_T$ will take a minimum value at $\lambda = \lambda_c$ so that

$$T^{-1}SSR(\lambda_c)_T = T^{-1}\sum_{t=1}^{T}\hat{e}(\lambda_c)_t^2 = T^{-1}\sum_{t=1}^{T}e_t^2 + o_p(1) \xrightarrow{p} \sigma_e^2.$$

Therefore, by the continuous mapping theorem,

$$T^{-1}SW = \left(\frac{T^{-1}SSR_T^r - T^{-1}SSR_T(\bar{\lambda})}{T^{-1}SSR_T(\bar{\lambda})}\right) \xrightarrow{p} \left(\frac{SSR^{r*} - \sigma_e^2}{\sigma_e^2}\right) = O_p(\|\theta\|^2),$$

where $\bar{\lambda}_T$ minimizes $SSR_T(\lambda)$. For the special case with $X_t = \{1\}$ for all t, it is clear that

$\bar{\alpha} = 1$ and $Q - \bar{\alpha}Q_1 = 0$ so that $p\lim_{T\to\infty}T^{-1}q\hat{L}L = 0$. For the sup-Wald test, the limit of SSR^{r*} reduces to $2\sigma_e^2$, which implies $p\lim_{T\to\infty}T^{-1}SW = 1$.