

Testing jointly for structural changes in the error variance and coefficients of a linear regression model*

Jing Zhou[†]

Pierre Perron[‡]

Boston University

Boston University

September 25, 2007

Abstract

We provide a comprehensive treatment of the problem of testing jointly for structural changes in both the regression coefficients and the variance of the errors in a single equation system involving stationary regressors. Our framework is quite general in that we allow for general mixing-type regressors and the assumptions on the errors are quite mild. Their distribution can be non-Normal and conditional heteroskedasticity is permitted. Extensions to the case with serially correlated errors are also treated. We provide the required tools to address the following testing problems, among others: a) testing for given numbers of changes in regression coefficients and variance of the errors; b) testing for some unknown number of changes within some pre-specified maximum; c) testing for changes in variance (regression coefficients) allowing for a given number of changes in the regression coefficients (variance); d) sequential procedures to estimate the number of changes present. These testing problems are important for practical applications as witnessed by recent interests in macroeconomics and finance where documenting structural changes in the variability of shocks to simple autoregressions or Vector Autoregressive Models has been a concern. Applications to such macroeconomic time series reinforces the prevalence of changes in both their mean and variance and the fact that for most series an important reduction in variance occurred in the 80s. In many cases, however, the so-called “great moderation” can instead be viewed as a “great reversion”.

JEL Classification: C22

Keywords: Change-point; Variance shift; Conditional heteroskedasticity; Likelihood ratio tests; the “Great moderation”.

*Perron acknowledges financial support from the National Science Foundation under Grant SES-0649350. We are grateful to Zhongjun Qu for comments and for pointing out an error in a previous draft.

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (jzhou112@bu.edu).

[‡]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

1 Introduction

Both the statistics and econometrics literature contain a vast amount of work on issues related to structural changes with unknown break dates, most of it specifically designed for the case of a single change (for an extensive review, see Perron, 2006). The problem of multiple structural changes has received more attention recently mostly in the context of a single regression. Bai and Perron (1998, 2003a) provide a comprehensive treatment of various issues: consistency of estimates of the break dates, tests for structural changes, confidence intervals for the break dates, methods to select the number of breaks and efficient algorithms to compute the estimates. Perron and Qu (2004) extend this analysis to the case where arbitrary linear restrictions are imposed on the coefficients of the model. Related contributions include Hawkins (1976) who presents a comprehensive treatment of estimation based on a dynamic programming algorithm. Also, Liu, Wu and Zidek (1997) consider multiple structural changes in the context of a more general threshold model and propose an information criterion for the selection of the number of changes. Bai, Lumsdaine and Stock (1998) consider asymptotically valid inference for the estimate of a single break date in multivariate time series allowing stationary or integrated regressors as well as trends with estimation carried using a quasi maximum likelihood (QML) procedure. Also, Bai (2000) considers the consistency, rate of convergence and limiting distribution of estimated break dates in a segmented stationary VAR model estimated again by QML when the break can occur in the parameters of the conditional mean, the variance of the error term or both. Kejriwal and Perron (2006a,b) provide a comprehensive treatment of issue related to testing and inference with multiple structural changes in a single equation cointegrated model.

With respect to testing for structural change in the variance of the regression error, the results are quite sparse. Qu and Perron (2007a) consider a multivariate system estimated by quasi maximum likelihood which provides methods to estimate models with structural changes in both the regression coefficients and the covariance matrix of the errors. They provide a limit distribution theory for inference about the break dates and also consider testing for multiple structural changes, though, in this case, their analysis is restricted to models with Normally distributed errors and a prior that the breaks in coefficients and in the variance occur at different dates. Horváth (1993) considers a change in the mean and variance (occurring at the same time) of a sequence of i.i.d. random variables with moments corresponding to those of a Normal distribution. Davis, Huang, and Yao (1995) extend the analysis to an autoregressive process under similar conditions.

We build on the work of Qu and Perron (2007a) to provide a comprehensive treatment of the problem of testing jointly for structural changes in both the regression coefficients and the variance of the errors in a single equation system involving stationary regressors, allowing the break dates to be different or overlap. Our framework is quite general in that we allow for general mixing-type regressors and the assumptions on the errors are quite mild. Their distribution can be non-Normal and conditional heteroskedasticity is permitted. Extensions to the case with serially correlated errors are also treated. We provide the required tools to address the following testing problems, among others: a) testing for given numbers of changes in regression coefficients and variance of the errors; b) testing for some unknown number of changes within some pre-specified maximum; c) testing for changes in variance (regression coefficients) allowing for a given number of changes in the regression coefficients (variance); d) sequential procedures to estimate the number of changes present.

These testing problems are important for practical applications as witnessed by recent interests in macroeconomics and finance where documenting structural changes in the variability of shocks to simple autoregressions or Vector Autoregressive Models has been a concern; see, among others, Blanchard and Simon (2001), Herrera and Pesavento (2005), Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Sensier and van Dijk (2004) and Stock and Watson (2002). Given the lack of proper testing procedures, a common approach is to apply standard sup-Wald type tests (e.g., Andrews, 1993, Bai and Perron, 1998) for changes in the mean of the absolute value of the estimated residuals; see, e.g., Herrera and Pesavento (2005) and Stock and Watson (2002). This is a rather ad hoc procedure. For the problem of testing for a change in variance only (imposing no change in the regression coefficients), Deng and Perron (2006) have recently extended the CUSUM of squares test of Brown, Durbin and Evans (1975) allowing very general conditions on the regressors and the errors (as suggested by Inclán and Tiao (1994) for Normally distributed time series). This test is, however, adequate only if no change in coefficient is present. As documented by, e.g., Stock and Watson (2002), it is often the case that changes in both coefficients and variance occur and the break dates need not be the same. A common method is to first test for changes in the regression coefficients and conditioning on the break dates found, then test for changes in variance. This is clearly inappropriate as in the first step the tests suffers for severe size distortions (see Section 2). Also, neglecting changes in regression coefficient when testing for changes in variance induces both size distortions and a loss of power. Hence, what is needed is a joint approach. To do so, our testing procedures are based on quasi likelihood ratio tests constructed using a likelihood function appropriate for identically

and independently distributed Normal errors. We then apply corrections to these likelihood ratio tests such that their limit distributions are free of nuisance parameters in the presence of non-Normal distribution and conditional heteroskedasticity. We also consider extensions that allow for serial correlation as well.

We apply our testing procedures to various macroeconomic time series studied by Stock and Watson (2002). On the one hand, our results reinforce the prevalence of changes in both mean, persistence and variance of the shocks in simple autoregressions. Most series have an important reduction in variance that occurred in the 80s. For many series, however, the evidence points to two breaks in the variance of the shocks with the feature that it increases at the first one and decreases at the second. Hence, the so-called “great moderation” may be qualified as a phenomenon where the high variance level of the 70s to early 80s are over and we are back to the level of (roughly) pre-70s; sometimes this reversion is exact (e.g., inflation), incomplete (e.g., interest rates) or magnified (real variables). Hence, the so-called “great-moderation” may rather be qualified as a “great-reversion”. We also present a number of interesting results pertaining to changes in level and persistence of the series.

The paper is structured as follows. Section 2 provides some motivations which show that commonly used procedures that do not treat the problem of changes in regression coefficient and in variance jointly suffer from important size distortions and power losses. Section 3 presents the class of models considered as well as the testing problems to be addressed. Section 4 presents the quasi-likelihood tests to be used as the basis of the various testing procedures. Section 5 discusses the main assumptions needed on the regressors and errors, derives the relevant limit distributions under the various null hypotheses and proposes corrected versions of the tests that have a limit distribution free of nuisance parameters. Section 5.1 deals with the case of martingale difference errors, Section 5.2 extends the analysis to serially correlated errors, Section 5.3 covers the case with an unknown number of breaks under the alternative hypothesis, Section 5.4 discusses tests for an additional break in either the regression coefficients or the variance. Section 6 provides simulation results to assess the adequacy of the suggested procedures in terms of their finite sample size and power and provides some guidelines for particular options. Section 7 presents a specific to general method to estimate the number of breaks in each of the regression coefficient and the variance. Section 8 provides empirical applications related to various macroeconomic time series for which changes in both the mean and the variance has been a concern. Section 9 provides brief concluding remarks and directions for future research, and a brief appendix contains some technical derivations.

2 Motivation

To motivate the importance of considering jointly the problem of testing for changes in the regression coefficients and the variance of the errors, we start with some simple simulation experiments. The data generating process (DGP) is a simple sequence of *i.i.d.* Normal random variable with mean and variance that can change at a single date. To analyze the effect of ignoring a variance break when testing for a change in the regression coefficients, the null hypothesis is specified by

$$y_t = \mu + e_t \quad (1)$$

where $e_t \sim i.i.d. N(0, 1 + \delta_1 I(t > T^v))$ with $I(\cdot)$, the indicator function. We consider 3 break dates, $T^v = \{[0.25T], [0.5T], [0.75T]\}$ and variance change δ_1 varying between 0 and 10 in steps of 0.05. The sample size is set to $T = 100$ and 5000 replications are used. The test considered is the standard *Sup-LR* test (see Andrews, 1993) for a one-time change in μ occurring at some unknown date. The size of the test is presented in Figure 1. The results show important size distortions unless the break occurs early at $T^v = [0.25T]$, and these are increasing with δ_1 . To assess the effect on power, the DGP is

$$y_t = \mu + \delta_2 I(t > T^c) + e_t \quad (2)$$

with e_t as specified above. We consider $T^v = \{[0.5T], [0.75T]\}$; $T^c = \{[0.3T]\}$, $T = 100$, $\delta_1 = \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$ and δ_2 varies between 0 and 2. The results are presented in Figure 2, which shows that power decreases as the magnitude of the unaccounted break in variance increases.

We now consider the effect of a change in mean on the size and power of tests for a change in variance that do not take into account the former change. We consider two testing procedures. One is based on the CUSUM of squares tests as originally proposed by Brown, Durbin and Evans (1975) and advocated as a test for a change in variance by Inclán and Tiao (1994), who showed that it is related to the likelihood ratio test for a change in variance in a sequence of *i.i.d.* Normal random variables (though the equivalence is not exact in finite samples). It is defined by

$$CUSQ = \max_{k+1 \leq r \leq T} \sqrt{T} \left| S_T^{(r)} - \frac{r-k}{T-k} \right|$$

where $S_T^{(r)} = (\sum_{t=k+1}^r \tilde{v}_t^2) / (\sum_{t=k+1}^T \tilde{v}_t^2)$, with \tilde{v}_t the recursive residuals. Its limit distribution under the null hypothesis is the supremum (over $[0, 1]$) of a Brownian Bridge process, for the

DGP considered here. To analyze the size of the test, DGP (2) with $\delta_1 = 0$ is used and we set $T^c = \{[.25T], [.5T], [.75T]\}$ with δ_2 varying between 0 and 10. The results are presented in Figure 3 which show that in all cases the size of the test increases to one rapidly as the magnitude of the change in mean increases. This is not surprising in view of the fact that the CUSQ test has power against a change in the regression coefficients as originally argued by Brown, Durbin and Evans (1975). For power, the DGP used is again (2) with δ_1 varying between 0 and 15 and $\delta_2 = \{0, 1, 1.5, 2, 2.5, 3, 3.5\}$. The results are presented in Figure 4, which show that a change in mean that is unaccounted for can increase the power of the CUSQ test. This results is, however, of little help given the large size distortions. Finally, we consider the two steps method used by Herrera and Pesavento (2005) and Stock and Watson (2002), among others, which applies a test for a change in the mean of the absolute value of the estimated residuals. Again, DGP (2) is used to assess the size ($\delta_1 = 0$) and power properties. For size, δ_2 varies between 0 and 10 and we set $T^c = \{[.25T], [.5T], [.75T]\}$, while for power δ_2 varies between 0 and 3.5 and we consider two sets of break dates, namely $\{T^c = [.5T], T^v = [.3T]\}$ and $\{T^c = [.75T], T^v = [.3T]\}$. The results are presented in Figures 5 and 6. They show that unless the unaccounted for change in mean is at mid-sample, the test suffers from serious size distortions, which increase as the magnitude of the change in mean increases. For the case of a break in mean at mid-sample, which suffers from no size distortions, Figure 6 shows that power decreases as the magnitude of the coefficient break increases.

While the setup considered is quite simple, it shows how inference can be misleading when changes in the coefficients of the conditional mean and changes in the variance of the errors are not analyzed jointly. The rest of the paper provides the necessary tools to do so.

3 Model and testing problems

We start with a description of the most general specification of the model considered where multiple breaks occur in both the coefficients of the conditional mean and the variance of the errors, at possibly different times. This will also allow us to set up the notation used throughout the paper.

The main framework of analysis can be described by the following multiple linear regression with m breaks (or $m + 1$ regimes) in the conditional mean equation:

$$y_t = x_t' \beta + z_t' \delta_j + u_t, \quad t = T_{j-1}^c + 1, \dots, T_j^c, \quad (3)$$

for $j = 1, \dots, m + 1$. In this model, y_t is the observed dependent variable at time t ; both

x_t ($p \times 1$) and z_t ($q \times 1$) are vectors of covariates and β and δ_j ($j = 1, \dots, m + 1$) are the corresponding vectors of coefficients; u_t is the disturbance at time t . The indices (T_1^c, \dots, T_m^c) , or the break points, are explicitly treated as unknown (the convention that $T_0^c = 0$ and $T_{m+1}^c = T$ is used). This is a partial structural change model since the parameter vector β is not subject to shifts and is estimated using the entire sample. When $p = 0$, we obtain a pure structural change model where all the model's coefficients are subject to change. Note that using a partial structural change model where only some coefficients are allowed to change can be beneficial both in terms of obtaining more precise estimates and more powerful tests. We also allow for n breaks (or $n + 1$ regimes) for the variance of the errors occurring at unknown dates (T_1^v, \dots, T_n^v) . Accordingly, the error term u_t has zero mean and variance σ_i^2 for $T_{i-1}^v + 1 \leq t \leq T_i^v$ ($i = 1, \dots, n + 1$), where again we use the convention that $T_0^v = 0$ and $T_{n+1}^v = T$. We allow the breaks in the variance and in the regression coefficients to happen at different times, hence the m -vector (T_1^c, \dots, T_m^c) and the n -vector (T_1^v, \dots, T_n^v) can have all distinct elements or they can overlap partly or completely. We let K denote the total number of break dates and $\max[m, n] \leq K \leq m + n$. When the breaks overlap completely, $m = n = K$.

The multiple linear regression system (3) may be expressed in matrix form as

$$Y = X\beta + \bar{Z}\delta + U,$$

where $Y = (y_1, \dots, y_T)'$, $X = (x_1, \dots, x_T)'$, $U = (u_1, \dots, u_T)'$, $\delta = (\delta'_1, \delta'_2, \dots, \delta'_{m+1})'$, and \bar{Z} is the matrix which diagonally partitions Z at (T_1^c, \dots, T_m^c) , i.e., $\bar{Z} = \text{diag}(Z_1, \dots, Z_{m+1})$ with $Z_i = (z_{T_{i-1}^c+1}, \dots, z_{T_i^c})'$. We denote the true value of a parameter with a 0 superscript. In particular, $\delta^0 = (\delta_1^0, \dots, \delta_{m+1}^0)'$ and $(T_1^{c0}, \dots, T_m^{c0})$ are used to denote, respectively, the true values of the parameters δ and the true break dates in the regression coefficients. The matrix \bar{Z}^{c0} is the one which diagonally partitions Z at $(T_1^{c0}, \dots, T_m^{c0})$. Hence, in its most general form, the data-generating process is

$$Y = X\beta^0 + \bar{Z}^0\delta^0 + U \tag{4}$$

with $E(UU') = \Omega^0$, where the diagonal elements of Ω^0 are σ_{i0}^2 for $T_{i-1}^{v0} + 1 \leq t \leq T_i^{v0}$ ($i = 1, \dots, n + 1$). We also consider cases with serial correlation in the errors for which the off-diagonal elements of Ω^0 need not be 0.

This model is a special case of the class of models considered by Qu and Perron (2007a). The method of estimation considered is quasi maximum likelihood (QML) assuming serially uncorrelated Gaussian errors. They prove consistency of the estimates of the break fractions $(\lambda_1^0, \dots, \lambda_K^0) \equiv (T_1^0/T, \dots, T_K^0/T)$, where T_i^0 ($i = 1, \dots, K$) denotes the union of the elements

of $(T_1^{c0}, \dots, T_m^{c0})$ and $(T_1^{v0}, \dots, T_n^{v0})$. This is done under general conditions on the regressors and the errors. Substantial heterogeneity in the distributions of the regressors is allowed across regimes, though unit root processes are not permitted. The series $z_t u_t$ and u_t are assumed to be short memory processes having bounded fourth moments. Otherwise, the conditions are mild in the sense that they allow for substantial conditional heteroskedasticity and autocorrelation. They also derive the limit distribution of the estimates of the break dates.

The testing problems to be considered are the following:

- TP-1: $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = 0, n = n_a\}$;
- TP-2: $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$;
- TP-3: $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a\}$;
- TP-4: $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$,

where m_a and n_a are some positive numbers selected a priori. We shall also consider testing problems where the alternatives specify some unknown numbers of breaks, up to some maximum. These are:

- TP-5: $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = 0, 1 \leq n \leq N\}$;
- TP-6: $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, 1 \leq n \leq N\}$;
- TP-7: $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{1 \leq m \leq M, n = n_a\}$;
- TP-8: $H_0 : \{m = n = 0\}$ versus $H_1 : \{1 \leq m \leq M, 1 \leq n \leq N\}$.

We shall also be concerned with the following testing problems:

- TP-9: $\{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a + 1, n = n_a\}$;
- TP-10: $\{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a + 1\}$,

where m_a and n_a non-negative integers. These are useful to assess the adequacy of a model with a particular number of breaks by looking at whether including one more break is warranted. In Section 7, we also consider sequential testing procedures that allow estimating the number of breaks in both the conditional mean regression and the variance of the errors.

4 The quasi-likelihood ratio tests

In this section we consider the likelihood ratio tests obtained assuming Normally distributed and serially uncorrelated errors, for the testing problems TP-1 to TP-4. We derive their limit distributions, which in general, are not free of nuisance parameters. We then propose, in Section 5, modifications whose asymptotic distributions are free of nuisance parameters. Results for the testing problems TP-5 to TP-8 follow as straightforward corollaries and are discussed in Section 5.3.

Consider TP-1 where one specifies no change in the regression coefficients ($m = q = 0$) but tests for a given number n_a of changes in the variance of the errors. Under the null hypothesis, the log-likelihood function is given by

$$\log \tilde{L}_T = -\frac{T}{2} (\log 2\pi + 1) - \frac{T}{2} \log \tilde{\sigma}^2 \quad (5)$$

where

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{1}{T} \sum_{t=1}^T (y_t - x_t' \tilde{\beta})^2 \\ \tilde{\beta} &= \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t y_t \right) \end{aligned}$$

Under the alternative hypothesis, we estimate the model using the Quasi-Maximum likelihood estimation method (QMLE). For a given partition $\{T_1^v, \dots, T_n^v\}$, the log-likelihood value is given by

$$\log \hat{L}_T(T_1^v, \dots, T_n^v) = -\frac{T}{2} (\log 2\pi + 1) - \sum_{i=1}^{n_a+1} \frac{T_i^v - T_{i-1}^v}{2} \log \hat{\sigma}_i^2, \quad (6)$$

where the QMLE jointly solve the system

$$\begin{aligned} \hat{\beta} &= \left(\sum_{i=1}^{n_a+1} \sum_{t=T_{i-1}^v+1}^{T_i^v} \frac{x_t x_t'}{\hat{\sigma}_i^2} \right)^{-1} \left(\sum_{i=1}^{n_a+1} \sum_{t=T_{i-1}^v+1}^{T_i^v} \frac{x_t y_t}{\hat{\sigma}_i^2} \right) \\ \hat{\sigma}_i^2 &= \frac{1}{T_i^v - T_{i-1}^v} \sum_{t=T_{i-1}^v+1}^{T_i^v} (y_t - x_t' \hat{\beta})^2 \end{aligned}$$

for $i = 1, \dots, n_a + 1$. Hence, the Sup-Likelihood ratio test considered is

$$\begin{aligned} \sup LR_{1,T}(n_a, \varepsilon | m = n = 0) &= \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}} 2 \left[\log \hat{L}_T(T_1^v, \dots, T_{n_a}^v) - \log \tilde{L}_T \right] \\ &= 2 \left[\log \hat{L}_T(\hat{T}_1^v, \dots, \hat{T}_{n_a}^v) - \log \tilde{L}_T \right] \end{aligned}$$

where the estimates $(\hat{T}_1^v, \dots, \hat{T}_{n_a}^v)$ are the QMLE obtained by imposing the restriction that there is no structural change in the coefficients and

$$\Lambda_{v,\varepsilon} = \{(\lambda_1^v, \dots, \lambda_{n_a}^v); |\lambda_{i+1}^v - \lambda_i^v| \geq \varepsilon (i = 1, \dots, n_a - 1), \lambda_1^v \geq \varepsilon, \lambda_{n_a}^v \leq 1 - \varepsilon\}.$$

The parameter ε acts as a truncation which imposes a minimal length for each segment and will affect the limiting distribution of the test.

For the testing problem TP-2, there are m_a breaks in the regression coefficients under both the null and alternative hypotheses, so that the test pertains to assess whether there are 0 or n_a breaks in variance. For a given partition $\{T_1^c, \dots, T_{m_a}^c\}$, the likelihood function under the null hypothesis is:

$$\log \tilde{L}_T(T_1^c, \dots, T_{m_a}^c) = -\frac{T}{2} (\log 2\pi + 1) - \frac{T}{2} \log \tilde{\sigma}^2$$

where

$$\tilde{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta}_j)^2$$

and

$$\begin{aligned} \tilde{\beta} &= (X' M_{\bar{Z}} X)^{-1} X' M_{\bar{Z}} Y \\ \tilde{\delta}_j &= (Z_j' Z_j)^{-1} Z_j (Y_j - X_j \tilde{\beta}) \end{aligned}$$

where $M_{\bar{Z}} = I - \bar{Z} (\bar{Z}' \bar{Z})^{-1} \bar{Z}'$, $\bar{Z} = \text{diag}(Z_1, \dots, Z_{m_a+1})$, and $Z_j = (z_{T_{j-1}^c+1}, \dots, z_{T_j^c})'$, $Y_j = (y_{T_{j-1}^c+1}, \dots, y_{T_j^c})'$, $X_j = (x_{T_{j-1}^c+1}, \dots, x_{T_j^c})'$ for $T_{j-1}^c < t \leq T_j^c$ ($j = 1, \dots, m_a + 1$). The log-likelihood value under the alternative hypothesis is, for given partitions $\{T_1^c, \dots, T_{m_a}^c\}$ and $\{T_1^v, \dots, T_{n_a}^v\}$,

$$\log \hat{L}_T(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v) = -\frac{T}{2} (\log 2\pi + 1) - \sum_{i=1}^{n_a+1} \frac{T_i^v - T_{i-1}^v}{2} \log \hat{\sigma}_i^2, \quad (7)$$

where the QMLE solves the following equations

$$\hat{\sigma}_i^2 = \frac{1}{T_i^v - T_{i-1}^v} \sum_{t=T_{i-1}^v+1}^{T_i^v} (y_t - x_t' \hat{\beta} - z_t' \hat{\delta}_{t,j})^2$$

for $i = 1, \dots, n_a + 1$, and

$$\hat{\beta} = (X' M_{\bar{Z}_\sigma} X)^{-1} X' M_{\bar{Z}_\sigma} Y$$

where $M_{\bar{Z}_\sigma} = I - \bar{Z}_\sigma (\bar{Z}'_\sigma \bar{Z}_\sigma)^{-1} \bar{Z}'_\sigma$ with $\bar{Z}_\sigma = \text{diag}(Z_1^\sigma, \dots, Z_{m_a+1}^\sigma)$, $Z_j^\sigma = (z_{T_{j-1}^c+1}^\sigma, \dots, z_{T_j^c}^\sigma)'$, and $z_t^\sigma = (z_t/\hat{\sigma}_i)$, for $T_{i-1}^v < t \leq T_i^v$ ($i = 1, \dots, n_a + 1$). Using the same notation,

$$\hat{\delta}_{t,j} = (Z_j^{\sigma'} Z_j^\sigma)^{-1} Z_j^{\sigma'} (Y_j^\sigma - X_j^\sigma \hat{\beta})$$

for $T_{j-1}^c \leq t \leq T_j^c$, where $Y_j^\sigma = (y_{T_{j-1}^c+1}^\sigma, \dots, y_{T_j^c}^\sigma)'$ and $X_j^\sigma = (x_{T_{j-1}^c+1}^\sigma, \dots, x_{T_j^c}^\sigma)'$ with $x_t^\sigma = (x_t/\hat{\sigma}_i)$ and $y_t^\sigma = (y_t/\hat{\sigma}_i)$. Hence, the quasi Sup-likelihood ratio test is

$$\begin{aligned} & \sup LR_{2,T}(m_a, n_a, \varepsilon | n = 0, m_a) \\ &= 2 \left[\sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \log \hat{L}_T(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v) - \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}} \log \tilde{L}_T(T_1^c, \dots, T_{m_a}^c) \right] \\ &= 2 \left[\log \hat{L}_T(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v) - \log \tilde{L}_T(\hat{T}_1^c, \dots, \hat{T}_{m_a}^c) \right] \end{aligned}$$

where

$$\Lambda_{c,\varepsilon} = \{(\lambda_1^c, \dots, \lambda_m^c); |\lambda_{j+1}^c - \lambda_j^c| \geq \varepsilon (j = 1, \dots, m_a - 1), \lambda_1^c \geq \varepsilon, \lambda_{m_a}^c \leq 1 - \varepsilon\}$$

and

$$\begin{aligned} \Lambda_\varepsilon &= \{(\lambda_1^c, \dots, \lambda_m^c, \lambda_1^v, \dots, \lambda_n^v); \text{for } (\lambda_1, \dots, \lambda_K) = (\lambda_1^c, \dots, \lambda_m^c) \cup (\lambda_1^v, \dots, \lambda_n^v) \\ & \quad |\lambda_{j+1} - \lambda_j| \geq \varepsilon (j = 1, \dots, K - 1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1 - \varepsilon\} \end{aligned} \quad (8)$$

Note that we denote the estimates of the break dates in coefficients and variance by a “ \sim ” when these are obtained jointly, as opposed to the estimates which are obtained separately and denoted by a “ $\hat{\cdot}$ ”.

Remark 1 *The set Λ_ε which defines the possible values of the break fractions in coefficients $(\lambda_1^c, \dots, \lambda_m^c)$ and in variance $(\lambda_1^v, \dots, \lambda_n^v)$ allows them to have some (or all) common elements or be completely different. What is important is that each break fraction be separated by a non-zero value ε . This does complicate inference since many cases need to be considered. To illustrate, consider the case with $m_a = n_a = 1$. We can have $K = 1$ in which case it is a one break model with both the coefficients and the variance of the errors changing at the same break date. On the other hand, if $K = 2$, the break date for the change in coefficients is different from that for the change in variance. This leads to two additional possible cases to consider: a) $\lambda_1^c \leq \lambda_1^v - \varepsilon$ (the break in the coefficients occurs before the break in the variance), b) $\lambda_1^c \geq \lambda_1^v + \varepsilon$ (the break in the coefficients occurs after the break in the variance). The maximized likelihood function for these two cases can easily be evaluated using the algorithm*

of Qu and Perron (2007a) since it permits the imposition of restrictions. For example, if $\lambda_1^c \leq \lambda_1^v - \varepsilon$, we have a two break model and the restrictions needed are that the variance of the errors in the first and second regimes is identical, and the regression coefficients are identical in the second and third regimes. Hence, for the case $m_a = n_a = 1$, there are three maximized likelihood values to construct and the test corresponds to the maximal value over these three cases. When m_a or n_a are greater than one, more cases need to be considered, but the principle remains the same.

For the testing problem TP-3, the null hypothesis specifies n_a breaks in variance and zero break in the regression coefficients so that, for a given partition $\{T_1^v, \dots, T_{n_a}^v\}$, the likelihood function is given by

$$\log \tilde{L}_T(T_1^v, \dots, T_{n_a}^v) = -\frac{T}{2} (\log 2\pi + 1) - \sum_{i=1}^{n_a+1} \frac{T_i^v - T_{i-1}^v}{2} \log \tilde{\sigma}_i^2,$$

where

$$\tilde{\sigma}_i^2 = \frac{1}{T_i^v - T_{i-1}^v} \sum_{t=T_{i-1}^v+1}^{T_i^v} (y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta})^2$$

for $i = 1, \dots, n_a + 1$, with

$$(\tilde{\beta}', \tilde{\delta}')' = (W^{\sigma'} W^\sigma)^{-1} W^{\sigma'} Y^\sigma,$$

where $W^\sigma = (w_1^\sigma, \dots, w_T^\sigma)'$ with $w_t^\sigma = (x_t^{\sigma'}, z_t^{\sigma'})'$. Under the alternative hypothesis, there are m_a breaks in the regression coefficients and n_a breaks in variance so that the likelihood function is given by (7). Hence, the Sup-Likelihood ratio test is

$$\begin{aligned} & \sup LR_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a) \\ &= 2 \left[\sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \log \hat{L}_T(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v) - \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}} \log \tilde{L}_T(T_1^v, \dots, T_{n_a}^v) \right] \\ &= 2 \left[\log \hat{L}_T(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v) - \log \tilde{L}_T(\hat{T}_1^v, \dots, \hat{T}_{n_a}^v) \right] \end{aligned}$$

For the testing problem TP-4, the null hypothesis specifies no break under both the null and alternative hypotheses and, hence, the log-likelihood value under the null hypothesis is given by $\log \tilde{L}_T$ as specified by (5). The alternative hypothesis specifies m_a breaks in the regression coefficients and n_a breaks in the variance of the errors and the log likelihood value

is given by (7). Hence, the Sup-Likelihood ratio test is then

$$\begin{aligned}
& \sup LR_{4,T}(m_a, n_a, \varepsilon | n = m = 0) \\
&= 2 \left[\sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \log \hat{L}_T(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v) - \log \tilde{L}_T \right] \\
&= 2 \left[\log \hat{L}_T(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v) - \log \tilde{L}_T \right] \tag{9}
\end{aligned}$$

5 The limiting distributions of the tests

We now consider the limit distribution of the tests. We start with the case where the errors are martingale differences in Section 5.1 and consider extensions to serially correlated errors in Section 5.2.

5.1 The case with martingale difference errors.

Since some testing problems imply a given non-zero number of breaks under the null hypothesis, we need some conditions to ensure that the estimates of the break fractions are consistent at a fast enough rate and that the estimates of the parameters are also consistent. This problem was analyzed in Qu and Perron (2007a) and we simply use the same set of assumptions. If breaks are allowed in the regression coefficients under both the null and alternative hypotheses, we specify the following conditions:

- Assumption A1: The conditions stated in Assumptions A1-A4 and A6-A8 of Qu and Perron (2007a) are assumed to hold.

When the null hypothesis specifies no change in the regression coefficients, we shall assume, with $w_t = (x_t', z_t)'$:

- Assumption A2: $T^{-1} \sum_{t=1}^{\lfloor Ts \rfloor} w_t w_t' \rightarrow_p sQ$, uniformly in $s \in [0, 1]$, with Q some positive definite matrix.

Assumption A2 rules out trending regressors and imposes the requirement that the limit moment matrix of the regressors be homogeneous throughout the sample. Hence, we avoid the case where the marginal distribution of the regressors may change while the coefficients do not (see, e.g., Hansen, 2000). This follows from our basic premise that regimes are defined by changes in some coefficients. When changes in the variance of the errors are allowed under both the null and alternative hypotheses, we specify:

- Assumption A3: The conditions stated in Assumption A5 of Qu and Perron (2007a) are assumed to hold with the addition that the errors $\{u_t\}$ form an array of martingale differences relative to $\mathcal{F}_t = \sigma\text{-field } \{\dots, z_{t-1}, z_t, \dots, x_{t-1}, x_t, \dots, u_{t-2}, u_{t-1}\}$.

When the null hypothesis imposes no changes in variance, we shall need:

- Assumption A4: The errors $\{u_t\}$ form an array of martingale differences relative to $\mathcal{F}_t = \sigma\text{-field } \{\dots, z_{t-1}, z_t, \dots, x_{t-1}, x_t, \dots, u_{t-2}, u_{t-1}\}$, and, additionally, $E(u_t^2) = \sigma_0^2$ for all t and $T^{-1/2} \sum_{t=1}^{[Ts]} z_t u_t \Rightarrow \sigma Q^{1/2} W_q(s)$, where $W_q(s)$ is a q -vector of independent Wiener processes. Also, $T^{-1/2} \sum_{t=1}^{[Ts]} (u_t^2/\sigma^2 - 1) \Rightarrow \psi W(s)$ where $W(s)$ is a Wiener process independent of $W_q(s)$ and $\psi = \lim_{T \rightarrow \infty} \text{var}(T^{-1/2} \sum_{t=1}^T (u_t^2/\sigma^2 - 1))$.

Assumption A4 rules out instability in the error process and states that a basic functional central limit theorem holds for the weighted partial sums of the errors and their squares. Note that A4 assumes no serial correlation in the errors u_t . This will be relaxed later.

The limiting distributions, under the relevant null hypothesis, of the likelihood ratio tests for the testing problems TP-1 to TP-4 are stated in the following Theorem, where “ \Rightarrow ” denotes weak convergence under the Skorohod topology and $\|\cdot\|$ is the Euclidian norm.

Theorem 1 *Under the relevant null hypothesis, we have, as $T \rightarrow \infty$,*

a) *For TP-1, under A2 and A4:*

$$\sup LR_{1,T}(n_a, \varepsilon | m = n = 0) \Rightarrow \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}} \frac{\psi}{2} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}$$

b) *For TP-2, under A1 and A4:*

$$\begin{aligned} \sup LR_{2,T}(m_a, n_a, \varepsilon | n = 0, m_a) &\Rightarrow \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}^c} \frac{\psi}{2} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \\ &\leq \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}} \frac{\psi}{2} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{aligned}$$

where

$$\begin{aligned} \Lambda_{v,\varepsilon}^c &= \{(\lambda_1^v, \dots, \lambda_{n_a}^v); \text{ for } (\lambda_1, \dots, \lambda_K) = (\lambda_1^{0c}, \dots, \lambda_m^{0c}) \cup (\lambda_1^v, \dots, \lambda_n^v) \\ &\quad |\lambda_{j+1} - \lambda_j| \geq \varepsilon \ (j = 1, \dots, K-1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1 - \varepsilon\} \end{aligned}$$

and

$$\Lambda_{v,\varepsilon} = \{(\lambda_1^v, \dots, \lambda_{n_a}^v); |\lambda_{i+1}^v - \lambda_i^v| \geq \varepsilon \ (i = 1, \dots, n_a - 1), \lambda_1^v \geq \varepsilon, \lambda_{n_a}^v \leq 1 - \varepsilon\}.$$

c) For TP-3, under A1-A3:

$$\begin{aligned} \sup LR_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a) &\Rightarrow \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}^v} \sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ &\leq \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}} \sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \end{aligned} \quad (10)$$

where

$$\begin{aligned} \Lambda_{c,\varepsilon}^v &= \{(\lambda_1^c, \dots, \lambda_m^c); \text{for } (\lambda_1, \dots, \lambda_K) = (\lambda_1^c, \dots, \lambda_m^c) \cup (\lambda_1^{0v}, \dots, \lambda_n^{0v}) \\ &\quad |\lambda_{j+1} - \lambda_j| \geq \varepsilon \ (j = 1, \dots, K - 1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1 - \varepsilon\} \end{aligned}$$

and

$$\Lambda_{c,\varepsilon} = \{(\lambda_1^c, \dots, \lambda_m^c); |\lambda_{j+1}^c - \lambda_j^c| \geq \varepsilon \ (j = 1, \dots, m_a - 1), \lambda_1^c \geq \varepsilon, \lambda_{m_a}^c \leq 1 - \varepsilon\}$$

d) For TP-4, under A2 and A4:

$$\sup LR_{4,T}(m_a, n_a, \varepsilon | n = m = 0) \Rightarrow \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \left[\begin{aligned} &\sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ &+ \frac{\psi}{2} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{aligned} \right]$$

where

$$\begin{aligned} \Lambda_\varepsilon &= \{(\lambda_1^c, \dots, \lambda_m^c, \lambda_1^v, \dots, \lambda_n^v); \text{for } (\lambda_1, \dots, \lambda_K) = (\lambda_1^c, \dots, \lambda_m^c) \cup (\lambda_1^v, \dots, \lambda_n^v) \\ &\quad |\lambda_{j+1} - \lambda_j| \geq \varepsilon \ (j = 1, \dots, K - 1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1 - \varepsilon\} \end{aligned}$$

Remark 2 For the testing problems TP-2 and TP-3, the limit distributions depend on the true unknown value of the relevant break fractions corresponding to the break dates allowed under both the null and alternative hypotheses. The results, however, indicate that these distributions can be bounded by limit random variables which do not depend on such unknown values. This follows since $\Lambda_{v,\varepsilon}^c \subseteq \Lambda_{v,\varepsilon}$ and $\Lambda_{c,\varepsilon}^v \subseteq \Lambda_{c,\varepsilon}$. Hence, a conservative testing procedure is possible. As we shall see, the test is barely conservative if the trimming parameter ε is small, though as ε gets large (e.g. 0.20) the test will be somewhat undersized.

The proof of this Theorem is given in the Appendix. For the testing problem TP-3, the bound is the same as the limit distribution in Bai and Perron (1998). Hence, the critical values provided by Bai and Perron (1998, 2003b) can be used. For the testing problems TP-1 and TP-2, the same limit distribution (for a one parameter change) applies except for the scaling factor $(\psi/2)$. This quantity can nevertheless still be consistently estimated. Consider the following class of estimates:

$$\hat{\psi} = \frac{1}{T} \sum_{j=-(T-1)}^{T-1} \omega(j, m) \sum_{t=|j|+1}^T \hat{\eta}_t \hat{\eta}_{t-j} \quad (11)$$

where $\hat{\eta}_t = (\hat{u}_t^2 / \hat{\sigma}^2) - 1$ where $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$ with \hat{u}_t the residuals under the null hypotheses. Here $w(j, m)$ is a weight function and m some bandwidth which can be selected using one of the many alternative ways that have been proposed; see, e.g., Andrews (1991). The estimate $\hat{\psi}$ will be consistent under some conditions on the choice of $w(j, m)$ and the rate of increase of m as a function of T . Following Kejriwal and Perron (2006a), we use the residuals under the null hypothesis to construct $\hat{\psi}$ but the residuals under the alternative hypothesis to select the bandwidth parameter m (see also Kejriwal, 2007). Simulations showed that using the residuals under the alternative hypothesis to select m and construct $\hat{\psi}$ leads to tests with important size distortions. Using the residuals under the null for both leads to conservative and less powerful tests. Using the hybrid method permits, as we shall see, to control the exact size in small samples without significant loss of power. In our simulations and empirical applications, we use the Quadratic Spectral kernel and to select m we adopt the method suggested by Andrews (1991) with an AR(1) approximation.

Remark 3 *If the errors are i.i.d., $\psi = \mu_4 / \sigma^4 - 1$, which can be consistently estimated using $\hat{\psi} = \hat{\mu}_4 / \hat{\sigma}^4 - 1$, where $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$ and $\hat{\mu}_4 = T^{-1} \sum_{t=1}^T \hat{u}_t^4$ with \hat{u}_t the residuals under the null or alternative hypotheses. Also, if the errors are Normal, $\psi = 2$ so that no adjustment is necessary, a case that was covered by Qu and Perron (2007a). Since these cases are of less relevance in practical applications, we shall only consider a correction involving $\hat{\psi}$ as defined by (11). But it is useful to note that a simpler correction is available if the i.i.d. assumption is reasonable.*

We then have the following corrected statistic with a nuisance parameter free limit distribution:

$$\begin{aligned} \sup LR_{1,T}^* &= (2/\hat{\psi}) \sup LR_{1,T} \Rightarrow \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\epsilon}} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \quad (12) \\ \sup LR_{2,T}^* &= (2/\hat{\psi}) \sup LR_{2,T} \Rightarrow \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\epsilon}} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \\ &\leq \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\epsilon}} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{aligned}$$

For the testing problem TP-4, it is possible to obtain a transformation that has a limit distribution free of nuisance parameters but the procedure is more involved. It is given by

$$\sup LR_{4,T}^* = \sup LR_{4,T} - \frac{\hat{\psi} - 2}{\hat{\psi}} LR_v \quad (13)$$

where LR_v is the likelihood ratio test for no break in variance versus n_a breaks evaluated using the estimates $\{\tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v\}$ obtained by maximizing the likelihood function jointly allowing for m_a breaks in coefficients, i.e.,

$$LR_v = 2 \left[\log \hat{L}_T(\tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v) - \log \tilde{L}_T \right]$$

where $\log \hat{L}_T(\cdot)$ and $\log \tilde{L}_T$ are defined by (6) and (5), respectively. Note that LR_v is not equivalent to $LR_{1,T}(n_a, \epsilon | m = n = 0)$ which is based on the estimates of the break dates for the changes in variance assuming no break in coefficients. Since $\{\tilde{T}_1^v/T, \dots, \tilde{T}_{n_a}^v/T\}$ are consistent estimates of the break fractions whether we have m_a breaks in coefficients or not, we deduce that

$$LR_v \Rightarrow \frac{\psi}{2} \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\epsilon} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}$$

and, hence,

$$\sup LR_{4,T}^* \Rightarrow \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\epsilon} \left[\sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} + \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \right] \equiv \sup LR_4^* \quad (14)$$

The limit distribution (14) is new. To obtain the relevant critical values we proceeded as follows. We first simulate a dependent variable and q regressors as independent $N(0, 1)$

random variables. This is without loss of generality since the limit distribution does not depend on the distribution of the regressors and using Normally distributed series will ensure a closer correspondence with the asymptotic distribution for a given sample size, which we set to $T = 500$. The algorithm of Qu and Perron (2007a) imposing appropriate restrictions is then used to obtain the estimates of the m_a break dates in coefficients and the n_a break dates in variance using the trimming specified by Λ_ϵ . We then simulate a $q \times 1$ vector of independent Wiener processes $W_{q+1}(\cdot)$ as partial sums of independent $N(0, 1)$ random variables, again with $T = 500$, and evaluate the quantity

$$\sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} + \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}$$

where $W_q(\cdot)$ contains the first q elements of the vector $W_{q+1}(\cdot)$ and $W(\cdot)$ is the $q+1$ th element of the simulated vector $W_{q+1}(\cdot)$. This is repeated 2,000 times to obtain the relevant quantiles corresponding to the distribution of the sum of the two terms. The critical values for tests of size 1%, 2.5%, 5% and 10% are presented in Table 1 for q between 1 and 5 and $\epsilon = 0.1, 0.15, 0.20$ and 0.25 . For $\epsilon = 0.1, 0.15, 0.2$, $m_a = 1, 2$ and $n_a = 1, 2$. For $\epsilon = 0.25$, $m_a = 1$, and $n_a = 1$ given that $\epsilon = 0.25$ imposes a maximal number of 2 breaks.

5.2 Extensions to serially correlated errors

We now consider the case where the errors u_t can be serially correlated. To that effect Assumptions A3 and A4 are replaced by:

- Assumption A3*: The conditions stated in Assumption A5 of Qu and Perron (2007a) are assumed to hold.

and when the null hypothesis imposes no changes in variance, we shall need:

- Assumption A4*: $E(u_t^2) = \sigma_0^2$ for all t and $T^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} z_t u_t \Rightarrow \sigma Q^{1/2} W_q(s)$, where $W_q(s)$ is a q -vector of independent Wiener processes. Also, $T^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} (u_t^2 / \sigma^2 - 1) \Rightarrow \psi W(s)$ where $W(s)$ is a Wiener process independent of $W_q(s)$ and

$$\psi = \lim_{T \rightarrow \infty} \text{var}(T^{-1/2} \sum_{t=1}^T (u_t^2 / \sigma^2) - 1).$$

For the testing problems TP-1 and TP-2, the results are the same and the sup $LR_{1,T}^*$ and sup $LR_{2,T}^*$ are statistics that will be asymptotically invariant to non-Normal errors, serial

correlation and conditional heteroskedasticity so that the limit distribution (12) still applies. For the testing problems TP-3 and TP-4, things are more complex. Consider first TP-3. When the errors u_t are serially correlated, the likelihood ratio type tests for changes in the coefficients of the conditional mean depend on nuisance parameters and would be hard to implement in practice. In such a case, structural changes in the regression coefficients can still be tested using the following Wald type statistics taking into account the presence of serial correlation: $\sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_\varepsilon} F_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a)$, where

$$F_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a) = \frac{(T - (m_a + 1)q - p)}{m_a q} \hat{\delta}' R' \left(R \hat{V}(\hat{\delta}) R' \right)^{-1} R \hat{\delta} \quad (15)$$

with $\hat{\delta} = (\delta'_1, \dots, \delta'_{m_a+1})'$ is the quasi-maximum likelihood estimate of the coefficients that are subject to change, under a given partition of the sample, R is the conventional matrix such that $(R\delta)' = (\delta'_1 - \delta'_2, \dots, \delta'_{m_a} - \delta'_{m_a+1})$ and $\hat{V}(\hat{\delta})$ is an estimate of the variance covariance matrix of $\hat{\delta}$ that is robust to serial correlation and heteroskedasticity, i.e, a consistent estimate of $V(\hat{\delta}) = \text{plim}_{T \rightarrow \infty} T (\bar{Z}_\sigma^{*'} \bar{Z}_\sigma^*)^{-1} \Omega_{\bar{Z}_\sigma^*} (\bar{Z}_\sigma^{*'} \bar{Z}_\sigma^*)^{-1}$, where $\bar{Z}_\sigma^* = M_{X_\sigma} \bar{Z}_\sigma$, $\Omega_{\bar{Z}_\sigma^*} = E(\bar{Z}_\sigma^{*'} U_b^* U_b^{*'} \bar{Z}_\sigma^*)$, $U_b^* = M_{X_\sigma} U_\sigma$, $M_{X_\sigma} = I_T - X_\sigma (X_\sigma' X_\sigma)^{-1} X_\sigma'$, with $\bar{Z}_\sigma = \text{diag}(Z_1^\sigma, \dots, Z_{m_a+1}^\sigma)$, $Z_j^\sigma = (z_{T_j^c-1+1}^\sigma, \dots, z_{T_j^c}^\sigma)'$, $U_\sigma = (u_1^\sigma, \dots, u_T^\sigma)'$, $z_t^\sigma = (z_t/\sigma_i)$ and $u_t^\sigma = (u_t/\sigma_i)$, for $T_{i-1}^{v0} < t \leq T_i^{v0}$ ($i = 1, \dots, n_a + 1$). Under A2, A3* and additional assumptions under which a consistent estimate of $V(\hat{\delta})$ can be obtained using kernel based methods as in Andrews (1991), the limiting distribution of $\sup F_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a)$ is the same as in the case with martingale difference errors, i.e, as stated in (10). In practice, the computation of the above tests could be very involved, especially if a data dependent method is used to construct the robust asymptotic covariance, $\hat{V}(\hat{\delta})$. Following Bai and Perron (1998), we suggest first to use the dynamic programming algorithm to get the break points corresponding to the global maximization of the likelihood function defined by (7), then plug the estimates into (15) to construct the test. This will not affect the consistency of the test since the break fractions are consistently estimated.

For the testing problem TP-4, things are more complex. We shall adopt a quasi Wald testing procedure. Note first that the information matrix is block diagonal with respect to δ and σ^2 , hence the test will involve one component for changes in δ and one component for changes in σ^2 . The first is the same as discussed above, namely $\sup F_{3,T}$ as defined by (15), except that one can use z_t instead of z_t^σ since the null hypothesis specifies no break in variance. The difficulty is with the second component. The Wald test for the equality in variance across regime is asymptotically different from the LR test even with martingale difference errors. In fact, its limit distribution is quite complex and would necessitate additional tables of critical

values. A compromise that is simple and yet still leads to a consistent test is to sum the individuals Wald tests for each successive pairs of regimes. This leads to the component:

$$\sup F_T^\sigma = \hat{\psi}^{-1} \sum_{i=1}^{n_a} (\hat{\sigma}_{i+1}^2 - \hat{\sigma}_i^2)^2 \left(\frac{\hat{\sigma}_{i+1}^4}{\tilde{\lambda}_{i+1}^v - \tilde{\lambda}_i^v} - \frac{\hat{\sigma}_i^4}{\tilde{\lambda}_i^v - \tilde{\lambda}_{i-1}^v} \right)^{-1}$$

where $\hat{\sigma}_i^2 = (\tilde{T}_i^v - \tilde{T}_{i-1}^v)^{-1} \sum_{\tilde{T}_{i-1}^v+1}^{\tilde{T}_i^v} \hat{u}_t^2$ and the estimates are constructed by maximizing the likelihood function (7) subject to the restrictions imposed by the set Λ_ε . The test statistic suggested is then

$$\sup F_{4,T}(m_a, n_a, \varepsilon | m = 0, n_a = 0) = \sup F_{3,T} + \sup F_T^\sigma$$

It is easy to show that, under A2 and A4*, the limit distribution of $\sup F_{4,T}$ is the same as the modified LR test in the case of martingale difference errors, i.e., given by the random variable (14).

5.3 A double maximum test

The tests discussed above need the prior information of the specification of the alternative hypothesis, i.e., the number of breaks in regression parameters and in the variance of the errors. However, in practice, researchers may lack such information, hence the need for the testing problems TP-5 to TP-8. Bai and Perron (1998) proposed so-called double maximum tests to solve this problem in a model with only breaks in the parameters. They are tests of no structural break against an unknown number of breaks given some upper bound. Bai and Perron (1998) suggested two versions of such tests. The first is an equal-weight version labelled *UD* max. It can be given a Bayesian interpretation in which the prior assigns equal weights to the possible number of changes. The second test applies weights to the individual tests such that the marginal p-values are equal across values of m and n and is denoted *WD* max. Bai and Perron (2006) showed via simulations that the two versions have similar finite sample properties. Hence, we shall only consider the *UD* max test given that it is simpler to construct.

The Double Maximum test can play a significant role in testing for structural changes and it is arguably the most useful tests to apply when trying to determine if structural changes are present. While the test for one break is consistent against alternatives involving multiple changes, its power in finite samples can sometimes be poor. First, there are types of multiple structural changes that are difficult to detect with a test for a single change (for example,

two breaks with the first and third regimes the same). Second, tests for a particular number of changes may have non monotonic power when the number of changes is greater than specified. Third, the simulations of Bai and Perron (2006) show, in the context of testing for changes in the regression coefficients, that the power of the double maximum tests is almost as high as the best power that can be achieved using the test that accounts for the correct number of breaks. All these elements strongly point to their usefulness.

For each testing problem, the tests and their limit distributions are presented in the following Theorem.

Theorem 2 *Under the relevant null hypothesis, we have, as $T \rightarrow \infty$,*

a) *For TP-5, under A2 and either A4 or A4* :*

$$\begin{aligned} UD \max LR_{1,T} &= \max_{1 \leq n_a \leq N} \sup LR_{1,T}^* (n_a, \varepsilon | m = n = 0) \\ &\Rightarrow \max_{1 \leq n_a \leq N} \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{aligned}$$

b) *For TP-6, under A1 and either A4 or A4* :*

$$\begin{aligned} UD \max LR_{2,T} &= \max_{1 \leq n_a \leq N} \sup LR_{2,T}^* (m_a, n_a, \varepsilon | n = 0, m_a) \\ &\Rightarrow \max_{1 \leq n_a \leq N} \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}^c} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \\ &\leq \max_{1 \leq n_a \leq N} \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{aligned}$$

c) *For TP-7, under A1-A3:*

$$\begin{aligned} UD \max LR_{3,T} &= \max_{1 \leq m_a \leq M} \sup LR_{3,T} (m_a, n_a, \varepsilon | m = 0, n_a) \\ &\Rightarrow \max_{1 \leq m_a \leq M} \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}^v} \sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ &\leq \max_{1 \leq m_a \leq M} \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}} \sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \end{aligned}$$

d) For TP-8, under A2 and A4:

$$\begin{aligned}
UD \max LR_{4,T} &= \max_{1 \leq n_a \leq N} \max_{1 \leq m_a \leq M} \sup LR_{4,T}^* (m_a, n_a, \varepsilon | n = m = 0) \\
&\Rightarrow \max_{1 \leq n_a \leq N} \max_{1 \leq m_a \leq M} \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \left[\begin{aligned} &\sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ &+ \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{aligned} \right]
\end{aligned}$$

For TP-5 to TP-7, the critical values of the limit distributions are available from Bai and Perron (1998, 2003b) for N or M equal to 5. Note that for the testing problems TP-5 and TP-6, the results are valid whether the errors are martingale differences or whether serial correlation is allowed. This is not the case for TP-7 and TP-8 for the same reasons as discussed above that the likelihood ratio tests are not applicable when the errors are serially correlated. In this case, we consider the maximum of the Wald-type test and the results are presented in the following Theorem.

Theorem 3 *Under the relevant null hypothesis, we have, as $T \rightarrow \infty$,*

a) For TP-7, under A2 and A3*:

$$\begin{aligned}
UD \max F_{3,T} &= \max_{1 \leq m_a \leq M} \sup F_{3,T} (m_a, n_a, \varepsilon | m = 0, n_a) \\
&\Rightarrow \max_{1 \leq m_a \leq M} \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}} \sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\
&\leq \max_{1 \leq m_a \leq M} \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}} \sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)}
\end{aligned}$$

b) For TP-8, under A2 and A4*:

$$\begin{aligned}
UD \max F_{4,T} &= \max_{1 \leq n_a \leq N} \max_{1 \leq m_a \leq M} \sup F_{4,T} (m_a, n_a, \varepsilon | n = m = 0) \\
&\Rightarrow \max_{1 \leq n_a \leq N} \max_{1 \leq m_a \leq M} \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \left[\begin{aligned} &\sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ &+ \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{aligned} \right]
\end{aligned}$$

The limit distribution applicable for the testing problem TP-8 is new. We obtained critical values using simulations as discussed above for the case of a fixed number of breaks under the alternative hypothesis. These are presented in Table 1 for $\varepsilon = 0.1, 0.15$, and 0.20 , and values of M and N up to 2.

5.4 Testing for an additional break

We now consider the testing problems TP-9 and TP-10, which looks at whether including an additional break is warranted. Let $(\tilde{T}_1^c, \dots, \tilde{T}_m^c; \tilde{T}_1^v, \dots, \tilde{T}_n^v)$ denote the estimates of the break dates in the regression coefficients and the variance of the errors obtained jointly by maximizing the quasi-likelihood function assuming m breaks in the coefficients and n breaks in the variance.

For the testing problem TP-9, the issue is whether an additional break in the regression coefficients is present. Following Bai and Perron (1998) and Qu and Perron (2007a), the test is

$$\begin{aligned} \sup Seq_T(m+1, n|m, n) &= \max_{1 \leq j \leq m+1} \sup_{\tau \in \Lambda_{j,\varepsilon}^c} LR_T(\tilde{T}_1^c, \dots, \tilde{T}_{j-1}^c, \tau, \tilde{T}_j^c, \dots, \tilde{T}_m^c; \tilde{T}_1^v, \dots, \tilde{T}_n^v) \\ &\quad - LR(\tilde{T}_1^c, \dots, \tilde{T}_m^c; \tilde{T}_1^v, \dots, \tilde{T}_n^v) \end{aligned}$$

where

$$\Lambda_{j,\varepsilon}^c = \{\tau; \tilde{T}_{j-1}^c + (\tilde{T}_j^c - \tilde{T}_{j-1}^c)\varepsilon \leq \tau \leq \tilde{T}_j^c - (\tilde{T}_j^c - \tilde{T}_{j-1}^c)\varepsilon\} \quad (16)$$

This amounts to performing $m+1$ tests for a single break in the regression coefficients for each of the $m+1$ regimes defined by the partition $\{\tilde{T}_1^c, \dots, \tilde{T}_m^c\}$. Note that there are different scenarios when allowing breaks in coefficients and in the variance to happen at different dates, since $(\tilde{T}_1^c, \dots, \tilde{T}_m^c)$ and $(\tilde{T}_1^v, \dots, \tilde{T}_n^v)$ can partly or completely overlap or be altogether different. This implies two possible cases: 1) if the n break dates in variance are a subset of the m break dates in coefficients, then there is no variance break between \tilde{T}_{j-1}^c and \tilde{T}_j^c ; 2) otherwise, there is one or more variance breaks between \tilde{T}_{j-1}^c and \tilde{T}_j^c . In either cases, one can appeal to the results of part (c) of Theorem 1 with $m_a = 1$ since any value of n_a (the number of breaks in variance) is allowed, including 0. It is then easy to deduce that, in the case of martingale errors, the limit distribution of the test is, under Assumptions A2 and A3,

$$\lim_{T \rightarrow \infty} P(\sup Seq_T(m+1, n|m, n) \leq x) = G_{q,\varepsilon}(x)^{m+1}$$

where $G_{q,\varepsilon}(x)$ is the cumulative distribution function of the random variable

$$\sup_{\lambda \in \Lambda_{1,\varepsilon}} \frac{(W_q(\lambda) - \lambda W_q(1))^2}{\lambda(1-\lambda)}. \quad (17)$$

where $\Lambda_{1,\varepsilon} = \{\lambda; \varepsilon < \lambda < 1 - \varepsilon\}$. The critical values of the distribution function $G_{q,\varepsilon}(x)^{m+1}$ can be found in Bai and Perron (1998, 2003b). When serial correlation in the error, the

principle is the same except that the statistic is based on the robust Wald test $\sup F_{3,T}$ as defined by (15) applied for a one break test to each segment.

For the testing problem TP-10, similar considerations apply. Here the issue is whether an additional break in the variance is present. The test statistic is

$$\begin{aligned} \sup Seq_T(m, n+1|m, n) &= \left(2/\hat{\psi}\right) \max_{1 \leq j \leq n+1} \sup_{\tau \in \Lambda_{j,\varepsilon}^v} LR_T(\tilde{T}_1^c, \dots, \tilde{T}_m^c; \tilde{T}_1^v, \dots, \tilde{T}_{j-1}^v, \tau, \tilde{T}_j^v, \dots, \tilde{T}_m^v) \\ &\quad - LR(\tilde{T}_1^c, \dots, \tilde{T}_m^c; \tilde{T}_1^v, \dots, \tilde{T}_n^v) \end{aligned}$$

where

$$\Lambda_{j,\varepsilon}^v = \{\tau; \tilde{T}_{j-1}^v + (\tilde{T}_j^v - \tilde{T}_{j-1}^v)\varepsilon \leq \tau \leq \tilde{T}_j^v - (\tilde{T}_j^v - \tilde{T}_{j-1}^v)\varepsilon\}.$$

The correction factor $(2/\hat{\psi})$ is needed to ensure that the limit distribution of the test is free of nuisance parameters when the errors are allowed to be non-Normal, serially correlated and conditionally heteroskedastic. One can then use part (b) of Theorem 1 to deduce that, under A1 and A4, or A1 and A3* applied to each segments under the null hypothesis,

$$\lim_{T \rightarrow \infty} P(\sup Seq_T(m, n+1|m, n) \leq x) = G_{1,\varepsilon}(x)^{n+1}.$$

6 Monte Carlo experiments

This section presents the results of simulation experiments to address the following issues: 1) which particular version of the correction factor $\hat{\psi}$ has better finite sample properties?; 2) whether applying a correction valid under more general conditions than needed is detrimental to the size and power of the test; 3) the finite sample size and power of the various tests proposed. Throughout, we use 1,000 replications.

6.1 The choice of $\hat{\psi}$

To address what specific version of the correction factor to use, we consider the size and power of the $\sup LR_{4,T}^*$ test under the following simple Data Generating Process (DGP) with ARCH(1) errors:

$$\begin{aligned} y_t &= \mu_1 + \mu_2 1(t > [.25T]) + e_t, \\ e_t &= u_t \sqrt{h_t}, \\ u_t &\sim i.i.d. N(0, 1), \\ h_t &= \delta_1 + \delta_2 1(t > [.75T]) + \gamma e_{t-1}^2, \end{aligned}$$

with $h_0 = \delta_1 / (1 - \gamma)$ and $\delta_1 = 1$. The sample size is $T = 100$ and we set the truncation to $\varepsilon = 0.20$. Under the null hypothesis of no change $\mu_2 = \delta_2 = 0$, while under the alternative hypothesis one break in mean and one break in variance are allowed. $\mu_1 = 0$ under both the null and alternative hypothesis. We consider three versions for the estimate $\hat{\psi}$ as defined by (11): 1) using the residuals under the alternative hypothesis to construct the bandwidth m and to estimate the relevant autocovariances of η_t (labelled “alternative”); 2) using the residuals under the null hypothesis instead (labelled “null”); and, as suggested by Kejriwal and Perron (2006a), 3) using a hybrid method that constructs the bandwidth m using the residuals under the alternative hypothesis but uses the residuals under the null to estimate the relevant autocovariances of η_t (labelled “hybrid”).

The results for the exact size of the test (using a 5% nominal size test) are presented in Table 2. They show the method “alternative” to exhibit substantial size distortions, that increase as γ , which indicates the extent of the correlation in the squared residuals, increases. The method “null”, on the other hand, shows conservative size distortions. Finally, the hybrid method shows an exact size close to the nominal level for all cases considered.

The results for power are presented in Table 3. We only consider the methods “null” and “hybrid” given the high size distortions of the method “alternative”. They show that substantial power gains can be achieved using the “hybrid” method as opposed to the “null” method, especially if the ARCH effect is pronounced. Hence, we recommend using the “hybrid” method and all results below will be based on it.

6.2 Should we always correct?

We now address the issue of whether it is costly in terms of power to use a correction valid under more general conditions than needed. To that effect we first consider the power of the sup $LR_{4,T}^*$ test under the following DGP with Normal errors:

$$\begin{aligned} y_t &= \mu_1 + \mu_2 1(t > T_1^c) + e_t, \\ e_t &\sim i.i.d. N(0, 1 + \delta 1(t > T_1^v)), \end{aligned}$$

where we set $\mu_1 = 0$ and $\mu_2 = \delta$. We consider three scenarios for the timing of the breaks: a common break in mean and variance at $T_1^c = T_1^v = [.5T]$, and disjoint breaks at $\{T_1^c = [.3T], T_1^v = [.6T]\}$ and $\{T_1^c = [.6T], T_1^v = [.3T]\}$. We use two sample sizes, $T = 100, 200$ and the power, for 5% nominal size tests, is evaluated at values of δ ranging from 0.25 to 1.5. Three versions of the sup $LR_{4,T}^*$ tests are evaluated: 1) with a full correction based on $\hat{\psi}$ as defined by (11) using the hybrid method, which is valid for errors that can be conditionally

heteroskedastic and serially correlated (labelled “full”); 2) a correction that is valid for *i.i.d.* errors, though not necessarily Normal, given by $\hat{\psi} = \hat{\mu}_4/\hat{\sigma}^4 - 1$, where $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$ and $\hat{\mu}_4 = T^{-1} \sum_{t=1}^T \hat{u}_t^4$ with \hat{u}_t the residuals under the null hypotheses (labelled “i.i.d.”); 3) no correction, i.e., using $\hat{\psi} = 2$, which is the appropriate value with Normal errors (labelled “NC”). The results are presented in Table 4. They show that the power is basically the same using any of the three methods. Hence, there is no cost in this case in using a full correction.

Table 5 presents related results for the test $\sup LR_{1,T}^*$, which tests for a single break in variance assuming no break in the regression coefficients. The DGP is $y_t = e_t$, with $e_t \sim i.i.d. N(0, 1 + \delta 1(t > [.5T]))$ and δ varies between 0 and 1.5. The full correction yields power similar to a correction that correctly assumes *i.i.d.* errors, though here imposing Normality can lead to tests with somewhat higher power.

Overall, using the full correction entails little power loss and, hence, we shall continue to use it in all results below. There may be cases where correctly imposing Normality can lead to tests with slightly higher power but this can be achieved only if one has the correct prior knowledge of the true distribution of the errors, a case that is unlikely to occur in practice.

6.3 Testing for variance breaks only

We now consider the case of testing only for variance breaks assuming no change in regression coefficients. To that effect we shall investigate the properties of the following tests: the $\sup LR_{1,T}^*(n_a, \varepsilon | m = n = 0)$, abbreviated $\sup LR_{1,T}^*(n_a, \varepsilon)$, the $UD \max LR_{1,T}$ for testing versus an unknown number of breaks up to $N = 5$, and a corrected version of the CUSQ given by

$$CUSQ^* = \frac{\sup_{\lambda \in [0,1]} \left| T^{-1/2} \left[\sum_{t=1}^{[T\lambda]} \tilde{v}_t^2 - \frac{[T\lambda]}{T} \sum_{t=1}^T \tilde{v}_t^2 \right] \right|}{\hat{\varphi}_a^{1/2}}$$

with

$$\hat{\varphi}_a = \sum_{j=-(T-1)}^{(T-1)} w(j, m) \sum_{t=|j|+1}^T \hat{\eta}_t \hat{\eta}_{t-j}$$

where $\hat{\eta}_t = \tilde{v}_t^2 - \hat{\sigma}^2$, with $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \tilde{v}_t^2$ and where \tilde{v}_t denotes the recursive residuals. Here also $w(j, m)$ is the Quadratic Spectral kernel and the bandwidth m is selected using Andrews’ (1991) method with an AR(1) approximation. The aim of the design is to address the following issues: a) the size of the $\sup LR_{1,T}^*(n_a, \varepsilon)$ and $UD \max LR_{1,T}$ tests; b) the relative power of the three tests; c) the power losses obtained when under-specifying the number of

breaks; d) the relative power of the $UD \max LR_{1,T}$ compared to the $\sup LR_{1,T}^*(n_a, \varepsilon)$ with n_a specified to be the true number of breaks.

We start with a simple design with Normal errors and the DGP is $y_t = e_t$; $e_t \sim i.i.d. N(0, 1 + \delta 1(t > [.5T]))$. We use $T = 100, 200$ and for the $\sup LR_{1,T}^*(1, \varepsilon)$ and $UD \max LR_{1,T}$ tests the trimming parameter is set to $\varepsilon = 0.25$. The coefficient δ varies between 0 (size) and 1.5. The results are presented in Table 6. They show the exact sizes of the $\sup LR_{1,T}^*(1, \varepsilon)$ and $UD \max LR_{1,T}$ tests to be close to the nominal 5% size. The $CUSQ^*$ test is slightly undersized. The highest power is achieved using the $\sup LR_{1,T}^*(1, \varepsilon)$ test. Interestingly, the $UD \max LR_{1,T}$ test has power very close to that of the $\sup LR_{1,T}^*(1, \varepsilon)$ test, even though the range of alternatives considered is broader.

We next consider a dynamic model with ARCH errors, for which the DGP is given by

$$\begin{aligned} y_t &= c + \alpha y_{t-1} + e_t, \\ e_t &= u_t \sqrt{h_t}, \\ u_t &\sim i.i.d. N(0, 1), \\ h_t &= \delta_1 + \delta_2 1(t > [.5T]) + \gamma e_{t-1}^2, \end{aligned}$$

where we set $h_0 = \delta_1 / (1 - \gamma)$, $c = 0.5$, $\delta_1 = 0.1$, and the trimming parameter is again $\varepsilon = 0.25$. We consider two values of the autoregressive parameter $\alpha = 0.2, 0.7$, the ARCH coefficient is set to $\gamma = 0.1, 0.3$ and 0.5 , and again the size and power of 5% nominal size tests are evaluated at $T = 100, 200$. The magnitude of the change δ_2 varies between 0 (size) and 0.30. The results are presented in Table 7. They show again the exact sizes of the $\sup LR_{1,T}^*(1, \varepsilon)$ and $UD \max LR_{1,T}$ tests to be close to the nominal 5% size. The $CUSQ^*$ test is again slightly undersized but more so as γ increases. The $UD \max LR_{1,T}$ test still has power very close to that of the $\sup LR_{1,T}^*(1, \varepsilon)$ test, even though the range of alternatives considered is broader. The power of the latter two tests dominates that of the $CUSQ^*$ especially when $T = 100$, in which case the discrepancies can be substantial.

We now turn to a case with two breaks in variance. The DGP is $y_t = e_t$; $e_t \sim i.i.d. N(0, 1 + \delta 1([.3T] < t \leq [.6T]))$. This specifies a model where the variance increases at $[.3T]$ and returns back to its original level at $[.6T]$. The sample size is set to $T = 200$. We consider four values of the trimming parameter $\varepsilon = 0.10, 0.15, 0.20$ and 0.25 . The magnitude of the break in variance varies between $\delta = 0$ (size) and $\delta = 3$. We again consider the $UD \max LR_{1,T}$ test with $N = 5$ but include both the $\sup LR_{1,T}^*(1, \varepsilon)$ test for a single break and the $\sup LR_{1,T}^*(2, \varepsilon)$ test for two breaks to assess the extent of power gains when specifying the correct number of breaks. The results are presented in Table 8. Consider first

the size of the tests. The $\sup LR_{1,T}^*(1, \varepsilon)$, $\sup LR_{1,T}^*(2, \varepsilon)$ and $UD \max LR_{1,T}$ are slightly conservative when ε is small but less so as ε increases. The $CUSQ^*$ is more conservative with an exact size of 0.026. As expected, power increases as ε increases since the class of alternatives is smaller. When comparing the $\sup LR_{1,T}^*(1, \varepsilon)$ and $\sup LR_{1,T}^*(2, \varepsilon)$ tests, the latter is substantially more powerful, indicating that allowing for the correct number of breaks is important for power considerations. Here, the $UD \max LR_{1,T}$ is not as powerful as the $\sup LR_{1,T}^*(2, \varepsilon)$ but more powerful than the $\sup LR_{1,T}^*(1, \varepsilon)$. All versions of these tests are considerably more powerful than the $CUSQ^*$, which shows little power.

6.4 Conditional tests

We now consider the properties of the tests that condition on either breaks in coefficients (resp., variance) when testing for changes in variance (resp., coefficients). Consider first the size and power of $\sup LR_{2,T}^*(m_a, n_a, \varepsilon | n = 0, m_a)$ which tests for n_a changes in variance conditional on m_a changes in regression coefficients. We set $m_a = n_a = 1$ and the DGP is a simple mean shift model with change in mean of magnitude μ_2 at mid-sample with *i.i.d.* Normal errors having a change in variance of magnitude δ (under the alternative hypothesis) that occurs at $[0.25T]$. The results for size are presented in Table 9. When there is no change in mean ($\mu_2 = 0$), the test exhibits liberal size distortions, as expected since the limit distribution is inappropriate. The exact size approaches the 5% nominal size rather quickly as μ_2 increases, and more so the larger the trimming ε and/or the sample size T . When the change in mean is very large, the test is conservative and more so as the trimming is larger. This is due to the fact that the limit distribution used is actually an upper bound as discussed in Remark 1. The results for power are presented in Table 10. Given the fact that the test is conservative as μ_2 increases, the power accordingly decreases, though not to a large extent. It increases rapidly with the sample size and marginally with the value of the trimming ε .

6.5 Size and power of the $\sup LR_{4,T}^*$ and $UD \max$ tests

We now present results about the properties of the $\sup LR_{4,T}^*$ and $UD \max$ tests. We first consider the size of the tests with normal *i.i.d.* errors, with the DGP set to $y_t = e_t \sim i.i.d. N(0, 1)$. We use three values of the trimming parameter $\varepsilon = 0.1, 0.15$ and 0.2 . For the $UD \max$ test, $M = N = 2$ and for the $\sup LR_{4,T}^*$ test, we consider the following combinations: a) $m_a = n_a = 1$, b) $m_a = 1, n_a = 2$, c) $m_a = 2, n_a = 1$. Two sample sizes are used, $T = 100, 200$. The results are presented in Table 11 and they show the size to be close to the nominal

5% level. Table 12 presents the results of a similar experiment but with ARCH(1) errors so that the DGP is $y_t = e_t$ with $e_t = u_t\sqrt{h_t}$, where $u_t \sim i.i.d. N(0, 1)$, $h_t = \delta_1 + \gamma\varepsilon_{t-1}^2$, $h_0 = \delta_1/(1 - \gamma)$, $\delta_1 = 1$ and γ takes values 0.1, 0.3 and 0.5. There are some cases with some slight size distortions when $T = 100$ but these quickly decrease when $T = 200$.

We now consider the power of the test. Since some partial results for the one break case are available in Tables 3 and 4 for the sup $LR_{4,T}^*$ test, we shall concentrate on the case with a different number of breaks in coefficients and in variance. We also only consider Normal errors in the DGP though the hybrid-type correction is still applied. Table 13 presents the results for the case with one break in coefficients and two breaks in variance, in which case the DGP is given by

$$\begin{aligned} y_t &= \mu_1 + \mu_2 1(t > T^c) + e_t, \\ e_t &\sim i.i.d. N(0, 1 + \delta 1(T_1^v < t \leq T_2^v)) \end{aligned}$$

with $\mu_1 = 0, \mu_2 = \delta$. For the tests the trimming parameter used is $\varepsilon = 0.1$. Five different configurations of break dates are considered. We analyze two forms of the sup $LR_{4,T}^*$ test: a) one that tests for a single break in both mean and variance, b) one that correctly tests for two changes in variance and one change in mean. This is done to investigate the extent of the power differences when underspecifying the number of breaks. As expected, the power increases rapidly with δ and with T . With the DGP used, the power is similar whether accounting for one or (correctly) two breaks in variance. The power of the UD max test is somewhat below the power of both versions of the sup $LR_{4,T}^*$ test. This may, however, be specific to the DGP considered.

Table 14 presents the results for the case with two breaks in coefficients and one break in variance, in which case the DGP is given by

$$\begin{aligned} y_t &= \mu_1 + \mu_2 1(T_1^c < t \leq T_2^c) + e_t, \\ e_t &\sim i.i.d. N(0, 1 + \delta 1(t > T^v)) \end{aligned}$$

with $\mu_1 = 0$ and $\mu_2 = \delta$. Again, we consider two forms of the sup $LR_{4,T}^*$ test: a) one that tests for a single break in both mean and variance, b) one that correctly tests for two changes in mean and one change in variance. A first difference is the fact that for given values of δ and T , the power is lower than in the case of one break in coefficient and two breaks in variance, indicating that changes in variance are easier to detect than changes in mean. The second difference is that the UD max test now has power in between that of the test that correctly specifies the type and number of breaks and the one that underspecifies the number

of changes in mean. The difference can indeed be substantial and, in line with the results of Bai and Perron (2006), the power of the UD max test is close to the power attainable when correctly specifying the type and number of breaks.

7 Estimating the numbers of breaks in coefficients and in variance

In this Section, we discuss a specific to general sequential procedure to estimate the number of breaks in the coefficients of the conditional mean and in the variance. The starting point is the use of a modification of the sequential procedure discussed in Qu and Perron (2007a). Our problem is, however, more complex since we wish to ascertain what types of break occur at any given selected break date, not only to know whether some kind of break did occur. Hence, the need for some refinements. The main difficulty is the fact that if a break occurs, it can be associated with a change in either or both the regression coefficients and the variance, and a method to decide which case it is in effect needs to be applied.

The starting point is to modify the $\sup LR_{4,T}^*$ so that it can be applied in a sequential manner to address the testing problem

$$H_0 : \{m = \ell, n = \ell\} \text{ versus } H_1 : \{m = \ell + 1, n = \ell + 1\}$$

The procedure is to test the null hypothesis of ℓ breaks versus the alternative hypothesis of $\ell + 1$ breaks by performing a one break test for each of the $\ell + 1$ segments defined by the partition $(\hat{T}_1, \dots, \hat{T}_l)$, which are the estimates of the break dates obtained by maximizing the Gaussian likelihood function defined by (7) with $T_j^c = T_i^v = T_k$. The test statistic is then the maximal value of the tests over all $\ell + 1$ segments, denoted $\sup Seq_T(\ell + 1|\ell)$. It follows that the limit distribution of the test is given by

$$\lim_{T \rightarrow \infty} P(\sup Seq_T(\ell + 1|\ell) \leq x) = G_4(x)^{\ell+1}$$

where $G_4(x)$ is the distribution function of the random variable defined by (17) with $q + 1$. Upon a rejection, a model with $\ell + 1$ breaks is considered with the additional break being inserted within the segment associated with the maximal value of the tests at the position that maximizes the likelihood function. This procedure is iterated until a non-rejection occurs. Let the number of breaks thus selected be denote by \bar{K} .

The next step is to decide whether a break in coefficients, in variance or in both has occurred at each of the selected break dates. We then perform standard hypothesis testing for the equality of the parameters across adjacent segments. Since the limit distribution of

the estimates of the parameters of the model are the same whether using estimates of the break dates or their true value, standard procedures can be applied. Consider first the case of testing whether the regression coefficients are equal across the two regimes $(\hat{T}_{k-1}, \hat{T}_k)$, regime k , and $(\hat{T}_k, \hat{T}_{k+1})$, regime $k+1$, separated by the k^{th} break ($k = 1, \dots, \bar{K}$). Denote the true value of the regression coefficients in regimes k and $k+1$ by β_k and β_{k+1} , respectively. The null hypothesis is then $H_0 : \beta_k = \beta_{k+1}$ and the alternative hypothesis, $H_1 : \beta_k \neq \beta_{k+1}$. Note that since there is a break in either the regression coefficients and/or the variance of the errors, under the null hypothesis there must be a change in the variance of the errors. Hence, the test to be applied is a standard Chow-type test allowing for a change in variance across regimes (see Goldfeld and Quandt, 1978).

Consider now the testing problem $H_0 : \sigma_k^2 = \sigma_{k+1}^2$ versus $H_1 : \sigma_k^2 \neq \sigma_{k+1}^2$, where σ_k^2 and σ_{k+1}^2 are the variances of the errors in regimes k and $k+1$, respectively. The Wald test corrected for potential non-normality and conditional heteroskedasticity is given by

$$W_k = \frac{(\hat{T}_k - \hat{T}_{k-1})(\hat{T}_{k+1} - \hat{T}_k)}{(\hat{T}_{k+1} - \hat{T}_{k-1})(\hat{\mu}_4 - \hat{\sigma}^4)} (\hat{\sigma}_{k+1}^2 - \hat{\sigma}_k^2)^2,$$

where $\hat{\sigma}_k^2$ and $\hat{\sigma}_{k+1}^2$ are the maximum likelihood estimates of σ_k^2 and σ_{k+1}^2 (for the same reasons discussed above these are constructed allowing the regression coefficients to be different in regimes k and $k+1$), and $\hat{\mu}_4$ is a consistent estimate of $E(u_t^4)$, e.g., $\hat{\mu}_4 = (\hat{T}_{k+1} - \hat{T}_{k-1})^{-1} \sum_{\hat{T}_{k-1}+1}^{\hat{T}_{k+1}} \hat{u}_t^4$, constructed under the alternative hypothesis to maximize power.

To assess the finite sample properties of this specific to general procedure, we performed a simple simulation experiment. The basic DGP is

$$\begin{aligned} y_t &= \mu_1 + \mu_2 \mathbf{1}(t > T^c) + e_t, \\ e_t &\sim i.i.d. N(0, 1 + \delta \mathbf{1}(t > T^v)), \end{aligned}$$

so that a maximum of one break in either mean or variance is allowed. The sample size is $T = 200$ and the tests are constructed with a trimming $\varepsilon = 0.15$. We consider the following scenarios: a) no change in either mean or variance, b) a change in mean only occurring at mid-sample, c) a change in variance only also occurring at mid sample, d) a change in both mean and variance occurring at a common date (mid-sample); e) a change in both mean and variance occurring at different but close dates ($T^c = [0.5T]$, $T^v = [0.7T]$); f) a change in both mean and variance occurring at different and distant dates ($T^c = [0.25T]$, $T^v = [0.75T]$). Whenever breaks occur, different magnitudes are considered.

The results are presented in Table 15. In general, the procedure works quite well in selecting the correct number and type of breaks. There are cases, however, where the probability

of making the correct selection is quite low. This occurs when both changes in mean and variance are not large and occur at different dates, especially when the respective break dates are far apart. The basic reason for that is the fact that the $\sup Seq_T(\ell + 1|\ell)$ statistic jointly tests whether a break in both regression coefficients and variance occur. Hence, if only one type of break occurs the power can be quite low unless the magnitudes of the breaks are large. Unfortunately, this situation is expected to be quite common in practice, as we shall see in the empirical applications reported in the next section. Hence, though this specific to general procedure is valid in large samples, it should not be applied mechanically. Care must be exercised to assess whether we are in a situation where its finite sample properties are rather poor.

An alternative approach is to use a general to specific type of procedure to determine the appropriate number and type of breaks. This involves using the battery of tests that we presented in this paper in a judicious way. The procedure cannot be mechanized but is likely to deliver better results. We shall illustrate how to apply it in the context of actual applications to be discussed in the next section.

8 Applications

The set of testing procedures we developed provide useful tools to detect jointly structural changes in the unconditional variance of the errors and the parameters of the conditional mean in a linear regression model. To our knowledge no such test is available under the level of generality that we consider. This is important for practical applications as witnessed by recent interest in macroeconomic and finance where documenting structural changes in the variability of shocks to simple autoregressions or Vector Autoregressive Models has been a concern; see, among others, Blanchard and Simon (2001), Herrera and Pesavento (2005), Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Sensier and van Dijk (2004) and Stock and Watson (2002).

Stock and Watson (2002) present an exhaustive analysis documenting facts about potential changes in the volatility of macroeconomic time series using the two step approach described in Section 2. Of interest here is the fact that many such series seem to have experienced a decline in volatility in the mid 80s. We reconsider the analysis presented in their Table 2 pertaining to 22 common macroeconomic variables. These are quarterly series covering the period 1960-2001, whose list is contained in Table 16, along with the relevant

transformation to eliminate trend and/or unit root ¹. Graphs of the series are presented in Figures 7 and 8. With the variables transformed to achieve stationarity the basic regression is a simple AR(4) with a fitted intercept.

We first discuss how we used our testing procedures to select the number of breaks in coefficients (intercept and autoregressive parameters) and in variance. With the types of breaks in the series analyzed, the sequential procedure did not perform well. This is due to the fact that in most cases changes in both the coefficients and the variance did occur and they did so at different times, a case for which the specific to general procedure performs poorly. Hence, we used a procedure more akin to a general to specific one. To start, we set an upper bound of two breaks in each of the coefficients and variance, which means a maximum of four breaks overall. This should be enough for the types of series analyzed. In any event, we also present evidence that two breaks are enough.

The first statistic used is the UD max test with $M = N = 2$. We report in Table 16 the outcome of this test with the trimming parameter $\varepsilon = 0.15$ and $\varepsilon = 0.20$. It shows significant evidence for at least some breaks for 14 of the 22 series. For the eight series for which this test is not significant, Stock and Watson reported evidence of some breaks for five of these: consumption, change in inventory investment, total production of goods, nondurable goods and non-agricultural employment.

For those series for which the UD max test shows a rejection, we computed a wide range of tests to decide which model to select. To illustrate, consider the case of the GDP series for which we select $m = 1$ and $n = 2$. The $\sup LR_{4,T}^*(1, 2)$ is indeed significant at the 1% level. We then consider the sequential test $\sup Seq_T(m, n + 1|m, n)$ to see if too many or too few breaks are included. The test $\sup Seq_T(2, 2|1, 2)$ is insignificant indicating that an additional break in coefficients is unwarranted. The $\sup Seq_T(1, 2|1, 1)$ test is significant indicating that a second break in variance is warranted. The $\sup LR_{3,T}(1, 2|0, 2)$ test is not significant at the 10% level but marginally, so that given the low power of this test induced by the fact that 5 parameters are allowed to change, we decided to keep one break in the coefficients (the parameter estimates will indeed show an important change). Finally, the $\sup Seq_T(1, 3|1, 2)$ is not significant indicating that a third break in variance is not needed. This is the basic procedure that is repeated for all series. The model selected is presented in the fourth column of Table 17. In all cases, at least one change in coefficients and in variance occurs and often two of each do.

¹The data source is the DRI-McGraw Hill Basic Economics database. The series were kindly posted by Mark Watson on his web page.

Table 17 presents the key parameter estimates: a) the break dates in coefficients (T_1^c and T_2^c) and in variance (T_1^v and T_2^v); b) the value of the intercept in each regime ($\alpha_i, i = 1, 2, 3$) to assess whether there are important level shifts; c) the sum of the autoregressive coefficients in each regime ($\beta_i, i = 1, 2, 3$) to assess whether there are changes in persistence induced by different propagation mechanisms; d) the standard deviation of the errors in each regime ($\sigma_i, i = 1, 2, 3$) to quantify the magnitude of the change in variance (we also present in the last columns various ratios to help gauge the relative magnitudes across regimes).

The first thing to note is that if one looks at the ratio of the standard deviation of the errors for the last regime compared to the preceding one, there is indeed strong evidence of a change mostly for a substantial decrease (GDP, consumption of durables and non-durables, fixed investment-total, residential investment, production of durable goods and structures, inflation, the T-bill and T-bond rates). There are, however, several cases where the evidence shows a substantial increase in the variance of the errors: the consumption and production of services, exports and imports. Hence, the so-called great moderation did not occur across all sectors. The last break date is estimated to be in the mid-80s for most series with some exceptions for which it occurred in the early 90s.

The results show many additional features that are of interest. Consider first the case when two breaks in the variance of the shocks did occur. What transpires is that there is a tendency to revert back to the level of the first regime. For example, for GDP, the ratio σ_3/σ_2 is 0.35 while σ_3/σ_1 is 0.67. For inflation, the variance after 1986:1 reverts back exactly to its level prior to 1971:3. For the interest rate series, there is actually an increase in variance after the mid-80s compared to before 1967:4 for the T-bill rate and before 1979:3 for the T-bond rate. So this so-called great moderation may be qualified as a phenomenon where the high variance level of the 70s to early 80s are over and we are back to the level of (roughly) pre-70s; sometimes this reversion is exact (e.g., inflation), incomplete (e.g., interest rates) or magnified (real variables). Hence, the so-called “great-moderation” may rather be qualified as a “great-reversion”.

With respect to the intercept of the regression (the long term level) there is not much evidence of significant changes with the following exceptions. For “exports”, there is a decrease in 1972:4 and an increase in 1992:1. For “imports”, we have a mirror image with an increase in 1967:1 and a decrease in 1990:4. The other series with important level shifts are the interest rate series: for the 90-day T-bill rate, an increase in 1967:4 and a decrease in 1983:4, the pattern is similar for the 10-year T-bond rate but the increase occurs in 1979:3 and the decrease in 1986:4.

With respect to the sum of the autoregressive coefficients which can be labelled as the persistence of the series, there are important changes. For GDP, the consumption and the investment series, the results point to a one-time increase, though the dates are different. For the production of services, there are two increases, in 1968:3 and 1982:4. For some other series the pattern is more complex and interesting. For “inflation”, there is a substantial increase in 1971:3 and a large decrease in 1986:1 (the pattern is similar for the series “production of structures”, though the dates and relative magnitudes are different). For the 90-day T-bill rate, there is a decrease in 1967:4 and an increase in 1983:4. Interestingly, for the 10-year T-bond rate, the pattern is reversed with an increase in 1979:3 and a decrease in 1986:4. The most peculiar results are, however, for the “imports” and “exports” series. For both, there are two changes in variance and in coefficients. In each regime, the variance of the shocks is the same for the two series. However, the pattern for the measure of persistence is completely different. For “exports”, there is a very large increase in 1972:4 followed by a large decrease in 1992:1, while for “imports” we have a large decrease in 1967:1 and a very large increase in 1990:4.

Since the statistic $UD \max LR_{4,T}$ tests jointly for the presence of changes in the regression coefficients and the variance of the errors, it may be the case that it lacks power if only changes in variance occur (especially if the number of regression coefficients allowed to change is large; e.g., 5 in the applications here). In that case, an alternative strategy is possible. It involves using the $UD \max LR_{1,T}$ and $SupLR_{1,T}^*$ tests to assess whether changes in variance are present assuming no change in coefficients, and then use the statistic $\sup LR_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a)$ where n_a is the number of changes in variance found in the first step, as well as the statistic $\sup Seq_T(m, n + 1 | m, n)$. Non-rejections with these tests are then viewed as confirmatory evidence that the results based on the $UD \max LR_{1,T}$ and $SupLR_{1,T}^*$ tests were adequate. We illustrate this approach using the series for which we could not obtain a rejection with the $UD \max LR_{4,T}$ test but for which Stock and Watson (2002) claimed evidence in favor a single change in variance. These series are: consumption, change in inventory-investment, total production of goods and production of nondurables. The results are presented in Table 18. The $UD \max LR_{1,T}$ test is significant at the 10% level at least, except for the total production of goods series for which there is no evidence of change in the variance of the errors. The use of the $SupLR_{1,T}^*$ leads us to conclude that there is one change for the consumption and production of nondurables series, while the evidence points to two changes for the change in inventory investment series. None of the tests $\sup LR_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a)$ and $\sup Seq_T(m, n + 1 | m, n)$, conditional on the num-

ber of breaks in variance found, are significant. Hence, we can be confident that the results based on $UD \max LR_{1,T}$ and $SupLR_{1,T}^*$ are not spurious. The parameter estimates yield the following features. For the consumption series, there is a substantial decrease in variance in 1983:2 a date which is quite different from the date 1992:1 found by Stock and Watson (2002). For the production of nondurables series, there is also a large decrease in variance in 1984:3. For the change in inventory-investment series, there is a large increase in variance in 1973:3 followed by a reversal to roughly the pre-1973 period in 1987:2. Hence, we again have that for series with two changes, the evidence indicates that the decrease in the 80s is indeed a reversal to a previous level.

There is undoubtedly a wealth of interesting features in these results that calls for explanations. This is obviously outside the scope of this paper but hopefully they can spur the interest of macroeconomists.

9 Conclusion

This paper provided tools for testing for multiple structural breaks in the error variance in the linear regression model with or without the presence of breaks in the regression coefficients. An innovation is that we do not impose any restrictions on the break dates, i.e., the breaks in the regression coefficients and in the variance can happen at the same time or at different times. We proposed statistics which have an asymptotic distribution invariant to nuisance parameters and are valid in the presence of non-normal errors and conditional heteroskedasticity, as well as serial correlation. Extensive simulations of the finite sample properties show that our procedures performs well in terms of size and power, though a specific to general procedure to estimate the number and type of breaks has some shortcomings when the breaks in coefficients and in the variance of the errors occur at different dates.

We applied our testing procedures to various macroeconomic time series studied by Stock and Watson (2002). On one hand, our results reinforce the prevalence of changes in both mean, persistence and variance of the shocks in simple autoregressions. Most series have an important reduction in variance that occurred in the 80s. For many series, however, the evidence points to two breaks in the variance of the shocks with the feature that it increases at the first one and decreases at the second. Hence, the so-called “great moderation” may be qualified as a phenomenon where the high variance level of the 70s to early 80s are over and we are back to the level of (roughly) pre-70s. Accordingly, the so-called “great-moderation” may rather be qualified as a “great-reversion”.

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.
- Bai, J. (1997): "Estimation of a change point in multiple regression models," *Review of Economics and Statistics*, 79, 551-563.
- Bai, J. (1999): "Likelihood ratio tests for multiple structural changes," *Journal of Econometrics*, 91, 299-323.
- Bai, J. (2000): "Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices," *Annals of Economics and Finance*, 1, 303-339.
- Bai, J., Lumsdaine, R.L., and Stock, J.H. (1998): "Testing for and dating breaks in multivariate time series," *Review of Economic Studies*, 65, 395-432.
- Bai, J., and Perron, P. (1998): "Estimating and testing linear models with multiple structural changes," *Econometrica*, 66, 47-78.
- Bai, J., and Perron, P. (2003a): "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18, 1-22.
- Bai, J., and Perron, P. (2003b): "Critical values for multiple structural change tests," *Econometrics Journal*, 6, 72-78.
- Bai, J., and Perron, P. (2006): "Multiple structural change models: a simulation analysis," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, D. Corbea, S. Durlauf and B. E. Hansen (eds.), Cambridge University Press, 2006, 212-237.
- Blanchard, O., and Simon, J. (2001): "The long and large decline in U.S. output volatility," *Brookings Papers on Economic Activity*, 1, 135-173.
- Brown, R.L., Durbin, J., and Evans, J.M. (1975): "Techniques for testing the constancy of regression relationships over time," *Journal of the Royal Statistical Society B*, 37, 149-163.
- Chen, J., and Gupta, A.K. (1997): "Testing and locating variance changepoints with application to stock prices," *Journal of the American Statistical Association*, 92, 739-747.
- Davis, R.A., Huang, D., and Yao, Y.C. (1995): "Testing for a change in the parameter values and order of an autoregressive model," *Annals of Statistics*, 23, 282-304.

- Deng, A., and Perron, P. (2006): “The limit distribution of the CUSUM of squares test under general mixing conditions,” forthcoming in *Econometric Theory*.
- Goldfeld, S.M., and Quandt, R.E. (1978): “Asymptotic tests for the constancy of regressions in the heteroskedastic case,” Econometric Research Program Research Memorandum No. 229, Princeton University.
- Hansen, B.E. (2000): “Testing for structural change in conditional models,” *Journal of Econometrics*, 97, 93-115.
- Hawkins, D.M. (1976): “Point estimation of the parameters of piecewise regression models,” *Applied Statistics*, 25, 51-57.
- Herrera, A.M., and Pesavento, E. (2005): “The decline in U.S. output volatility: structural changes and inventory investment,” *Journal of Business & Economic Statistics*, 23, 462-472.
- Horváth, L. (1993): “The maximum likelihood method for testing changes in the parameters of normal observations,” *Annals of Statistics*, 21, 671-680.
- Inclán, C., and Tiao, G.C. (1994): “Use of cumulative sums of squares for retrospective detection of changes of variance,” *Journal of the American Statistical Association*, 89, 913-923.
- Kejriwal, M. (2007): “Tests for a mean shift with good size and monotonic power,” Unpublished Manuscript, Krannert School of Management, Purdue University.
- Kejriwal, M, and Perron, P. (2006a): “Testing for multiple structural changes in cointegrated regression models,” Unpublished Manuscript, Department of Economics, Boston University.
- Kejriwal, M, and Perron, P. (2006b): “The limit distribution of the estimates in cointegrated regression models with multiple structural changes,” Unpublished Manuscript, Department of Economics, Boston University.
- Kim, C.-J., and Nelson, C.R. (1999): “Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle,” *Review of Economics and Statistics*, 81, 608-616.
- Liu, J., Wu, S., and Zidek, J.V. (1997): “On segmented multivariate regressions,” *Statistica Sinica*, 7, 497-525.
- Lumsdaine, R.L., and Ng, S. (1999): “Testing for ARCH in the presence of a possibly misspecified conditional mean,” *Journal of Econometrics*, 93, 257-279.
- McConnell, M.M., and Perez-Quiros, G. (2000): “Output fluctuations in the United States: What has changed since the early 1980’s?” *American Economic Review*, 90, 1464-1476.

- 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, 278-352.
- Perron, P., and Qu, Z. (2006): “Estimating restricted structural change models,” *Journal of Econometrics* 134 (2006), 373-399.
- Pitarakis, J.Y. (2004): “Least squares estimation and tests of breaks in mean and variance under misspecification,” *Econometrics Journal*, 7, 32-54.
- Qu, Z., and Perron, P. (2007a): “Estimating and testing multiple structural changes in multivariate regressions,” *Econometrica*, 75, 459-502.
- Qu, Z., and Perron, P. (2007b): “Supplemental material to: Estimating and testing multiple structural changes in multivariate regressions,” (<http://www.econometricsociety.org/ecta/supmat/ECTA5733SUPP.pdf>).
- Sansó, A., Arago, V., and Carion-i-Silvestre, J.L. (2004), “Testing for changes in the unconditional variance of financial time series,” *Revista de Economia Financiera*, 4, 32-53.
- Sensier, M., and van Dijk, D. (2004): “Testing for volatility changes in U.S. macroeconomic time series,” *Review of Economics and Statistics*, 86, 833-839.
- Stock, J.H., and Watson, M.W. (2002): “Has the business cycle changed and why?” in *NBER Macroeconomics Annual* 17, M. Gertler & K. Rogoff (eds.), MIT press, 159-218.
- Stock, J.H., and Watson, M.W. (2003): “Has the business cycle changed? Evidence and explanations,” in *Monetary Policy and Uncertainty: Adapting to a Changing Economy*, Proceedings of a symposium sponsored by Federal Reserve Bank of Kansas, 9-56.
- Summers, P. M. (2005), “What caused the great moderation? Some cross-country evidence”, *Economic Review* QII, Federal Reserve Bank of Kansas, 5-32.

Appendix

Proof of Theorem 1: Part (a) follows from Theorem 5 of Perron and Qu (2007). For part (b),

$$\begin{aligned}
& \sup LR_{2,T}(m_a, n_a, \varepsilon | n = 0, m_a) \\
&= 2 \left[\log \hat{L}_T \left(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v \right) - \log \tilde{L}_T \right] = T \log \tilde{\sigma}^2 - \sum_{i=1}^{n_a+1} \left(\tilde{T}_i^v - \tilde{T}_{i-1}^v \right) \log \hat{\sigma}_i^2 \\
&= \sum_{i=1}^{n_a} \left[\tilde{T}_{i+1}^v \log \tilde{\sigma}_{1,i+1}^2 - \tilde{T}_i^v \log \tilde{\sigma}_{1,i}^2 - \left(\tilde{T}_{i+1}^v - \tilde{T}_i^v \right) \log \hat{\sigma}_{i+1}^2 \right] + \tilde{T}_1^v \left(\log \tilde{\sigma}_{1,1}^2 - \log \hat{\sigma}_1^2 \right),
\end{aligned}$$

where

$$\tilde{\sigma}_{1,i}^2 = \frac{1}{\tilde{T}_i^v} \sum_{t=1}^{\tilde{T}_i^v} (y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta}_{t,j})^2.$$

Applying a Taylor expansion to $\log \tilde{\sigma}_{1,i+1}^2$, $\log \tilde{\sigma}_{1,i}^2$ and $\log \hat{\sigma}_{i+1}^2$, we obtain after some algebra,

$$\sup LR_{2,T}(m_a, n_a, \varepsilon | n = 0, m_a) = \sum_{i=1}^{n_a} (F_{1,T}^i + F_{2,T}^i) + o_p(1),$$

where

$$\begin{aligned}
& F_{1,T}^i \\
&= \frac{1}{\sigma_0^2} \left[\tilde{T}_{i+1}^v \tilde{\sigma}_{1,i+1}^2 - \tilde{T}_i^v \tilde{\sigma}_{1,i}^2 - \left(\tilde{T}_{i+1}^v - \tilde{T}_i^v \right) \hat{\sigma}_{i+1}^2 \right] \\
&= \frac{1}{\sigma_0^2} \sum_{t=\tilde{T}_i^v+1}^{\tilde{T}_{i+1}^v} \left[(y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta}_{t,j})^2 - (y_t - x_t' \hat{\beta} - z_t' \hat{\delta}_{t,j})^2 \right]
\end{aligned}$$

and

$$F_{2,T}^i = -\frac{1}{2} \left[\tilde{T}_{i+1}^v \left(\frac{\tilde{\sigma}_{1,i+1}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - \tilde{T}_i^v \left(\frac{\tilde{\sigma}_{1,i}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - \left(\tilde{T}_{i+1}^v - \tilde{T}_i^v \right) \left(\frac{\hat{\sigma}_{i+1}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 \right].$$

Now we show that $F_{1,T}^i = o_p(1)$. We can express $F_{1,T}^i$ as

$$\begin{aligned}
& F_{1,T}^i \\
&= \frac{1}{\sigma_0^2} \left[\begin{array}{l} (U_{i+1} + X_{i+1}(\beta - \tilde{\beta}) + Z_{i+1}(\delta - \tilde{\delta}_{t,j}))'(U_{i+1} + X_{i+1}(\beta - \tilde{\beta}) + Z_{i+1}(\delta - \tilde{\delta}_{t,j})) \\ -(U_{i+1} + X_{i+1}(\beta - \hat{\beta}) + Z_{i+1}(\delta - \hat{\delta}_{t,j}))'(U_{i+1} + X_{i+1}(\beta - \hat{\beta}) + Z_{i+1}(\delta - \hat{\delta}_{t,j})) \end{array} \right] \\
&= \frac{1}{\sigma_0^2} \left[\begin{array}{l} (\hat{\beta} - \tilde{\beta})' X'_{i+1} X_{i+1} (\hat{\beta} - \tilde{\beta}) + (\hat{\delta}_{t,j} - \tilde{\delta}_{t,j})' Z'_{i+1} Z_{i+1} (\hat{\delta}_{t,j} - \tilde{\delta}_{t,j}) \\ + (\hat{\beta} - \tilde{\beta})' X'_{i+1} Z_{i+1} (\hat{\delta}_{t,j} - \tilde{\delta}_{t,j}) + 2(\beta - \hat{\beta})' X'_{i+1} X_{i+1} (\hat{\beta} - \tilde{\beta}) \\ + 2(\delta - \hat{\delta}_{t,j})' Z'_{i+1} Z_{i+1} (\hat{\delta}_{t,j} - \tilde{\delta}_{t,j}) + 2(\hat{\beta} - \tilde{\beta})' X'_{i+1} Z_{i+1} (\delta - \hat{\delta}_{t,j}) \\ + 2(\beta - \hat{\beta})' X'_{i+1} Z_{i+1} (\hat{\delta}_{t,j} - \tilde{\delta}_{t,j}) + 2(\hat{\beta} - \tilde{\beta})' X'_{i+1} U_{i+1} + 2(\hat{\delta}_{t,j} - \tilde{\delta}_{t,j})' Z'_{i+1} U_{i+1} \end{array} \right].
\end{aligned}$$

The result follows using the facts that $X'_{i+1} X_{i+1} = O_p(T)$, $Z'_{i+1} Z_{i+1} = O_p(T)$, $X'_{i+1} Z_{i+1} = O_p(T)$, $X'_{i+1} U_{i+1} = O_p(T^{1/2})$, and $Z'_{i+1} U_{i+1} = O_p(T^{1/2})$. Also, since, under the null hypothesis, with A1 the estimates of the break fractions converge to the true break fractions at a fast enough rate so that estimates of the parameters of the models are consistent and have the same limit distribution as when the break dates are known, we have: $\beta - \hat{\beta} = O_p(T^{-1/2})$, $\delta - \hat{\delta} = O_p(T^{-1/2})$, $\hat{\beta} - \tilde{\beta} = o_p(T^{-1/2})$, and $\hat{\delta} - \tilde{\delta} = o_p(T^{-1/2})$. The last two quantities are $o_p(T^{-1/2})$ since $\sqrt{T}(\hat{\beta} - \beta)$ and $\sqrt{T}(\hat{\delta} - \delta)$ have the same limit distribution under the null hypothesis; and likewise for $\sqrt{T}(\hat{\delta}_{t,j} - \delta)$ and $\sqrt{T}(\tilde{\delta}_{t,j} - \delta)$. Part (b) of Theorem 1 follows since by Lemma S.1 in Qu and Perron (2007b), we have

$$F_{2,T}^i \Rightarrow \frac{\psi(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{2 \lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}.$$

The proof of part (c) is similar. We have,

$$\begin{aligned}
\sup LR_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a) &= 2 \left[\log \hat{L}_T(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v) - \log \tilde{L}_T(\hat{T}_1^v, \dots, \hat{T}_{n_a}^v) \right] \\
&= \sum_{i=1}^{n_a+1} \left(\hat{T}_i^v - \hat{T}_{i-1}^v \right) \log \tilde{\sigma}_i^2 - \sum_{i=1}^{n_a+1} \left(\tilde{T}_i^v - \tilde{T}_{i-1}^v \right) \log \hat{\sigma}_i^2,
\end{aligned}$$

where

$$\tilde{\sigma}_i^2 = \frac{1}{\hat{T}_i^v - \hat{T}_{i-1}^v} \sum_{t=\hat{T}_{i-1}^v+1}^{\hat{T}_i^v} (y_t - x'_t \tilde{\beta} - z'_t \tilde{\delta})^2$$

and

$$\hat{\sigma}_i^2 = \frac{1}{\tilde{T}_i^v - \tilde{T}_{i-1}^v} \sum_{t=\tilde{T}_{i-1}^v+1}^{\tilde{T}_i^v} (y_t - x'_t \hat{\beta} - z'_t \hat{\delta}_{t,j})^2.$$

Applying a Taylor extension on $\log \tilde{\sigma}_i^2$ and $\log \hat{\sigma}_i^2$ around σ_0^2 , we have

$$\begin{aligned} & \sup LR_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a) \\ &= \frac{1}{\sigma_0^2} \sum_{i=1}^{n_a+1} \left[\left(\hat{T}_i^v - \hat{T}_{i-1}^v \right) \tilde{\sigma}_i^2 - \left(\tilde{T}_i^v - \tilde{T}_{i-1}^v \right) \hat{\sigma}_i^2 \right] \\ & \quad - \frac{1}{2} \sum_{i=1}^{n_a+1} \left[\left(\hat{T}_i^v - \hat{T}_{i-1}^v \right) \left(\frac{\tilde{\sigma}_i^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - \left(\tilde{T}_i^v - \tilde{T}_{i-1}^v \right) \left(\frac{\hat{\sigma}_i^2 - \sigma_0^2}{\sigma_0^2} \right)^2 \right] + o_p(1). \end{aligned}$$

Using arguments similar to those in part (b), under A1 and A2 it can be shown that the second term is $o_p(1)$. For the first term,

$$\frac{1}{\sigma_0^2} \sum_{i=1}^{n_a+1} \left[\left(\hat{T}_i^v - \hat{T}_{i-1}^v \right) \tilde{\sigma}_i^2 - \left(\tilde{T}_i^v - \tilde{T}_{i-1}^v \right) \hat{\sigma}_i^2 \right] = \frac{1}{\sigma_0^2} \sum_{j=1}^{m_a} \left[\tilde{T}_j^c \tilde{\sigma}_{1,j+1}^2 - \tilde{T}_j^c \tilde{\sigma}_{1,j}^2 - \left(\tilde{T}_{j+1}^c - \tilde{T}_j^c \right) \hat{\sigma}_{j+1}^2 \right],$$

where

$$\tilde{\sigma}_{1,j}^2 = \frac{1}{\tilde{T}_j^c} \sum_{t=1}^{\tilde{T}_j^c} \left(y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta} \right)^2.$$

Following the proof of Theorem 5 in Qu and Perron (2007b), we have

$$\frac{1}{\sigma_0^2} \sum_{j=1}^{m_a} \left[\tilde{T}_j^c \tilde{\sigma}_{1,j+1}^2 - \tilde{T}_j^c \tilde{\sigma}_{1,j}^2 - \left(\tilde{T}_{j+1}^c - \tilde{T}_j^c \right) \hat{\sigma}_{j+1}^2 \right] \Rightarrow \sum_{j=1}^{m_a} \frac{\| \lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c) \|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)}$$

For part (d)

$$\begin{aligned} & \sup LR_{4,T}(m_a, n_a, \varepsilon | n = m = 0) \\ &= 2 \left[\sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \log \hat{L}_T(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v) - \log \tilde{L}_T \right] \\ &= 2 \left[\log \hat{L}_T(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v) - \log \tilde{L}_T \right] \\ &= T \log \tilde{\sigma}^2 - \sum_{i=1}^{n_a+1} \left(\tilde{T}_i^v - \tilde{T}_{i-1}^v \right) \log \hat{\sigma}_i^2 \\ &= \sum_{i=1}^{n_a} \left[\tilde{T}_{i+1}^v \log \tilde{\sigma}_{1,i+1}^2 - \tilde{T}_i^v \log \tilde{\sigma}_{1,i}^2 - \left(\tilde{T}_{i+1}^v - \tilde{T}_i^v \right) \log \hat{\sigma}_{i+1}^2 \right] + \tilde{T}_1^v \left(\log \tilde{\sigma}_{1,1}^2 - \log \hat{\sigma}_1^2 \right), \end{aligned}$$

where

$$\tilde{\sigma}_{1,i}^2 = \frac{1}{\tilde{T}_i^v} \sum_{t=1}^{\tilde{T}_i^v} \left(y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta} \right)^2.$$

Applying a Taylor expansion to $\log \tilde{\sigma}_{1,i+1}^2$, $\log \tilde{\sigma}_{1,i}^2$ and $\log \hat{\sigma}_{i+1}^2$, we obtain after some algebra,

$$\sup LR_{4,T}(m_a, n_a, \varepsilon | n = 0, m_a) = \sum_{i=1}^{n_a} (F_{1,T}^i + F_{2,T}^i) + o_p(1),$$

where

$$\begin{aligned} & \sum_{i=1}^{n_a} F_{1,T}^i \\ &= \sum_{i=1}^{n_a} \frac{1}{\sigma_0^2} \left[\tilde{T}_{i+1}^v \tilde{\sigma}_{1,i+1}^2 - \tilde{T}_i^v \tilde{\sigma}_{1,i}^2 - (\tilde{T}_{i+1}^v - \tilde{T}_i^v) \hat{\sigma}_{i+1}^2 \right] \\ &= \frac{1}{\sigma_0^2} \sum_{j=1}^{m_a} \left[\tilde{T}_j^c \tilde{\sigma}_{1,j+1}^2 - \tilde{T}_j^c \tilde{\sigma}_{1,j}^2 - (\tilde{T}_{j+1}^c - \tilde{T}_j^c) \hat{\sigma}_{j+1}^2 \right], \end{aligned}$$

$$\tilde{\sigma}_{1,j}^2 = \frac{1}{\tilde{T}_j^c} \sum_{t=1}^{\tilde{T}_j^c} \left(y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta} \right)^2$$

and

$$F_{2,T}^i = -\frac{1}{2} \left[\tilde{T}_{i+1}^v \left(\frac{\tilde{\sigma}_{1,i+1}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - \tilde{T}_i^v \left(\frac{\tilde{\sigma}_{1,i}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - (\tilde{T}_{i+1}^v - \tilde{T}_i^v) \left(\frac{\hat{\sigma}_{i+1}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 \right].$$

From the proof of part (c), we have

$$\sum_{i=1}^{n_a} F_{1,T}^i \Rightarrow \sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)},$$

and from that of part (b),

$$F_{2,T}^i \Rightarrow \frac{\psi \left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v) \right)^2}{2 \lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)},$$

under assumptions A2 and A4. Hence, we obtain

$$\sup LR_{4,T}(m_a, n_a, \varepsilon | n = m = 0) \Rightarrow \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \left[\begin{aligned} & \sum_{j=1}^{m_a} \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ & + \frac{\psi}{2} \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v) \right)^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{aligned} \right].$$

Table 1: Asymptotic critical values of the sup $LR_{4,T}^*$ test (the entries are quantiles x such that $P(\text{sup } LR_4^* \leq x) = \alpha$)

		$\varepsilon = 0.10$				$\varepsilon = 0.15$				$\varepsilon = 0.20$		$\varepsilon = 0.25$	$UDmaxLR_4^*$			
		$n_a = 1$		$n_a = 2$		$n_a = 1$		$n_a = 2$		$n_a = 1$		$n_a = 1$	$M = N = 2$			
q	α	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 1$	$\varepsilon = 0.10$	$\varepsilon = 0.15$	$\varepsilon = 0.20$
1	.90	12.11	17.58	18.21	22.92	11.53	15.91	16.15	20.04	10.83	14.14	14.55	10.08	22.92	20.04	17.14
	.95	14.09	19.91	20.49	25.37	13.54	18.33	18.38	22.54	12.94	16.54	16.44	12.04	25.37	22.54	19.47
	.975	16.59	22.06	22.82	27.66	15.47	20.50	20.88	24.79	14.82	18.54	18.80	13.61	27.66	24.79	22.26
	.99	19.97	24.79	25.48	30.12	18.86	22.83	23.96	27.40	17.21	21.60	22.23	16.45	30.12	27.40	24.41
2	.90	14.94	22.52	20.44	27.36	14.05	20.71	18.56	24.59	13.07	18.32	16.34	12.19	27.36	24.59	20.83
	.95	17.34	24.72	22.72	29.90	16.56	23.42	20.59	27.06	15.24	21.17	18.81	14.02	29.89	27.06	23.29
	.975	19.04	26.84	25.06	32.91	18.41	25.59	22.66	29.55	17.62	23.42	20.65	15.84	32.91	29.55	26.15
	.99	20.82	29.94	27.49	34.72	19.99	27.87	25.13	32.06	18.99	26.15	23.39	18.33	34.72	32.06	29.34
3	.90	16.76	26.61	22.62	31.99	16.10	24.40	20.98	28.35	15.35	22.55	18.97	14.60	31.99	28.35	25.17
	.95	18.79	28.99	25.14	34.36	17.96	27.03	23.01	31.19	17.17	25.01	20.75	16.26	34.36	31.19	27.32
	.975	20.36	30.63	26.96	36.29	19.73	29.61	25.01	33.42	18.67	26.81	22.39	17.86	36.29	33.42	29.33
	.99	22.28	33.93	29.51	39.14	21.99	31.31	27.81	36.13	20.24	29.07	24.60	19.87	39.14	36.13	31.96
4	.90	19.31	30.63	25.07	36.07	18.31	28.14	22.66	31.94	17.37	26.05	20.57	16.41	36.07	31.94	28.33
	.95	21.54	33.71	27.34	38.91	20.49	30.84	24.81	34.34	19.43	28.48	22.84	18.75	38.91	34.34	30.91
	.975	23.81	36.50	29.78	41.25	22.52	33.50	26.84	37.31	21.54	30.88	25.11	20.55	41.25	37.01	33.20
	.99	26.37	39.79	31.87	44.50	24.84	37.10	29.50	41.07	24.31	34.31	26.88	22.69	44.51	41.07	36.08
5	.90	21.35	34.69	26.76	39.78	20.22	32.18	24.40	35.76	19.37	29.92	22.09	18.06	39.78	35.76	31.82
	.95	23.74	37.53	29.34	43.03	22.38	34.62	26.44	38.19	21.57	32.15	24.42	20.36	43.03	38.19	34.21
	.975	26.51	39.75	32.16	45.89	24.32	37.32	28.98	41.09	23.53	34.38	26.70	22.54	45.89	41.09	36.71
	.99	29.23	43.38	35.04	49.63	28.65	40.70	32.53	44.94	27.10	38.11	30.04	24.86	49.63	44.94	40.19

Table 2: Size of the sup $LR_{4,T}^*$ using different estimates of ψ in the case of ARCH(1) errors
(DGP: $y_t = e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim i.i.d. N(0, 1)$, $h_t = \delta_1 + \gamma e_{t-1}^2$, $h_0 = \delta_1 / (1 - \gamma)$,
 $\delta_1 = 1$, $T = 100$, $\varepsilon = 0.20$, Alternative hypothesis: $m = 1, n = 1$).

γ	alternative	null	hybrid
0.1	0.083	0.040	0.042
0.2	0.102	0.042	0.049
0.3	0.116	0.038	0.048
0.4	0.139	0.031	0.040
0.5	0.161	0.027	0.042

Note: "alternative" specifies that the unrestricted residuals are used to construct $\hat{\psi}$ and m ; "null" specifies that the residuals imposing the null hypothesis are used to construct $\hat{\psi}$ and m , and "hybrid" specifies that the residuals under the alternative are used to construct m and the residuals under the null hypothesis are used to construct $\hat{\psi}$.

Table 3: Power of the sup $LR_{4,T}^*$ using different estimates of ψ in the case of ARCH(1) errors
(DGP: $y_t = \mu_1 + \mu_2 1(t > [.25T]) + e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim i.i.d. N(0, 1)$,
 $h_t = \delta_1 + \delta_2 1(t > [.75T]) + \gamma e_{t-1}^2$, $h_0 = \delta_1 / (1 - \gamma)$, $\delta_1 = 1$, $T = 100$, $\varepsilon = 0.20$).

a) small change in variance, large change in mean

	$\gamma = 0.1$				$\gamma = 0.3$				$\gamma = 0.5$			
	null		hybrid		null		hybrid		null		hybrid	
$\mu_2 \setminus \delta_2$	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5
0.5	0.267	0.299	0.281	0.318	0.222	0.231	0.230	0.250	0.161	0.169	0.169	0.181
1	0.859	0.859	0.863	0.862	0.758	0.752	0.762	0.760	0.612	0.616	0.619	0.631
1.5	0.999	0.998	0.999	0.998	0.986	0.986	0.987	0.986	0.930	0.929	0.932	0.932
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.993	0.992	0.993	0.992

b) small change in mean, large change in variance

	$\gamma = 0.1$				$\gamma = 0.3$				$\gamma = 0.5$			
	null		hybrid		null		hybrid		null		hybrid	
$\delta_2 \setminus \mu_2$	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5
1	0.202	0.394	0.246	0.429	0.142	0.293	0.173	0.324	0.094	0.216	0.121	0.231
3	0.512	0.682	0.655	0.771	0.332	0.483	0.438	0.569	0.210	0.346	0.277	0.398
5	0.652	0.805	0.822	0.903	0.464	0.592	0.600	0.715	0.299	0.422	0.406	0.495
7	0.731	0.853	0.887	0.945	0.532	0.671	0.693	0.791	0.360	0.477	0.493	0.574

Note: "null" specifies that the residuals imposing the null hypothesis are used to construct $\hat{\psi}$ and m , and "hybrid" specifies that the residuals under the alternative are used to construct m and the residuals under the null hypothesis are used to construct $\hat{\psi}$.

Table 4: Power of the sup $LR_{4,T}^*$ test using different corrections in the case of Normal errors (DGP: $y_t = \mu_1 + \mu_2 1(t > T_1^c) + e_t$; $e_t \sim i.i.d. N(0, 1 + \delta 1(t > T_1^v))$, $\mu_1 = 0$, $\mu_2 = \delta$, $\varepsilon = 0.15$)

T = 100									
	$T_1^c = T_1^v = [.5T]$			$T_1^c = [.3T], T_1^v = [.6T]$			$T_1^c = [.6T], T_1^v = [.3T]$		
δ	full	i.i.d.	NC	full	i.i.d.	NC	full	i.i.d.	NC
0.25	0.125	0.126	0.125	0.112	0.115	0.123	0.120	0.112	0.121
0.5	0.425	0.439	0.455	0.406	0.401	0.414	0.382	0.377	0.396
0.75	0.780	0.779	0.783	0.750	0.752	0.753	0.685	0.686	0.703
1	0.946	0.947	0.953	0.949	0.948	0.952	0.889	0.890	0.898
1.25	0.992	0.992	0.993	0.991	0.992	0.991	0.978	0.978	0.982
1.5	0.998	0.998	0.999	0.999	0.999	0.999	0.995	0.995	0.995
T = 200									
	$T_1^c = T_1^v = [.5T]$			$T_1^c = [.3T], T_1^v = [.6T]$			$T_1^c = [.6T], T_1^v = [.3T]$		
δ	full	i.i.d.	NC	full	i.i.d.	NC	full	i.i.d.	NC
0.25	0.228	0.224	0.239	0.213	0.210	0.211	0.206	0.207	0.210
0.5	0.783	0.788	0.779	0.745	0.748	0.732	0.709	0.711	0.719
0.75	0.981	0.982	0.993	0.982	0.982	0.985	0.955	0.953	0.960
1	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.997	0.998
1.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: The nominal size is 5% and 1,000 replications are used. The column "full" refers to the tests using the correction $\hat{\psi}$ which allows for non-Normal, conditionally heteroskedastic and serially correlated errors, as defined by (11); the column "i.i.d." refers to a correction that only allows for i.i.d. non-Normal errors, i.e., $\hat{\psi} = \hat{\mu}_4 / \hat{\sigma}^4 - 1$, where $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$ and $\hat{\mu}_4 = T^{-1} \sum_{t=1}^T \hat{u}_t^4$ with \hat{u}_t the residuals under the null hypotheses; the column "NC" applies no correction and sets $\hat{\psi} = 2$, which is valid with Normal errors.

Table 5: Size and Power of the sup $LR_{1,T}^*$ test using different corrections in the case of Normal errors (DGP: $y_t = e_t$; $e_t \sim i.i.d. N(0, 1 + \delta 1(t > [.5T]))$, $\varepsilon = 0.25$)

	T = 100			T = 200		
δ	full	i.i.d.	NC	full	i.i.d.	NC
0	0.046	0.039	0.038	0.041	0.044	0.049
0.25	0.053	0.073	0.065	0.129	0.112	0.137
0.5	0.159	0.159	0.190	0.363	0.348	0.383
0.75	0.308	0.297	0.365	0.618	0.598	0.609
1	0.462	0.453	0.533	0.803	0.806	0.848
1.25	0.573	0.603	0.668	0.932	0.908	0.944
1.5	0.761	0.690	0.795	0.969	0.967	0.983

Note: see notes to Table 5.

Table 6: Size and Power of the sup $LR_{1,T}^*(n_a = 1)$, UDmax and $CUSQ^*$ tests in the case of Normal errors (DGP: $y_t = e_t$; $e_t \sim i.i.d. N(0, 1 + \delta 1(t > [.5T]))$, $\varepsilon = 0.25$)

	T = 100			T = 200		
δ	sup $LR_{1,T}^*$	UDmax	$CUSQ^*$	sup $LR_{1,T}^*$	UDmax	$CUSQ^*$
0	0.046	0.041	0.030	0.041	0.041	0.031
0.25	0.053	0.050	0.060	0.129	0.126	0.098
0.5	0.159	0.159	0.134	0.363	0.350	0.345
0.75	0.308	0.300	0.291	0.618	0.607	0.595
1	0.462	0.448	0.427	0.803	0.796	0.805
1.25	0.573	0.558	0.560	0.932	0.928	0.905
1.5	0.761	0.614	0.645	0.969	0.965	0.964

Table 7: Size and Power of the sup $LR_{1,T}^*(n_a = 1, \varepsilon)$, UDmax and CUSQ* tests in a dynamic model with ARCH(1) errors

(DGP: $y_t = c + \alpha y_{t-1} + e_t$, $e_t = u_t \sqrt{h_t}$, with $u_t \sim i.i.d. N(0, 1)$, $h_t = \delta_1 + \delta_2 1(t > [.5T]) + \gamma e_{t-1}^2$, $h_0 = \delta_1 / (1 - \gamma)$, $c = 0.5$, $\delta_1 = 0.1$, $\varepsilon = 0.25$).

$T = 100$																		
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
δ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.056	0.052	0.028	0.054	0.052	0.031	0.050	0.050	0.031	0.052	0.050	0.027	0.050	0.055	0.028	0.052	0.049	0.019
0.05	0.165	0.156	0.134	0.138	0.136	0.083	0.125	0.126	0.054	0.170	0.160	0.139	0.140	0.147	0.082	0.116	0.096	0.059
0.10	0.434	0.417	0.268	0.302	0.293	0.151	0.209	0.201	0.149	0.429	0.415	0.297	0.303	0.283	0.149	0.209	0.209	0.147
0.15	0.620	0.608	0.528	0.452	0.440	0.324	0.318	0.309	0.196	0.623	0.608	0.555	0.452	0.453	0.317	0.306	0.282	0.202
0.20	0.811	0.807	0.649	0.617	0.602	0.413	0.434	0.411	0.315	0.809	0.801	0.678	0.611	0.594	0.399	0.415	0.430	0.319
0.30	0.916	0.911	0.828	0.784	0.775	0.570	0.562	0.552	0.407	0.916	0.909	0.854	0.775	0.727	0.558	0.544	0.542	0.423
$T = 200$																		
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
δ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.043	0.043	0.034	0.052	0.050	0.029	0.038	0.035	0.023	0.044	0.042	0.032	0.043	0.044	0.030	0.035	0.028	0.034
0.05	0.351	0.341	0.313	0.214	0.209	0.190	0.139	0.133	0.114	0.350	0.340	0.309	0.209	0.203	0.192	0.123	0.122	0.120
0.10	0.753	0.749	0.726	0.497	0.485	0.501	0.313	0.300	0.299	0.758	0.749	0.728	0.522	0.506	0.507	0.277	0.312	0.266
0.15	0.927	0.923	0.930	0.740	0.727	0.723	0.512	0.496	0.442	0.929	0.921	0.933	0.729	0.745	0.709	0.448	0.467	0.404
0.20	0.981	0.981	0.984	0.885	0.876	0.837	0.621	0.614	0.616	0.980	0.979	0.982	0.878	0.839	0.825	0.627	0.631	0.576
0.30	0.999	0.999	0.998	0.954	0.950	0.931	0.780	0.773	0.719	0.999	0.998	0.997	0.949	0.940	0.924	0.759	0.749	0.698

Table 8: Size and Power of the sup $LR_{1,T}^*(n_a, \varepsilon)$, UD_{\max} and $CUSQ^*$ tests with Normal errors and two variance breaks
(DGP: $y_t = e_t$; $e_t \sim i.i.d. N(0, 1 + \delta 1([.3T] < t \leq [.6T]))$, $T = 200$)

δ	$\varepsilon = 0.10$			$\varepsilon = 0.15$			$\varepsilon = 0.20$			$\varepsilon = 0.25$			$CUSQ^*$
	$n_a = 1$	$n_a = 2$	UD _{max}	$n_a = 1$	$n_a = 2$	UD _{max}	$n_a = 1$	$n_a = 2$	UD _{max}	$n_a = 1$	$n_a = 2$	UD _{max}	
0	0.033	0.033	0.032	0.036	0.033	0.033	0.040	0.035	0.039	0.041	0.035	0.041	0.026
0.25	0.071	0.057	0.069	0.080	0.062	0.071	0.080	0.076	0.081	0.087	0.091	0.081	0.039
0.5	0.124	0.133	0.117	0.137	0.167	0.142	0.141	0.197	0.154	0.138	0.234	0.150	0.069
0.75	0.174	0.252	0.193	0.182	0.321	0.218	0.184	0.369	0.219	0.198	0.439	0.242	0.089
1	0.202	0.394	0.271	0.241	0.484	0.325	0.266	0.570	0.364	0.287	0.621	0.374	0.118
1.25	0.280	0.546	0.403	0.328	0.631	0.466	0.367	0.704	0.507	0.387	0.774	0.532	0.154
1.5	0.372	0.673	0.521	0.418	0.760	0.586	0.454	0.825	0.629	0.477	0.868	0.660	0.186
2	0.502	0.866	0.721	0.535	0.915	0.783	0.572	0.950	0.819	0.624	0.965	0.837	0.300
2.5	0.592	0.934	0.827	0.675	0.968	0.878	0.714	0.982	0.909	0.750	0.990	0.922	0.348
3	0.681	0.977	0.909	0.749	0.986	0.938	0.780	0.992	0.960	0.823	0.998	0.971	0.397

Note: The columns $n_a = 1$ and $n_a = 2$ correspond to the sup $LR_{1,T}^*(n_a = 1, \varepsilon)$ and sup $LR_{1,T}^*(n_a = 2, \varepsilon)$ tests, respectively.

Table 11: Size of the sup $LR_{4,T}^*(m_a, n_a)$ and UD max tests in the case of Normal errors
(DGP: $y_t = e_t, e_t \sim i.i.d. N(0, 1)$)

T=100				
ε	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax
0.2	0.041	0.050	0.045	0.052
0.15	0.046	0.053	0.043	0.046
0.1	0.057	0.058	0.052	0.054
T=200				
ε	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax
0.2	0.050	0.044	0.052	0.046
0.15	0.054	0.050	0.047	0.047
0.1	0.048	0.040	0.046	0.045

Table 12: Size of sup $LR_{4,T}^*(m_a, n_a)$ and UD max tests in the case of ARCH(1) errors
(DGP: $y_t = e_t, e_t = u_t \sqrt{h_t}$, with $u_t \sim i.i.d. N(0, 1)$, $h_t = \delta_1 + \gamma \varepsilon_{t-1}^2$, $h_0 = \delta_1 / (1 - \gamma)$, $\delta_1 = 1$)

T=100								
$\varepsilon = 0.1$					$\varepsilon = 0.2$			
γ	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax
0.1	0.057	0.058	0.058	0.062	0.042	0.057	0.046	0.056
0.3	0.057	0.067	0.059	0.072	0.048	0.068	0.048	0.070
0.5	0.058	0.063	0.057	0.076	0.042	0.064	0.043	0.056
T=200								
$\varepsilon = 0.1$					$\varepsilon = 0.2$			
γ	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax
0.1	0.058	0.039	0.049	0.053	0.051	0.049	0.051	0.056
0.3	0.054	0.031	0.043	0.050	0.042	0.039	0.040	0.038
0.5	0.044	0.019	0.038	0.044	0.044	0.030	0.037	0.031

Table 13: Size of the sup $LR_{4,T}^*(m_a, n_a)$ and UD max tests for DGPs with one break in coefficients and two breaks in variance

(DGP: $y_t = \mu_1 + \mu_2 1(t > T^c) + e_t$, $e_t \sim i.i.d. N(0, 1 + \delta 1(T_1^v < t \leq T_2^v))$, $\mu_1 = 0, \mu_2 = \delta, \varepsilon = 0.1$)

	$m_a=1$ $n_a=1$	$m_a=1$ $n_a=2$	UDmax	$m_a=1$ $n_a=1$	$m_a=1$ $n_a=2$	UDmax	$m_a=1$ $n_a=1$	$m_a=1$ $n_a=2$	UDmax	$m_a=1$ $n_a=1$	$m_a=1$ $n_a=2$	UDmax	$m_a=1$ $n_a=1$	$m_a=1$ $n_a=2$	UDmax
	$T^c = T_1^v = [.3T], T_2^v = [.6T]$			$T^c = T_2^v = [.6T], T_1^v = [.3T]$			$T^c = [.3T], T_1^v = [.5T], T_2^v = [.6T]$			$T^c = [.5T], T_1^v = [.3T], T_2^v = [.6T]$			$T^c = [.6T], T_1^v = [.3T], T_2^v = [.5T]$		
δ	$T = 100$														
0.25	0.119	0.087	0.083	0.125	0.099	0.094	0.117	0.093	0.086	0.131	0.104	0.093	0.126	0.102	0.096
0.5	0.328	0.283	0.263	0.367	0.317	0.273	0.331	0.268	0.239	0.391	0.316	0.288	0.368	0.306	0.277
0.75	0.667	0.604	0.570	0.715	0.672	0.588	0.610	0.591	0.555	0.734	0.683	0.631	0.726	0.649	0.601
1	0.906	0.891	0.847	0.933	0.917	0.891	0.924	0.883	0.837	0.943	0.927	0.906	0.929	0.916	0.894
1.25	0.984	0.982	0.970	0.994	0.989	0.977	0.985	0.974	0.972	0.994	0.995	0.985	0.995	0.988	0.983
1.5	1.000	0.999	0.998	1.000	1.000	1.000	0.999	0.999	0.998	1.000	1.000	0.999	1.000	1.000	0.999
	$T = 200$														
0.25	0.162	0.131	0.123	0.192	0.164	0.142	0.158	0.128	0.109	0.191	0.166	0.146	0.189	0.161	0.138
0.5	0.610	0.583	0.518	0.686	0.662	0.597	0.598	0.510	0.468	0.698	0.667	0.605	0.666	0.631	0.572
0.75	0.958	0.948	0.921	0.970	0.964	0.946	0.958	0.924	0.889	0.964	0.966	0.944	0.969	0.962	0.947
1	1.000	0.998	0.996	1.000	0.998	0.995	1.000	0.995	0.992	1.000	0.998	0.998	0.999	0.999	0.997
1.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 14: Size of the sup $LR_{4,T}^*(m_a, n_a)$ and UD max tests for DGPs with two breaks in coefficients and one break in variance

(DGP: $y_t = \mu_1 + \mu_2 1(T_1^c < t \leq T_2^c) + e_t$, $e_t \sim i.i.d. N(0, 1 + \delta 1(t > T^v))$, $\mu_1 = 0, \mu_2 = \delta, \varepsilon = 0.1$).

	$m_a=1$ $n_a=1$	$m_a=2$ $n_a=1$	UDmax	$m_a=1$ $n_a=1$	$m_a=2$ $n_a=1$	UDmax	$m_a=1$ $n_a=1$	$m_a=2$ $n_a=1$	UDmax	$m_a=1$ $n_a=1$	$m_a=2$ $n_a=1$	UDmax	$m_a=1$ $n_a=1$	$m_a=2$ $n_a=1$	UDmax
	$T_1^c = T^v = [.3T], T_2^c = [.6T]$			$T_1^c = [.3T], T_2^c = T^v = [.6T]$			$T_1^c = [.5T], T_2^c = [.6T], T^v = [.3T]$			$T_1^c = [.3T], T_2^c = [.6T], T^v = [.5T]$			$T_1^c = [.3T], T_2^c = [.5T], T^v = [.6T]$		
δ	$T = 100$														
0.25	0.086	0.097	0.084	0.087	0.089	0.090	0.073	0.074	0.071	0.089	0.094	0.086	0.081	0.078	0.083
0.5	0.160	0.240	0.201	0.194	0.270	0.248	0.110	0.121	0.098	0.181	0.264	0.235	0.144	0.197	0.194
0.75	0.300	0.480	0.408	0.380	0.569	0.525	0.183	0.225	0.167	0.370	0.554	0.500	0.248	0.453	0.405
1	0.453	0.725	0.660	0.602	0.850	0.827	0.273	0.350	0.272	0.588	0.825	0.797	0.382	0.733	0.694
1.25	0.660	0.877	0.836	0.791	0.973	0.962	0.377	0.502	0.424	0.771	0.962	0.943	0.513	0.912	0.888
1.5	0.796	0.965	0.936	0.919	0.999	0.995	0.482	0.624	0.545	0.880	0.991	0.989	0.623	0.981	0.971
	$T = 200$														
0.25	0.122	0.169	0.131	0.135	0.172	0.147	0.091	0.089	0.080	0.123	0.175	0.144	0.101	0.146	0.125
0.5	0.326	0.512	0.433	0.399	0.574	0.520	0.192	0.234	0.182	0.397	0.560	0.505	0.297	0.460	0.409
0.75	0.636	0.834	0.798	0.745	0.939	0.909	0.392	0.477	0.417	0.733	0.913	0.890	0.540	0.848	0.813
1	0.871	0.978	0.959	0.948	1.000	0.993	0.600	0.724	0.667	0.930	0.998	0.994	0.775	0.990	0.985
1.25	0.967	0.999	0.996	0.992	1.000	1.000	0.775	0.884	0.851	0.986	1.000	1.000	0.911	1.000	1.000
1.5	0.995	1.000	1.000	1.000	1.000	1.000	0.881	0.949	0.950	0.997	1.000	1.000	0.970	1.000	1.000

Table15: Finite sample performance of the specific to general sequential procedure to select the number of breaks in coefficients and variance.

(DGP: $y_t = \mu_1 + \mu_2 1(t > T^c) + e_t$, $e_t \sim i.i.d. N(0, 1 + \delta 1(t > T^v))$, $\varepsilon = 0.15$, $T = 200$).

	$m = n = 0$	$m = n = 1$ $T^c = [0.5T], T^v = [0.7T]$				$m = n = 1$ $T^c = [0.25T], T^v = [0.75T]$		
		$\mu_2 = \delta = 1$	$\mu_2 = 1, \delta = 3$	$\mu_2 = 1, \delta = 5$	$\mu_2 = \delta = 2$	$\mu_2 = \delta = 1$	$\mu_2 = \delta = 2$	$\mu_2 = 1, \delta = 3$
$prob(m = 0, n = 0)$	0.966	0.010	0.001	0.005	0.000	0.019	0.000	0.002
$prob(m = 0, n = 1)$	0.028	0.018	0.167	0.206	0.000	0.021	0.000	0.055
$prob(m = 0, n = 2)$	0.001	0.003	0.007	0.010	0.000	0.003	0.000	0.005
$prob(m = 1, n = 0)$	0.005	0.419	0.010	0.004	0.079	0.612	0.218	0.044
$prob(m = 1, n = 1)$	0.000	0.512	0.778	0.734	0.883	0.329	0.757	0.868
$prob(m = 1, n = 2)$	0.000	0.031	0.035	0.040	0.032	0.013	0.022	0.025
$prob(m = 2, n = 0)$	0.000	0.004	0.000	0.000	0.001	0.003	0.001	0.000
$prob(m = 2, n = 1)$	0.000	0.003	0.002	0.001	0.004	0.000	0.002	0.000
$prob(m = 2, n = 2)$	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
$prob(\bar{K} = 0)$	0.944	0.000	0.000	0.000	0.000	0.004	0.000	0.002
$prob(\bar{K} = 1)$	0.053	0.805	0.299	0.109	0.487	0.681	0.238	0.065
$prob(\bar{K} = 2)$	0.003	0.195	0.701	0.891	0.513	0.315	0.762	0.933
	$m = n = 1$ $T^c = T^v = [0.5T]$		$m = 1, n = 0$ $T^c = [0.5T]$			$m = 0, n = 1$ $T^v = [0.5T]$		
	$\mu_2 = \delta = 1$	$\mu_2 = 1, \delta = 3$	$\mu_2 = 1$	$\mu_2 = 2$	$\mu_2 = 3$	$\delta = 1$	$\delta = 2$	$\delta = 3$
$prob(m = 0, n = 0)$	0.005	0.000	0.000	0.001	0.000	0.354	0.032	0.002
$prob(m = 0, n = 1)$	0.045	0.003	0.000	0.000	0.000	0.628	0.940	0.971
$prob(m = 0, n = 2)$	0.002	0.000	0.001	0.000	0.000	0.009	0.022	0.021
$prob(m = 1, n = 0)$	0.126	0.002	0.934	0.933	0.928	0.002	0.000	0.000
$prob(m = 1, n = 1)$	0.801	0.966	0.044	0.047	0.051	0.005	0.006	0.005
$prob(m = 1, n = 2)$	0.015	0.024	0.011	0.010	0.009	0.001	0.000	0.001
$prob(m = 2, n = 0)$	0.001	0.000	0.007	0.008	0.011	0.001	0.000	0.000
$prob(m = 2, n = 1)$	0.004	0.005	0.002	0.001	0.001	0.000	0.000	0.000
$prob(m = 2, n = 2)$	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$prob(\bar{K} = 0)$	0.000	0.000	0.000	0.000	0.000	0.342	0.030	0.002
$prob(\bar{K} = 1)$	0.946	0.948	0.956	0.946	0.943	0.632	0.919	0.945
$prob(\bar{K} = 2)$	0.054	0.052	0.044	0.054	0.057	0.026	0.051	0.053

Note: $prob(m = j, n = i)$ represents the probability of choosing j breaks in mean and i breaks in variance, and $prob(\bar{K} = j)$ denotes the probability of selecting j total breaks in either mean or variance. The upper bound for the total number of breaks is set to 2.

Table 16: Empirical results for US macroeconomic series: Outcome of the tests.

series	UDmax(2,2)		Model	sup LR* _{4,T}	Sup-SEQ					SupLR* _{2,T}			SupLR _{3,T}		
	$\varepsilon = 0.15$	$\varepsilon = 0.20$			(1,2 1,1)	(2,1 1,1)	(1,3 1,2)	(2,2 1,2)	(2,2 2,1)	(1,1 1,0)	(1,2 1,0)	(2,1 2,0)	(1,1 0,1)	(2,1 0,1)	(1,2 0,2)
GDP	39.03 ^c	33.91 ^c	(1,2)	39.03 ^a	12.02 ^b	6.36	3.66	9.10	9.43 ^c	15.68 ^a	15.65 ^a	11.74 ^b	7.84	8.14	14.19
Consumption	30.75	29.16													
durables	49.65 ^a	46.18 ^a	(1,1)	34.51 ^a	3.97	8.42	2.62	11.54	1.30	29.23 ^a	16.27 ^a	38.67 ^a	8.98	12.31	10.02
nondurables	41.63 ^b	41.63 ^a	(1,1)	30.17 ^a	2.29	11.62	1.05	11.62	2.58	7.44 ^c	3.75	5.57	25.55 ^a	18.55 ^a	14.03
services	45.85 ^a	45.79 ^a	(1,1)	34.31 ^a	1.95	14.83	0.99	15.92	1.47	5.47	3.10	6.43	31.88 ^a	21.70 ^a	30.62 ^a
Investment (total)	25.37	24.27													
fixed investment-total	35.77 ^c	29.10	(1,2)	26.39 ^c	8.55 ^c	9.52	4.05	10.78	8.95 ^c	18.06 ^a	11.80 ^a	17.88 ^a	3.02	4.95	3.71
nonresidential	27.72	28.09													
residential	44.22 ^b	31.82 ^c	(1,1)	28.19 ^a	8.04	16.48	2.45	15.88	6.94	18.56 ^a	11.80 ^a	20.43 ^a	11.97	13.19	11.54
Δ inventory-inv/GDP	31.60	24.92													
Exports	60.22 ^a	61.96 ^a	(2,2)	65.77 ^a	6.97	8.47	4.43	9.91	9.82 ^c	18.13 ^a	14.84 ^a	21.96 ^a	30.71 ^a	19.11 ^a	13.14
Imports	56.64 ^a	57.79 ^a	(2,2)	51.39 ^a	2.70	14.32	3.09	13.71	1.65	16.04 ^a	12.58 ^a	12.47 ^a	28.96 ^a	16.81 ^b	30.17 ^a
Government spending	31.89	28.66													
Production															
goods (total)	25.94	26.04													
nondurable	31.57	30.66													
durable	34.68	34.99 ^b	(1,2)	26.22 ^c	10.43 ^b	8.63	4.06	11.14	6.18	10.82 ^b	9.77 ^a	11.25 ^b	4.35	12.09	8.70
services	40.55 ^b	36.23 ^b	(2,1)	41.36 ^a	1.79	8.63	2.24	10.69	0.20	2.12	1.48	3.24	26.64 ^a	19.71 ^a	27.71 ^a
structures	35.05	33.44 ^c	(2,2)	41.87 ^b	8.05	7.59	2.79	9.52	7.89	20.35 ^a	12.54 ^a	20.14 ^a	12.95	9.36	10.82
Employment	31.90	26.96													
Price inflation	47.82 ^a	47.82 ^a	(2,2)	47.82 ^a	11.97 ^b	8.47	4.11	8.47	9.53 ^c	22.68 ^a	15.42 ^a	18.06 ^a	6.73	8.35	12.65
90-day T-bill rate	43.82 ^b	38.07 ^b	(2,2)	43.82 ^b	8.22	13.75	8.00	16.89	8.14	9.81 ^b	7.24 ^b	7.41	16.41 ^c	14.19	12.95
10-year T-bond rate	42.84 ^b	44.17 ^a	(2,2)	43.39 ^b	19.26 ^a	8.02	9.98 ^c	8.02	18.18 ^a	17.32 ^a	13.87 ^a	15.11 ^a	7.52	5.24	12.04

Note: The test results are based on an AR(4) model. The subscripts a,b, and c indicate a statistic significant at the 1%, 5% and 10% significance level, respectively. For the SupLR* tests, the trimming parameter is $\varepsilon = 0.15$. The first 19 series are annual growth rates (i.e., $100\ln(x_t/x_{t-4})$), except for the change in inventory investment, which is the annual difference of the quarterly change in inventories as a fraction of GDP. Inflation is the four-quarter change in the annual inflation rate (i.e., $100[\ln(P_t/P_{t-1}) - \ln(P_{t-4}/P_{t-5})]$), with P_t the GDP deflator and the two interest rates series are in four-quarter changes (i.e., $x_t - x_{t-4}$).

Table 17: Empirical results for US macroeconomic series: Parameter estimates.

series	SW (2002)																	
	T^c	T^v	T_1^c	T_2^c	T_1^v	T_2^v	α_1	α_2	α_3	β_1	β_2	β_3	σ_1	σ_2	σ_3	$\frac{\sigma_2}{\sigma_1}$	$\frac{\sigma_3}{\sigma_2}$	$\frac{\sigma_3}{\sigma_1}$
GDP	.	1983:2	1968:2		1975:4	1983:1	0.019	0.013		0.604	0.722		0.009	0.017	0.006	1.89	0.35	0.67
Consumption	.	1992:1																
durables	1987:3	1987:3	1991:1		1991:1		0.017	0.02		0.647	0.70		0.046	0.018		0.39		
nondurables	1991:4	.	1992:1		1982:1		0.008	0.005		0.692	0.826		0.010	0.006		0.6		
services	1969:4	.	1970:1		1977:4		0.02	0.008		0.585	0.758		0.004	0.006		1.5		
Investment (total)	.	.																
fixed investment-total	.	1983:3	1983:3		1974:3	1983:3	0.013	0.008		0.726	0.821		0.025	0.039	0.018	1.53	0.48	0.73
nonresidential	.	.																
residential	.	1983:2	1991:1		1983:4		-0.002	0.019		0.642	0.812		0.075	0.028		0.37		
Δ inventory-inv/GDP	.	1988:1																
Exports	.	1975:4	1972:4	1992:1	1983:2	1992:1	0.071	0.021	0.041	-0.169	0.730	0.342	0.047	0.020	0.028	0.43	1.4	0.6
Imports	1972:4	1986:2	1967:1	1990:4	1985:2	1995:3	0.037	0.097	0.016	0.470	-0.247	0.747	0.048	0.020	0.027	0.42	1.35	0.56
Government spending													
Production																		
goods (total)	.	1983:4	.		.													
nondurable	.	1983:4																
durable	.	1985:2	1993:4		1973:3	1982:3	0.01	0.027		0.556	-0.079		0.017	0.031	0.012	1.82	0.39	0.71
services	1968:3	.	1968:3	1982:4	1995:3		0.032	0.006	0.005	0.334	0.789	0.839	0.004	0.007		1.75		
structures	1991:3	1984:2	1973:3	1991:3	1973:3	1983:3	0.016	-0.001	0.018	0.539	0.846	0.467	0.026	0.044	0.018	1.69	0.41	0.69
Employment	1981:2	1983:2																
Price inflation	1973:2	.	1971:3	1986:1	1971:3	1986:1	0.001	-0.000	0.000	-0.394	0.561	0.123	0.002	0.004	0.002	2	0.5	1
90-day T-bill rate	1981:1	1984:4	1967:4	1983:4	1967:4	1983:4	0.089	0.167	-0.039	0.663	0.566	0.766	0.223	1.317	0.593	5.91	0.45	2.66
10-year T-bond rate	1981:1	1979:3	1979:3	1986:4	1979:3	1986:4	0.045	0.116	-0.111	0.423	0.646	0.421	0.303	1.072	0.487	3.54	0.45	1.61

Note: The estimation results are based on an AR(4) model and the trimming parameter is $\varepsilon = 0.15$. The first 19 series are annual growth rates (i.e., $100\ln(x_t/x_{t-4})$), except for the change in inventory investment, which is the annual difference of the quarterly change in inventories as a fraction of GDP. Inflation is the four-quarter change in the annual inflation rate (i.e.,

$$100[\ln(P_t/P_{t-1}) - \ln(P_{t-4}/P_{t-5})], \text{ with } P_t \text{ the GDP deflator and the two interest rates series are in four-quarter changes (i.e., } x_t - x_{t-4}).$$

Table 18: Empirical results for series with variance break only: tests and parameter estimates.

	Consumption	Δ inventory-inv/GDP	Production	
			goods (total)	nondurable
Tests				
UDmaxLR _{1,T}	8.75 ^c	12.30 ^b	7.90	16.02 ^a
Model Selected	(0,1)	(0,2)	.	(0,1)
SupLR _{1,T} [*]	8.75 ^c	9.97 ^a	.	16.02 ^a
SupLR _{3,T}				
(1,1 0,1)	11.27	7.32	.	10.68
(2,1 0,1)	11.17	4.86	.	7.80
(1,2 0,2)	12.11	13.14	.	7.79
Sup-SEQ				
(0,2 0,1)	2.60	7.73	.	0.67
(0,3 0,2)	1.77	0.32	.	0.65
Estimates				
T_1^v	1983:2	1973:3	.	1984:3
T_2^v	.	1987:2	.	.
α	0.008	0.000	.	0.020
β	0.763	0.009	.	0.638
σ_1	0.010	0.005	.	0.050
σ_2	0.006	0.008	.	0.023
σ_3	.	0.004	.	.
(σ_2/σ_1)	0.6	1.6	.	0.46
(σ_3/σ_2)	.	0.5	.	.
(σ_3/σ_1)	.	0.8	.	.
Stock_Waton (2002)				
T^v	1992:1	1988:1	1983:4	1983:4

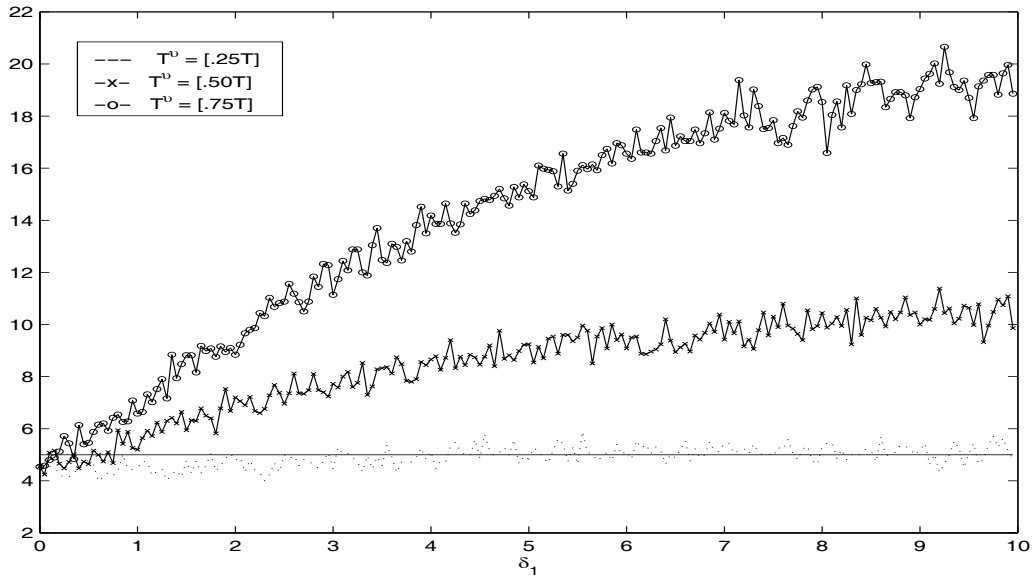


Figure 1: Size of the Sup-LR test for a coefficient change ignoring a variance change

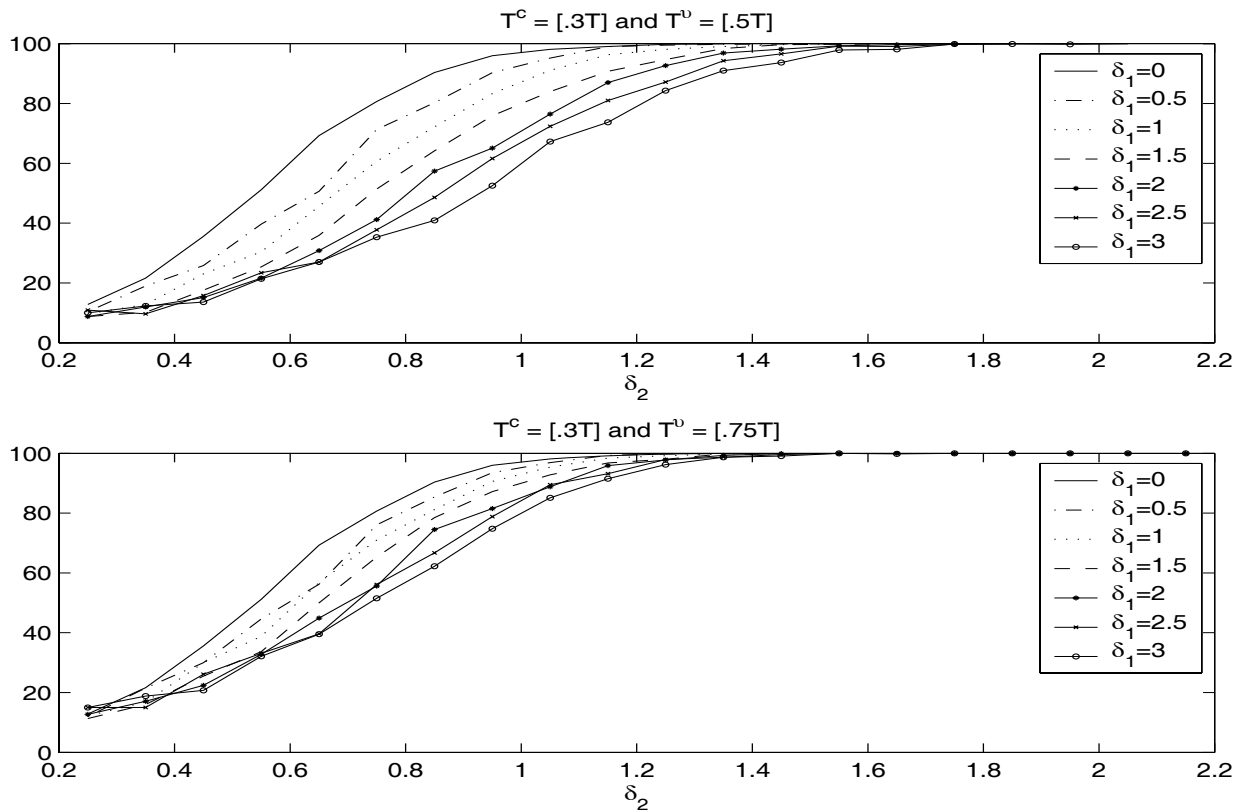


Figure 2: Power of the Sup-LR test for a coefficient change ignoring a variance change

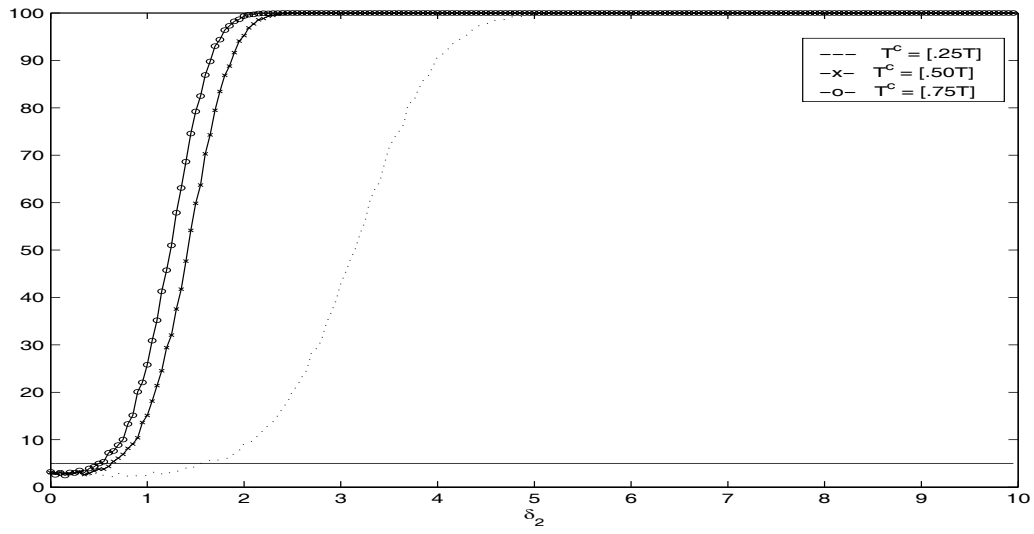


Figure 3: Size of the CUSQ test for a variance change ignoring a coefficient change

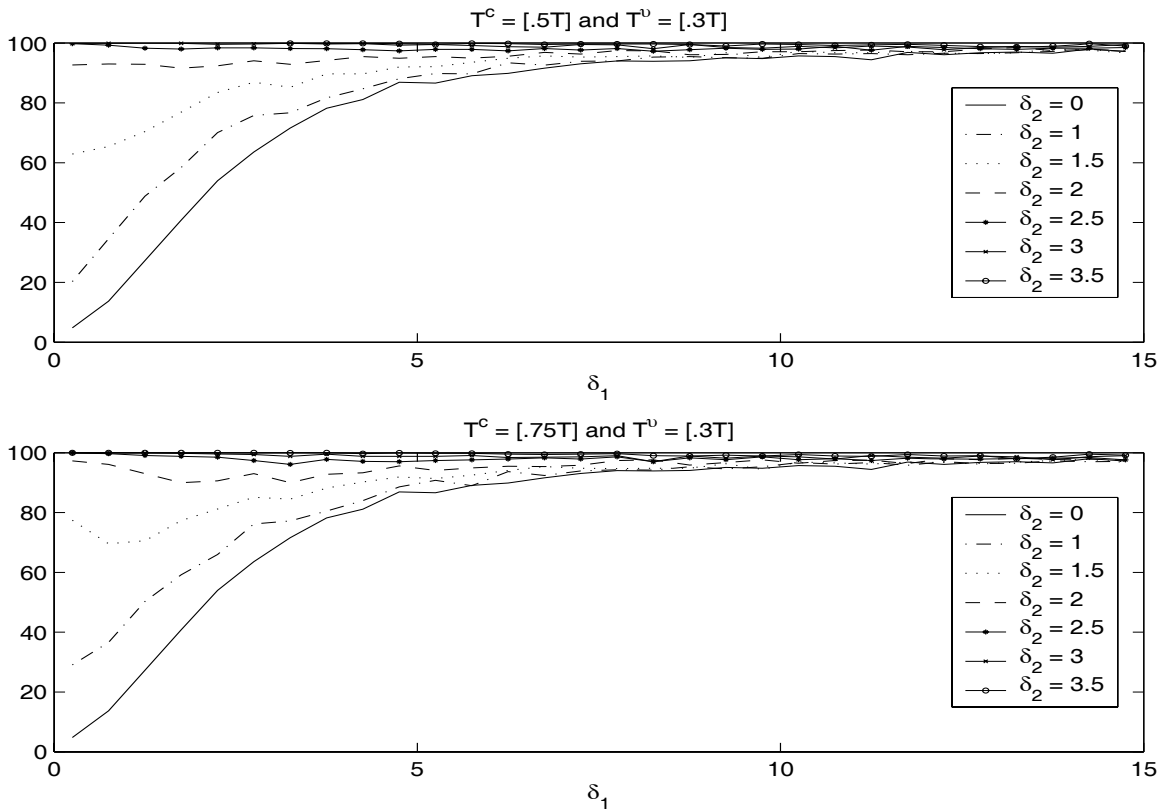


Figure 4: Power of the CUSQ test for a variance change ignoring a coefficient change

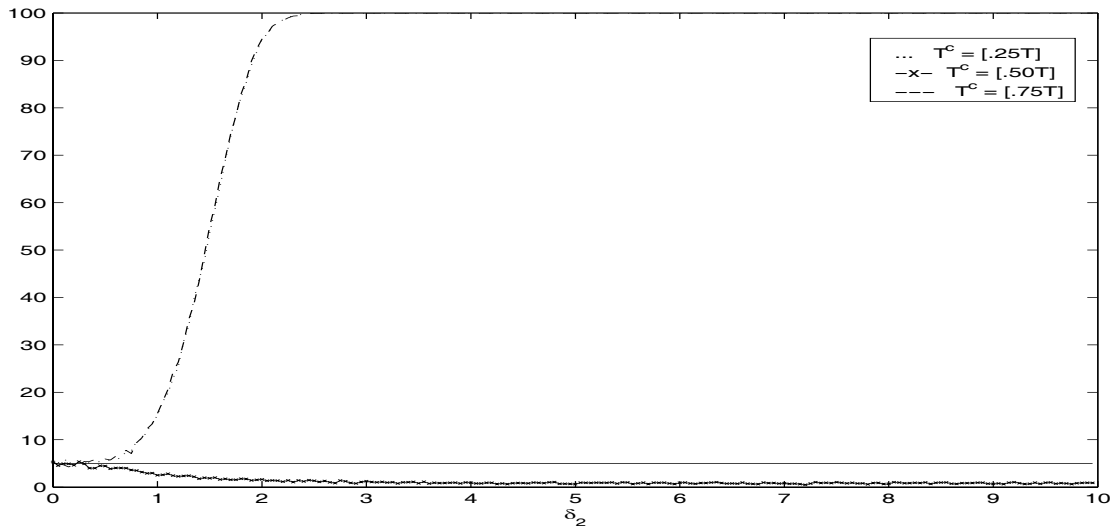


Figure 5: Size of the two-step test for a variance change ignoring a coefficient change

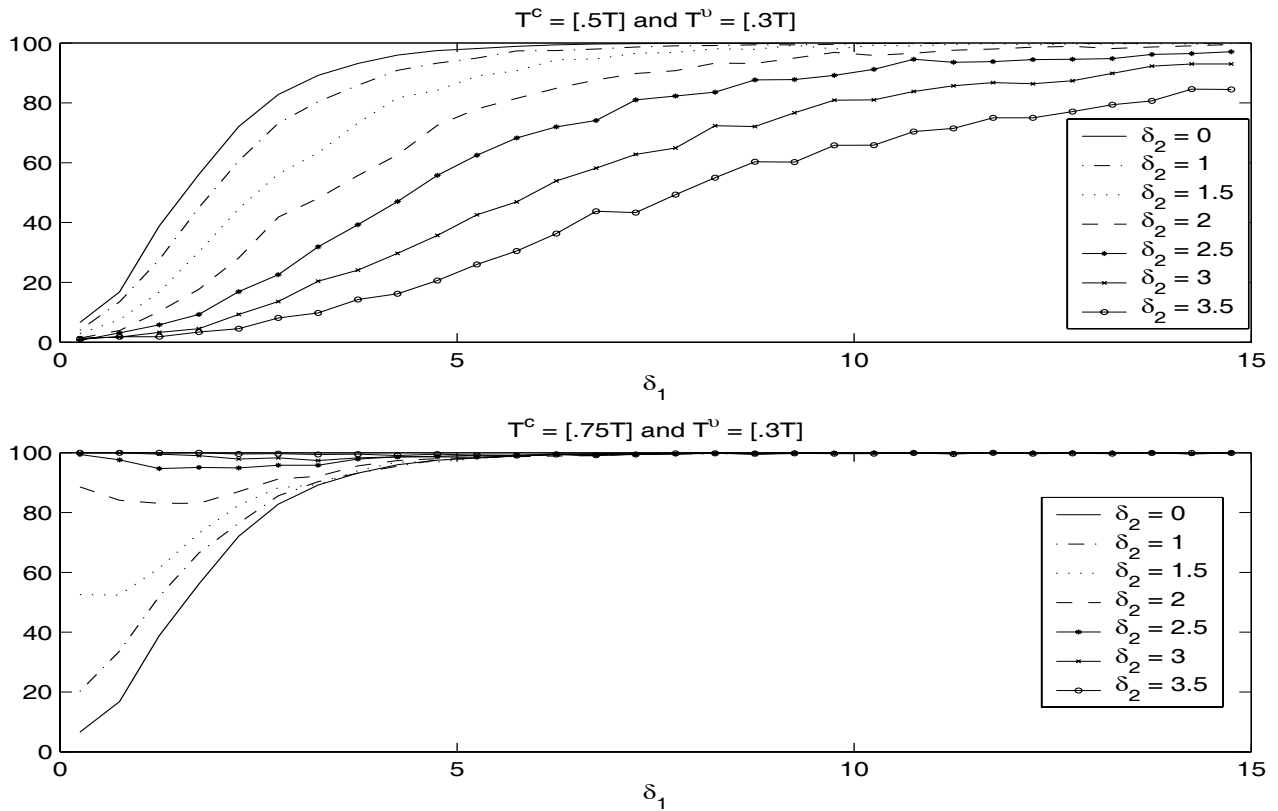


Figure 6: Power of the two-step test for a variance change ignoring a coefficient change

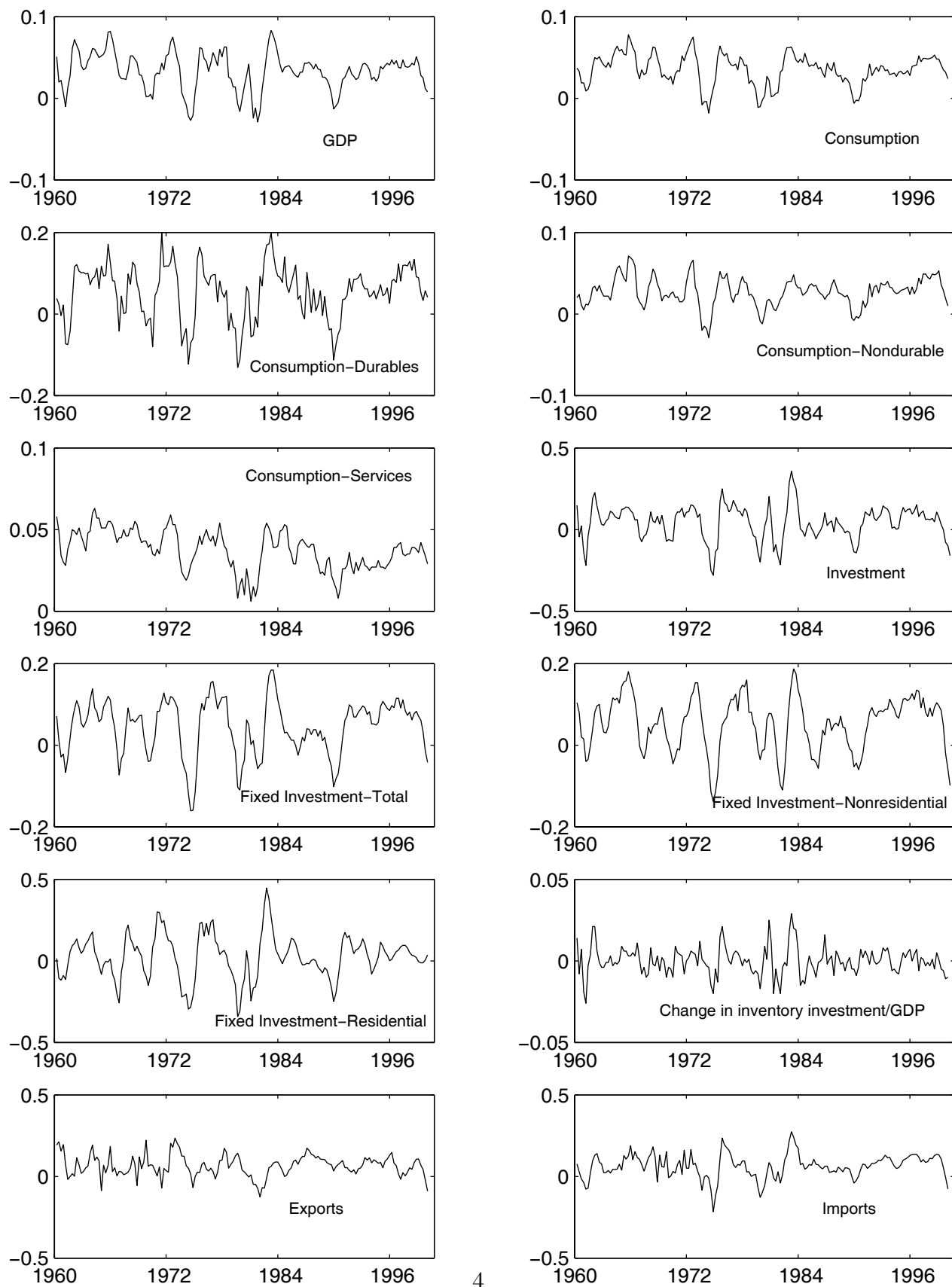


Figure 7: 22 Macro Time Series_1

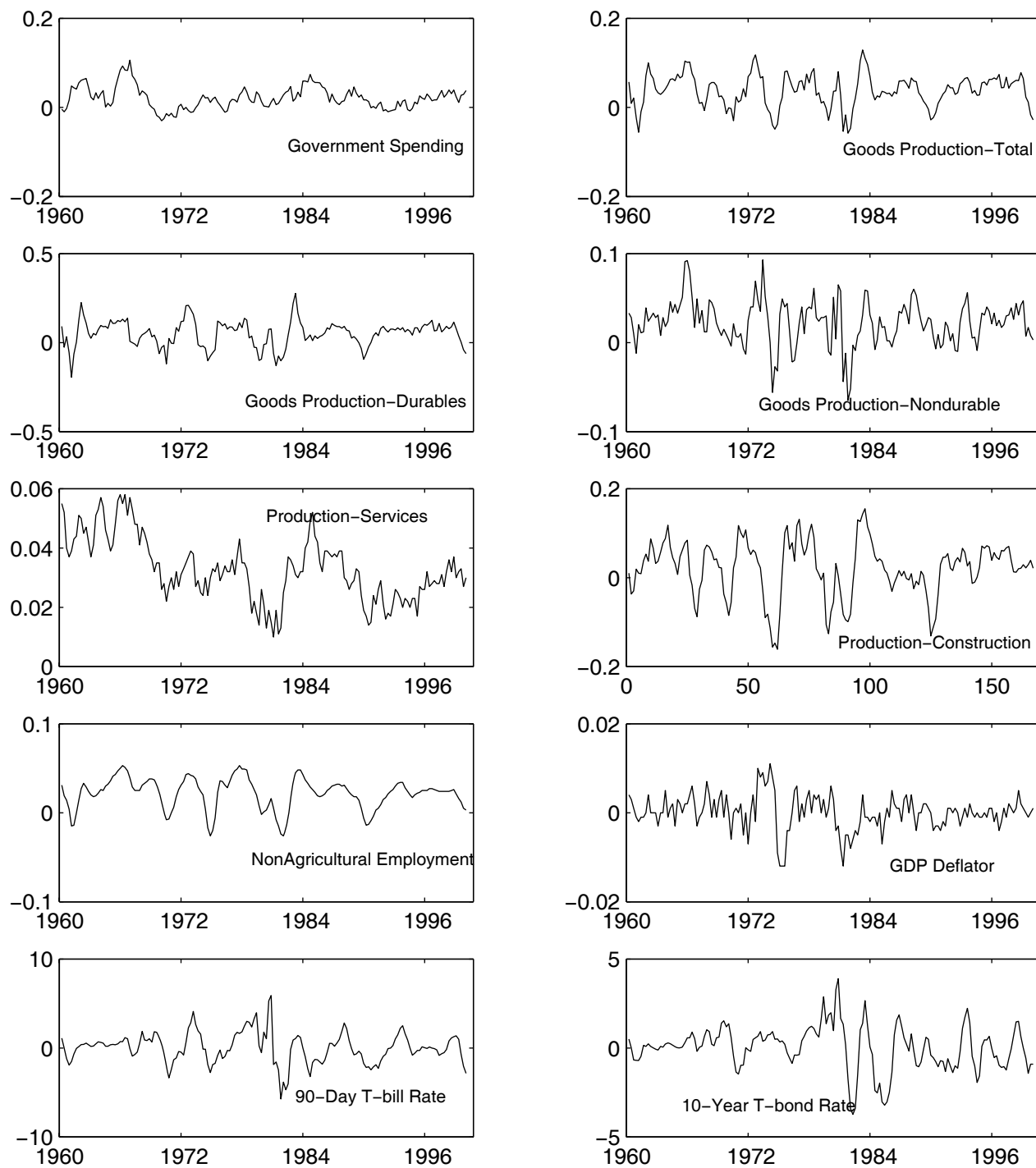


Figure 8: 22 Macro Time Series_2