Essays on Structural Breaks in Time Series and Panel Data Models

by

Daisuke Yamazaki

Submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in
Economics

Graduate School of Economics
Hitotsubashi University

2016

Acknowledgements

This dissertation contains research papers that I wrote at Graduate School of Economics, Hitotsubashi University. This thesis is composed of three articles on structural breaks in time series and panel data models. The first paper is concerned with testing for parameter constancy in panel data models. The second paper derives the first-order bias of the long-run variance estimator with structural breaks. The third paper proposes bias-corrected tests for a shift in mean of a time series. Chapter 2 is based on Yamazaki and Kurozumi (2015a), "Testing for Parameter Constancy in the Time Series Direction in Panel Data Models", *Journal of Statistical Computation and Simulation*, Vol. 85, pp. 2874–2902. Chapter 4 is based on Yamazaki and Kurozumi (2015b), "Improving the Finite Sample Performance of Tests for a Shift in Mean," *Journal of Statistical Planning and Inference*, Vol. 167, pp. 144–173.

I am deeply grateful to my supervisor, Professor Eiji Kurozumi for valuable advices and helpful comments. Without his continuing support and encouragement, I would never have completed this dissertation. I also appreciate Professor Yohei Yamamoto, Professor Toshio Honda, Professor Toshiaki Watanabe, Professor Toshihiro Yamada, and all the participants of my presentations for their constructive comments. I also thank my parents for enduring financial support.

Finally, this research has been supported by the Japan Society for the Promotion of Science Research Fellowship for Young Scientists.

March 2016

Daisuke Yamazaki

Contents

Chapter 1

Overview

1.1 Introduction

Time series models with structural breaks have been intensively investigated over the last fifty years, and various kinds of estimation methods and testing procedures have been proposed in the econometric and statistical literature. For the structural change tests, the CUSUM test by Brown, Durbin and Evans (1975) and Ploberger and Kramer (1992), the sup-type test by Andrews (1993), and the mean- and exponential-type tests by Andrews and Ploberger (1994) and Andrews, Lee and Ploberger (1996) are widely used in empirical analyses.

In practice, when we apply structural break tests, we need to take serial correlation into account, and thus we have to estimate the long-run variance of the error term. However, it is known in the literature that the finite sample performance of the tests is poor when we assume serial correlation in the error term. For example, if we estimate the long-run variance under the null hypothesis of no structural breaks, the tests suffer from the so-called "non-monotonic power" problem, as explained in Vogelsang (1999), Crainiceanu and Vogelsang (2007), Deng and Perron (2008) and Perron and Yamamoto (2014). The "non-monotonic power" problem is that the power decreases as the break magnitude increases, so that we cannot detect big structural breaks. The reason for this problem is that the long-run variance estimator using the residuals under the null hypothesis takes extremely large values when the break magnitude is large, and thus the test statistics take small values under the alternative hypothesis. On the other hand, if we estimate the long-run variance under the alternative hypothesis, the tests suffer from the size distortion because the long-run variance estimator has downward bias under the null hypothesis.

In order to cope with the problems, several methods have been proposed in the literature.

Sayginsoy and Vogelsang (2011) and Yang and Vogelsang (2011) proposed tests with good size by employing the fixed-b method. Shao and Zhang (2010) applied the self-normalization method to the CUSUM test to improve the size of the tests. However, the fixed-b and the self-normalizing methods use an inconsistent long-run variance estimator, so that the tests suffer from asymptotic power loss. On the other hand, Juhl and Xiao (2009) proposed to estimate the long-run variance using nonparametrically demeaned residuals to mitigate the non-monotonic power problem, but the finite sample performance of their test is very sensitive to the choice of the bandwidth used in the nonparametric estimation. Kejriwal (2009) proposed to estimate the long-run variance using the residuals both under the null and alternative hypotheses, but the test has extremely low power when the error term has strong serial correlation. Overall, the existing methods are not satisfactory, in view of both size and power.

While most of the existing literature consider the time series models, as the macro panel data become available, it is necessary to test for the constancy of parameters in panel data models. For panel data models, we need to consider the cases where the parameters are time-varying and heterogeneous. The tests for slope heterogeneity in panel data models are studied by Swamy (1970), Pesaran and Yamagata (2008) and Juhl and Lugovskyy (2014), but the tests for parameter constancy in the time series direction has not been widely studied in the literature.

In this thesis, we investigate the theoretical properties of structural break models, and propose solutions to the problems associated with structural breaks in time series and panel data models. In Chapter 2, we develop tests for parameter constancy in panel data models, taking heterogeneity into account. In Chapter 3, we derive the bias of the long-run variance estimator in the presence of structural breaks in mean, and propose a bias-corrected long-run variance estimator. In Chapter 4, we propose a bias-corrected test for a shift in mean.

1.2 Overview: Chapter 2

In Chapter 2, we propose tests for parameter constancy in the time series direction in the following heterogeneous-slope panel data model:

$$y_{it} = \alpha_i + x_{it}'\beta_{it} + u_{it}, \qquad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{1.1}$$

$$\beta_{it} = \beta_{i,t-1} + e_{it}, \tag{1.2}$$

where α_i is the individual effect, N is the number of cross sections, and T is the number of time series observations. We assume that u_{it} is heteroskedastic across cross-sections, and cross-sectionally dependent. The testing problem which we consider in this chapter is given by

$$H_0 : Var(e_{it}) = 0 \qquad \text{vs.} \qquad H_1 : Var(e_{it}) > 0.$$

Under the null hypothesis, β_{it} is constant across time, whereas under the alternative, β_{it} is time-varying.

We construct a locally optimal test based on Tanaka (1996) and an asymptotically point optimal test based on Elliott and Müller (2006), and derive the asymptotic distribution of the test statistics as $T \to \infty$ while N is fixed. We also consider the case where the parameter is homogeneous across cross-sections (which we call the "homogeneous-slope" model).

Since the asymptotic distribution depends on N in the heterogeneous-slope case, we need to calculate critical values and the optimal localizing parameter for each values of N. Therefore, we obtain the characteristic function of the limiting distributions, and derive the response surface of the critical values, and the optimal localizing parameter for the point-optimal test.

By Monte Carlo simulations, we find that the tests based on the homogeneous-slope model have serious size distortion when the true model has heterogeneous slopes. On the other hand, the tests based on the heterogeneous-slope model have good size for both the homogeneous- and heterogeneous-slope models, although these tests have lower power when the true model has homogeneous slopes. Therefore, we need to pay careful attention to the existence of heterogeneity in the slopes when we apply these tests.

1.3 Overview: Chapter 3

In Chapter 3, we consider the following time series model with multiple shifts in mean:

$$y_t = \begin{cases} \mu_1 + u_t & \text{for } t = 1, \cdots, T_1, \\ \mu_2 + u_t & \text{for } t = T_1 + 1, \cdots, T_2, \\ \qquad \vdots & \\ \mu_{m+1} + u_t & \text{for } t = T_m + 1, \cdots, T, \end{cases}$$

and consider estimating the long-run variance of the error term u_t. We estimate the long-run variance by the autoregressive spectral density estimator based on the AR(p) model, which

is defined as

$$\hat{\omega}_{AR} = \frac{\hat{\sigma}_{\varepsilon}^2}{\left(1 - \sum_{j=1}^{p} \hat{\phi}_j\right)^2},$$

where $\hat{u}_t = \sum_{j=1}^{p} \hat{\phi}_j \hat{u}_{t-j} + \hat{\varepsilon}_t$ with $\hat{\phi}_j$ $(j = 1, \cdots, p)$ being the OLS estimator, and $\hat{\sigma}_{\varepsilon}^2 = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t^2$.

It is known that the autoregressive spectral density estimator has downward bias in finite samples. In this chapter, we derive the bias of the autoregressive spectral density estimator up to $O(T^{-1})$, under the assumption that u_t follows a stationary AR(p) model. In order to derive the first-order bias of the long-run variance estimator, we first obtain the bias of the OLS estimator of the AR(p) regression in the presence of multiple shifts in mean. Then, we obtain the bias of the long-run variance estimator, and propose a bias-correction method. We find that the downward bias of the OLS estimator gets larger as the number of structural breaks increases, which leads to the downward bias of the long-run variance estimator.

When the error term u_t follows a stationary infinite-order autoregressive process, we need to truncate the lag order at p_T to implement the autoregression, and we let p_T go to infinity at an appropriate rate. In this case, we show that the first-order bias of the long-run variance estimator is exactly the same as the case with fixed p, so that our bias correction method can also be applied in such cases.

We perform simulations to investigate the finite sample properties of the long-run variance estimators. We find that the bias-corrected long-run variance estimator has much smaller bias than other estimators, and the mean squared error of the bias-corrected estimator is comparable to that of other estimators. Overall, we can see that our bias correction works well in finite samples.

1.4 Overview: Chapter 4

Chapter 4 considers the following mean-shift model:

$$y_t = \mu + \delta DU_t(T_b^0) + u_t, \quad t = 1, \cdots, T,$$

where $DU_t(T_b^0) = 1\{t > T_b^0\}$, and $1\{\cdot\}$ is the indicator function. We are interested in the following testing problem:

$$H_0 : \ \delta = 0 \quad \text{vs.} \quad H_1 : \ \delta \neq 0.$$

Under H_0, there is no shift in mean, while under H_1, there is a one-time break.

When u_t is serially correlated, we need to estimate the long-run variance of u_t for the scale adjustment in order to test for a shift in mean. If we estimate the long-run variance under the null hypothesis of no structural breaks, it is known that the tests suffer from the non-monotonic power problem because the long-run variance is over-estimated when the break magnitude is large (Vogelsang 1999, Crainiceanu and Vogelsang 2007, Perron and Yamamoto 2014). On the other hand, if we estimate the long-run variance under the alternative hypothesis, the tests suffer from the over-size distortion because the long-run variance is under-estimated under the null hypothesis.

In order to improve the finite sample properties of the tests, we propose bias correction to the long-run variance estimator, which is estimated under the alternative hypothesis. First, we derive the bias of the reciprocal of the autoregressive spectral density estimator based on the AR(p) model, under the assumption that the correct specification of u_t is the AR(p) process. Then, we propose bias correction to the test statistics.

We also discuss the cases where u_t follows a stationary AR(∞) process, and we find that the first-order bias is exactly the same as in the AR(p) case.

Simulation results show that the bias-corrected tests have much less size distortion than the tests without bias correction. Moreover, the bias-corrected tests have higher size-adjusted power than the existing tests. Since our proposed tests use a consistent long-run variance estimator, there is no asymptotic power loss due to bias correction. Thus, the bias-corrected tests have good finite sample property, in terms of both size and power.

Chapter 2

Testing for Parameter Constancy in the Time Series Direction in Panel Data Models

We propose tests for parameter constancy in the time series direction in panel data models. We construct a locally best invariant test based on Tanaka (1996) and an asymptotically point optimal test based on Elliott and Müller (2006). We derive the limiting distributions of the test statistics as $T \to \infty$ while N is fixed, and calculate the critical values by applying numerical integration and response surface regression. Simulation results show that the proposed tests perform well if we apply them appropriately.[1]

2.1 Introduction

This study proposes tests for parameter constancy in panel data models given by

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}, \tag{2.1}$$

where α_i is the individual effect and x_{it} is the vector of regressors. It is often the case that the parameter β is assumed to be constant across cross-sections and over time, but this assumption does not always hold; the violation of this assumption leads to a problem. If β varies across i and/or t, then the estimation based on (2.1) results in misleading statistical inference because the estimator of β is inconsistent. The variation in β across i is likely

[1]The published version is Yamazaki and Kurozumi (2015a), "Testing for Parameter Constancy in the Time Series Direction in Panel Data Models", *Journal of Statistical Computation and Simulation* 85, 2874–2902. (DOI:10.1080/00949655.2014.945089)

to happen when we use panel data with large N, while the unstable β in the time series direction typically comes from data with long T, such as financial data and macro panel data. From this reason, it is important to test whether or not the parameter β is constant across cross-sections and/or across time.

The motivation of this research is to detect parameter instability in the time series direction in macro panel data models or seemingly unrelated regression (SUR) models with large T and small to moderately-sized N. Therefore, we discuss the asymptotics as $T \to \infty$ while N is fixed.

There are several works related to this problem. For example, Swamy (1970), Pesaran and Yamagata (2008), and Juhl and Lugovskyy (2014) constructed tests for slope homogeneity in panel data models. On the other hand, tests for parameter constancy in the time series direction are rarely studied in the literature. Bai (2010) proposed a method of estimating the break point in mean in heterogeneous panel data, while Horváth and Hušková (2012) and Chan, Horváth and Hušková (2013) considered testing a shift in mean. On the other hand, Kim (2011) extended Bai's (2010) result for nonstationary panel data, but they did not consider tests for parameter constancy. One of the possible reasons for the lack of studies is that tests for parameter constancy in the time series direction may be seen as a simple extension of a univariate model to a multivariate one. This is partly true; however, as we show in this chapter, we have to carefully deal with slope heterogeneity across cross-sections when we test for parameter constancy in the time series direction.

In the time series literature, there have been many studies on testing for parameter constancy. For example, tests for shifts in slope are proposed by Andrews (1993), Andrews and Ploberger (1994), Andrews, Lee and Ploberger (1996), Bai (1997, 1999, 2000), Bai, Lumsdaine and Stock (1998), Bai and Perron (1998), Andrews and Kim (2006), Qu and Perron (2007), Kejriwal and Perron (2010), and Kim (2010). Further, the optimality is discussed by Andrews and Ploberger (1994), Andrews, Lee and Ploberger (1996), Sowell (1996), Kim and Perron (2009), and Perron and Yamamoto (2014). On the other hand, testing for parameter constancy using a time-varying parameter under the alternative is considered by Nyblom and Mäkeläinen (1983), King and Hillier (1985), Nyblom (1986, 1989), Nabeya and Tanaka (1988), Hansen (1992a, b), Kurozumi (2003), Elliott and Müller (2006), among others, and the optimal properties of the tests are studied in, for example, Nyblom and Mäkeläinen (1983), King and Hillier (1985), Nyblom (1989), Nabeya and Tanaka (1988), Kurozumi (2003), and Elliott and Müller (2006). An overall review of structural changes is given by Perron (2006) and Aue and Horváth (2013).

In this research, we extend a time series model with time-varying parameter to the corresponding panel data model and develop locally optimal and asymptotically point optimal tests for parameter constancy. We consider both the homogeneous-slope and heterogeneous-slope models and derive the limiting distributions of the test statistics as $T \to \infty$ with N fixed. Then, we derive the characteristic functions of the test statistics, which are used to obtain critical values by numerical integration based on Lévy's inversion formula, and implement response surface regressions to obtain critical values for the case of large degrees of freedom. We show that critical values are well approximated by response surface regressions. Through simulations, we show the importance of the careful treatment of the homogeneous/heterogeneous-slope cases in view of size and power.

The remainder of this chapter is organized as follows. In Section 2.2, we develop tests for parameter constancy. Calculation of critical values is discussed in Section 2.3, and the finite sample properties are investigated via Monte Carlo simulations in Section 2.4. Concluding remarks are given in Section 2.5. All mathematical proofs are delegated to the appendix.

2.2 Model, Assumptions and Test Statistics

2.2.1 Model and assumptions

In this section we consider the following panel data model with heterogeneous slopes:

$$y_{it} = \alpha_i + x_{it}'\beta_{it} + u_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{2.2}$$

where α_i is an individual effect, x_{it} is a $k \times 1$ vector of strictly exogenous regressors, and β_{it} is a time-varying $k \times 1$ vector of parameters. For each i, suppose that β_{it} evolves as follows:

$$\beta_{it} = \beta_{i,t-1} + e_{it}, \tag{2.3}$$

and the initial values β_{i0}, $i = 1, \cdots, N$ are unknown and nonstochastic. The time-varying specification in (2.3) implies the parameter varies smoothly, and that we cannot expect the future change in the parameter, such as the direction and the magnitude of change, based on the past information.[2] We call model (2.2)-(2.3) the "heterogeneous-slope" model.

For notational convenience, we stack equation (2.2) as follows. First, the equations at

[2]When β_{it} follows a stationary autoregressive process (i.e., $\beta_{it} - \bar{\beta}_i = \sum_{j=1}^{p} \Phi_{ij}(\beta_{i,t-j} - \bar{\beta}_i) + e_{it}$ where $|I - \sum_{j=1}^{p} \Phi_{ij} z^j| \neq 0$ for $|z| \leq 1$), our tests cannot detect parameter instability. In such cases, we need to use tests based on Shively (1988) and Lin and Teräsvirta (1999).

time t for individuals $i = 1, \cdots, N$ are stacked as

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} x'_{1t} & & & 0 \\ & x'_{2t} & & \\ & & \ddots & \\ 0 & & & x'_{Nt} \end{bmatrix} \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \vdots \\ \beta_{Nt} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Nt} \end{bmatrix},$$

or

$$y_t = \alpha + D_{xt}\beta_t + u_t, \tag{2.4}$$

where $D_{xt} = \text{diag}\{x'_{1t}, x'_{2t}, \cdots, x'_{Nt}\}$ is an $N \times kN$ matrix, and $\beta_t = [\beta'_{1t}, \beta'_{2t}, \cdots, \beta'_{Nt}]'$ is a $kN \times 1$ vector.

Then, stacking equation (2.4) from $t = 1$ to T, we have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} I_N \\ I_N \\ \vdots \\ I_N \end{bmatrix} \alpha + \begin{bmatrix} D_{x1} & & & 0 \\ & D_{x2} & & \\ & & \ddots & \\ 0 & & & D_{xT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix},$$

or

$$y = F_N\alpha + D_{dx}\beta + u, \tag{2.5}$$

where $F_N = \iota_T \otimes I_N$ with ι_T being a $T \times 1$ vector of ones and $D_{dx} = \text{diag}\{D_{x1}, D_{x2}, \cdots, D_{xN}\}$.

On the other hand, by letting $e_t = [e'_{1t}, e'_{2t}, \cdots, e'_{Nt}]'$ and $\beta_0 = [\beta'_{10}, \beta'_{20}, \cdots, \beta'_{N0}]'$, β can also be expressed as

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{bmatrix} = \begin{bmatrix} I_{kN} \\ I_{kN} \\ \vdots \\ I_{kN} \end{bmatrix} \beta_0 + \begin{bmatrix} I_{kN} & & & 0 \\ I_{kN} & I_{kN} & & \\ \vdots & \vdots & \ddots & \\ I_{kN} & I_{kN} & \dots & I_{kN} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix},$$

or

$$\beta = F_{kN}\beta_0 + (L \otimes I_{kN})e, \tag{2.6}$$

where $F_{kN} = \iota_T \otimes I_{kN}$ and L is a $T \times T$ random walk generating matrix, which is a lower triangular matrix with the diagonal and lower elements equal to 1.

Substituting (2.6) into (2.5) yields

$$\begin{aligned} y &= F_N\alpha + X\beta_0 + D_{dx}(L \otimes I_{kN})e + u \\ &= Z\gamma + v + u, \end{aligned}$$

where $Z = [F_N, X]$ with $X = D_{dx}F_{kN} = [D'_{x1}, D'_{x2}, \cdots, D'_{xT}]'$, $\gamma = [\alpha', \beta'_0]'$, and $v = D_{dx}(L \otimes I_{kN})e$. Note that v is the error term caused by the variation of β_{it} whereas u is the idiosyncratic error.

In order to construct test statistics and derive the limiting distributions as $T \to \infty$, we make the following assumptions.

Assumption A1 *For each i, $\{x_{it}\}_{t=1}^{T}$ is covariance stationary with finite fourth moments.*

Assumption A2 *$\{u_t\}$ is an i.i.d. sequence with $E[u_t] = 0$ and $E[u_t u'_t] = V$, where V is an $N \times N$ known positive definite matrix.*

Assumption A3 *$\{e_t\}$ is an i.i.d. sequence with $E[e_t] = 0$ and $E[e_t e'_t] = \rho\Sigma_e$, where Σ_e is a $kN \times kN$ known positive definite matrix.*

Assumption A4 *e is independent of both X and u.*

Assumption A5 *The individual effect α is fixed or independent of u and e.*

Assumption A6 *The variant $D'_{xt}V^{-1}u_t$ is a martingale difference sequence with respect to $\mathcal{F}_t = \sigma\{D_{xt}, u_t, D_{x,t-1}, u_{t-1}, \cdots\}$ with $E(D'_{xt}V^{-1}u_t u'_t V^{-1}D_{xt}|\mathcal{F}_{t-1}) = Q$ where $Q = E(D'_{xt}V^{-1}D_{xt}|\mathcal{F}_{t-1})$ is positive definite.*

Assumption A7 *$T^{-1}\sum_{t=1}^{[Tr]} D'_{xt}V^{-1}D_{xt} \xrightarrow{p} rQ$ and $T^{-1}\sum_{t=1}^{[Tr]} D_{xt} \xrightarrow{p} r\mu$ hold as $T \to \infty$ uniformly in $r \in [0,1]$, where $\mu = E(D_{xt})$, and $[a]$ denotes the largest integer less than or equal to a.*

Assumption A8 *For all $r \in [0,1]$, the following weak convergences hold jointly as $T \to \infty$:*

$$(a) \qquad \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} D'_{xt}V^{-1}u_t \xrightarrow{d} Q^{1/2}W_1(r),$$

$$(b) \qquad \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} D'_{xt}V^{-1}v_t \xrightarrow{d} cQ\Sigma_e^{1/2}\int_0^r W_2(s)ds,$$

where $\rho = c^2/T^2$ $(c \geq 0)$ with c fixed, and $W_1(r)$ and $W_2(r)$ are independent kN-dimensional standard Brownian motions.

Assumption A1 excludes a nonstationary regressor with a unit root. When x_{it} has a unit root, it should be first-differenced to be applied in this test. In our model, the regressor x_{it} is

allowed to be correlated across cross-sections. Assumption A2 means that the error term u_{it} is possibly heteroskedastic and cross-sectionally dependent. If we further assume that u is cross-sectionally independent, then V is simplified as $V = \text{diag}\ \{\sigma_{11}, \cdots, \sigma_{NN}\}$. By Assumption A3, the magnitude of the fluctuation in β is determined by the parameter ρ as well as Σ_e. We will localize the parameter ρ to consider the asymptotic local power functions in later sections. The assumption of the known variances of V and Σ_e will be relaxed later for practical purpose. The innovations driving the fluctuation in β_{it} are supposed to be independent of the regressors and the idiosyncratic errors in Assumption A4. Assumption A5 is required for deriving the likelihood function but this can be relaxed in practical analysis because our test statistics are invariant to α. We make Assumptions A6-A8 to derive limiting distributions under both the null and alternative hypotheses. For example, when x_t is stationary with finite fourth moment and x_t is independent of u_t and e_t, Assumptions A6-A8 are satisfied.

We are interested in a testing problem given by

$$H_0 : \ \rho = 0 \quad \text{vs.} \quad H_1 : \ \rho > 0. \tag{2.7}$$

Under H_0, $Var(e_{it}) = 0$ and then β_{it} is constant across t. On the other hand, β_{it} is smoothly time-varying under H_1. Such time-varying parameter models have been considered in the econometric and statistical literature.

Note that our test can be applied to models with multiple structural changes. For example, suppose that e_t is independently distributed and

$$e_{it} = \begin{cases} m_{it} & \text{with probability } p \\ -m_{it} & \text{with probability } p \\ 0 & \text{with probability } 1 - 2p \end{cases}$$

where p is close to zero, and m_{it} is large. This can be viewed as a model with multiple structural breaks. Therefore, the alternative hypothesis in our model is more general than that of Bai (1997) and Qu and Perron (2007) because they only assume multiple structural breaks, and they exclude the time-varying parameter model as the alternative.

2.2.2 The LM test

To derive the Lagrange multiplier (LM) test statistic, we assume that u and e are normally distributed and X and u are independent. This assumption is made only for the derivation of the test statistic and the discussion of the optimality; theorems and corollaries in the following do not require this assumption.

By noting that $e \sim N(0, \rho I_T \otimes \Sigma_e)$ under Assumptions A1 to A5 with normality, the distribution of y conditional on α and X is given by

$$y|\alpha, X \sim N\left(Z\gamma, \Sigma(\rho)\right) \quad \text{where} \quad \Sigma(\rho) = I_T \otimes V + \rho D_{dx}(LL' \otimes \Sigma_e)D'_{dx}.$$

Then, by letting $w = u + v$, the model can be represented as

$$y = Z\gamma + w, \quad w \sim N(0, \Sigma(\rho)),$$

and this corresponds to model (9.5) in Tanaka (1996). Therefore, following Section 9.2 in Tanaka (1996), the one-sided LM test for the testing problem (2.7) rejects the null hypothesis when

$$y' M'_Z \Sigma^{-1}(0) \left(\frac{d\Sigma(\rho)}{d\rho}\bigg|_{\rho=0} \right) \Sigma^{-1}(0) M_Z y > \text{constant},$$

where $M_Z = I_{NT} - Z(Z'(I_T \otimes V^{-1})Z)^{-1}Z'(I_T \otimes V^{-1})$ and $\frac{d\Sigma(\rho)}{d\rho}\big|_{\rho=0} = D_{dx}(LL' \otimes \Sigma_e)D'_{dx}$. Therefore, the LM test statistic is given by

$$LM_{hetero} = \frac{1}{T^2} y' M'_Z (I_T \otimes V^{-1}) D_{dx}(LL' \otimes \Sigma_e) D'_{dx} (I_T \otimes V^{-1}) M_Z y. \tag{2.8}$$

For computational purpose, it would be convenient to express the LM test statistic as

$$LM_{hetero} = \frac{1}{T^2} \sum_{t=1}^{T} s'_t \Sigma_e s_t,$$

where $s_t = \sum_{s=1}^{t} D'_{xs} V^{-1} \hat{v}_s$, $\hat{v}_t = [\hat{v}_{1t}, \cdots, \hat{v}_{Nt}]'$, and $\hat{v}_{it}$ is the residual of the GLS regression of

$$y_{it} = \alpha_i + x'_{it}\beta_i + v_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T.$$

Remark 1 *Under the assumption of normality, the LM test is equivalent to the locally best invariant (LBI) test as shown by Tanaka (1996). Moreover, even without the assumption of normality, the LM test is an asymptotically LBI test because the limiting distribution in Theorem 1 does not depend on specific distributions of u_t and e_t.*

In practice, V is replaced with a consistent estimator $\hat{V}$, where the (i, j)-th element of $\hat{V}$ is given by $\hat{\sigma}_{ij}$ in equation (2.10).

The limiting distribution of (2.8) is derived under the local alternative given by

$$H_1 : \ \rho = \frac{c^2}{T^2} \quad (c \geq 0)$$

as T goes to infinity while N is fixed.

Theorem 1 *Under Assumptions A1 to A8 with* $\rho = c^2/T^2$*, the LM test statistic weakly converges to*

$$LM_{hetero} \xrightarrow{d} \int_0^1 V_{kN}(r;c)' V_{kN}(r;c) dr, \tag{2.9}$$

as $T \to \infty$ *while* N *is fixed, where*

$$\begin{aligned} V_{kN}(r;c) &= \Sigma_e^{1/2} \left(B_1(r) + cB_2(r) \right), \\ B_1(r) &= Q^{1/2} \left(W_1(r) - rW_1(1) \right), \\ B_2(r) &= Q\Sigma_e^{1/2} \left(\int_0^r W_2(s)ds - r \int_0^1 W_2(s)ds \right), \end{aligned}$$

and $W_1(r)$ *and* $W_2(r)$ *are independent* kN*-dimensional standard Brownian motions.*

Remark 2 *In the case of the partial structural change model given by*

$$y_{it} = \alpha_i + x_{it}'\beta_{it} + w_{it}'\delta_i + u_{it},$$

where w_{it} *is a covariance stationary and strictly exogenous regressor, the LM test statistic is the same as* (2.8) *with* Z *replaced by* $[F_N, X, W]$*,* $W = [D_{w1}', D_{w2}', \cdots, D_{wT}']'$*, and* $D_{wt} = diag\{w_{1t}', w_{2t}', \cdots, w_{Nt}'\}$*. In this case, the limiting distribution is exactly the same as in Theorem 1.*

Remark 3 *In Theorem 1, the limiting distribution depends on* Σ_e *and* Q*, but when* $\Sigma_e = Q^{-1}$ *holds, we can easily see that the limiting distribution is free of nuisance parameters and is given by (2.9) with* $V_{kN}(r;c)$ *replaced by*

$$V_{kN}(r;c) = W_1(r) - rW_1(1) + c\left(\int_0^r W_2(s)ds - r\int_0^1 W_2(s)ds \right).$$

In practice, we do not know the true value of V and Σ_e and we need to modify the test statistic LM_{hetero}. We replace V with $\hat{V}$, whose (i,j)-th element is given by

$$\hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} (y_{it} - \hat{\alpha}_i - x_{it}'\hat{\beta}_i)(y_{jt} - \hat{\alpha}_j - x_{jt}'\hat{\beta}_j) \tag{2.10}$$

and $\hat{\alpha}_i, \hat{\beta}_i$ are based on the following OLS regression:

$$y_{it} = \alpha_i + x_{it}'\beta_i + u_{it}, \quad t = 1, \cdots, T.$$

The proof of consistency of $\hat{\sigma}_{ij}$ is given in the appendix. On the other hand, to obtain the asymptotic null distribution that is free of nuisance parameters, we replace Σ_e not with the consistent estimator of Σ_e but with $\hat{Q}^{-1}$, where $\hat{Q}$ is the consistent estimator of Q given by

$$\hat{Q} = \frac{1}{T}\sum_{t=1}^{T} D_{xt}'\hat{V}^{-1}D_{xt}. \tag{2.11}$$

In this case, the test statistic is modified as

$$\begin{aligned}\widetilde{LM}_{hetero} &= \frac{1}{T^2} y' M_Z' (I_T \otimes \hat{V}^{-1}) D_{dx} (LL' \otimes \hat{Q}^{-1}) D_{dx}' (I_T \otimes \hat{V}^{-1}) M_Z y \\ &= \frac{1}{T^2} \sum_{t=1}^{T} \hat{s}_t' \hat{Q}^{-1} \hat{s}_t.\end{aligned}$$

Since $\hat{V} \xrightarrow{p} V$ and $\hat{Q} \xrightarrow{p} Q$ by the law of large numbers, we obtain the following corollary.

Corollary 1 *Under Assumptions A1 to A8 with $\rho = c^2/T^2$, the modified LM test statistic weakly converges to*

$$\widetilde{LM}_{hetero} \xrightarrow{d} \int_0^1 \tilde{V}_{kN}(r;c)' \tilde{V}_{kN}(r;c) dr, \tag{2.12}$$

as $T \to \infty$ while N is fixed, where

$$\tilde{V}_{kN}(r;c) = W_1(r) - rW_1(1) + cQ^{1/2}\Sigma_e^{1/2} \left(\int_0^r W_2(s)ds - r \int_0^1 W_2(s)ds \right).$$

From this result, we can see that the null distribution becomes $\int_0^1 (W_1(r) - rW_1(1))'(W_1(r) - rW_1(1))dr$, which is free of nuisance parameters and is known as the generalized Von Mises distribution with kN degrees of freedom. Therefore, the modified LM test is feasible, but it is not locally optimal unless $\Sigma_e = Q^{-1}$ holds. The calculation of critical values for this distribution will be discussed in a later section.

2.2.3 The asymptotically point optimal test

In this subsection, we extend the asymptotically point optimal test for parameter constancy in time series models proposed by Elliott and Müller (2006) to panel data models. Following their result, the test statistic we consider is given by

$$qLL_{hetero} = \sum_{\ell=1}^{kN} \hat{v}_\ell' (G_{\bar{c}} - M_e) \hat{v}_\ell, \tag{2.13}$$

where we reject the null hypothesis when qLL_{hetero} takes small values, and

$$\begin{aligned}
\hat{v}_\ell &= (I_T \otimes \iota'_{kN,\ell} Q^{-1/2}) D'_{dx} (I_T \otimes V^{-1}) M_Z y, \\
M_e &= I_T - T^{-1} \iota_T \iota'_T, \\
G_{\bar{c}} &= H_{\bar{c}}^{-1} - H_{\bar{c}}^{-1} \iota_T (\iota'_T H_{\bar{c}}^{-1} \iota_T)^{-1} \iota'_T H_{\bar{c}}^{-1}, \\
H_{\bar{c}} &= r_{\bar{c}}^{-1} L A_{\bar{c}} A'_{\bar{c}} L', \\
A_{\bar{c}} &= \begin{bmatrix} 1 & & & 0 \\ -r_{\bar{c}} & \ddots & & \\ & \ddots & \ddots & \\ 0 & & -r_{\bar{c}} & 1 \end{bmatrix} \qquad (T \times T \text{ matrix}), \\
r_{\bar{c}} &= \frac{1}{2} \left(2 + \frac{\bar{c}^2}{T^2} - \frac{1}{T} \sqrt{4\bar{c}^2 + \frac{\bar{c}^4}{T^2}} \right),
\end{aligned}$$

and $\iota_{kN,\ell}$ is a $kN \times 1$ vector with 1 in the ℓ-th element and 0s elsewhere. Note that we need to prespecify the value of the localizing parameter $\bar{c}$.

To derive the limiting distribution of qLL_{hetero}, we assume $\Sigma_e = Q^{-1}$ in order for the test statistic to be asymptotically free of nuisance parameters under the local alternative, which is a useful result when we choose an optimal value of $\bar{c}$ in a later section.

Theorem 2 *Under Assumptions A1 to A8 with $\rho = c^2/T^2$ and $\Sigma_e = Q^{-1}$, the qLL_{hetero} statistic weakly converges to*

$$qLL_{hetero} \xrightarrow{d} \sum_{\ell=1}^{kN} R_\ell(c, \bar{c}), \tag{2.14}$$

as $T \to \infty$ while N is fixed, where

$$\begin{aligned}
R_\ell(c, \bar{c}) &= \left[-\bar{c} L_\ell(1)^2 - \bar{c}^2 \int_0^1 L_\ell(s)^2 ds - \frac{2\bar{c}}{1 - e^{2\bar{c}}} \left(e^{-\bar{c}} L_\ell(1) + \bar{c} \int_0^1 e^{-\bar{c}s} L_\ell(s) ds \right)^2 \right. \\
&\quad \left. + \left(L_\ell(1) + \bar{c} \int_0^1 L_\ell(s) ds \right)^2 \right], \qquad (2.15) \\
L_\ell(r) &= J_\ell(r) + c K_\ell(r), \\
J_\ell(r) &= W_{1,\ell}(r) - \bar{c} \int_0^r e^{-\bar{c}(r-s)} W_{1,\ell}(s) ds, \\
K_\ell(r) &= \int_0^r W_{2,\ell}(s) ds - \bar{c} \int_0^r e^{-\bar{c}(r-s)} \left(\int_0^s W_{2,\ell}(\lambda) d\lambda \right) ds,
\end{aligned}$$

and $W_m(r) = (W_{m,1}(r), \cdots, W_{m,kN}(r))'$, $m = 1, 2$ are independent kN-dimensional standard Brownian motions.

Remark 4 *As is seen in the previous section, the LM test is asymptotically locally best irrespective of whether or not* $\Sigma_e = Q^{-1}$ *is assumed, but its null limiting distribution becomes asymptotically free of nuisance parameters only under the assumption of* $\Sigma_e = Q^{-1}$*. On the other hand, the* qLL_{hetero} *statistic is asymptotically free of nuisance parameters under the null hypothesis even without this assumption, but it is an asymptotically point optimal test only under the assumption of* $\Sigma_e = Q^{-1}$*. The asymptotic optimality is proved in exactly the same way as in Elliott and Müller (2006), and we omit the proof.*

In practice, Q and V are replaced with consistent estimators $\hat{Q}$ and $\hat{V}$ as in the case of the LM test and in this case, the test statistic is denoted as $\widetilde{qLL}_{hetero}$. The choice of the localizing parameter $\bar{c}$ and the calculation of critical values will be discussed in a later section.

2.2.4 The case with serially correlated errors

The above asymptotic results crucially depend on the assumption that $D_{xt}'V^{-1}u_t$ or u_t is serially uncorrelated but we may need to take serial correlation into account in practical analysis. When the error term u_t is serially correlated, the matrix Q in Assumption A8(a) must be replaced by the long-run variance of $D_{xt}'V^{-1}u_t$.

One of the possible candidates for the consistent estimator is the heteroskedasticity-autocorrelation consistent (HAC) estimator. However, it is known that the tests based on the HAC estimator has non-monotonic power as discussed in Vogelsang (1999) and Perron and Yamamoto (2014). Although the non-monotonic power problem has not been completely solved, some methods have been proposed to mitigate the problem (cf. Kejriwal 2009, Juhl and Xiao 2009).

2.2.5 Tests under the homogeneous-slope model

We have so far considered the heterogeneous-slope model where the parameter β varies across cross-sections. However, when the slopes are homogeneous across cross-sections, it is better to apply tests based on the following homogeneous-slope model:

$$y_{it} = \alpha_i + x_{it}'\beta_t + u_{it}, \tag{2.16}$$

$$\beta_t = \beta_{t-1} + e_t. \tag{2.17}$$

In the following, we allow an abuse of notation by defining e_t, β_t, X, Σ_e, μ and Q in a different way from the previous subsections to save notation. For the homogeneous-slope model, we modify Assumptions A3 and A6–A8 as follows:

Assumption B3 *$\{e_t\}$ is an i.i.d. sequence with $E[e_t] = 0$ and $E[e_t e_t'] = \rho\Sigma_e$, where Σ_e is a $k \times k$ known positive definite matrix.*

Assumption B6 *The variant $X_t'V^{-1}u_t$ is a martingale difference sequence with respect to $\mathcal{F}_t = \sigma\{X_t, u_t, X_{t-1}, u_{t-1}, \cdots\}$ with $E(X_t'V^{-1}u_t u_t'V^{-1}X_t|\mathcal{F}_{t-1}) = Q$ where $X_t = [x_{1t}, \cdots, x_{Nt}]'$ and $Q = E(X_t'V^{-1}X_t|\mathcal{F}_{t-1})$ is positive definite.*

Assumption B7 *$T^{-1}\sum_{t=1}^{[Tr]} X_t'V^{-1}X_t \xrightarrow{p} rQ$ and $T^{-1}\sum_{t=1}^{[Tr]} X_t \xrightarrow{p} r\mu$ hold as $T \to \infty$ uniformly in $r \in [0,1]$, where $\mu = E(X_t)$.*

Assumption B8 *For all $r \in [0,1]$, the following weak convergences hold jointly as $T \to \infty$:*

$$(a) \quad \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} X_t'V^{-1}u_t \xrightarrow{d} Q^{1/2}W_1(r),$$

$$(b) \quad \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} X_t'V^{-1}v_t \xrightarrow{d} cQ\Sigma_e^{1/2}\int_0^r W_2(s)ds,$$

where $\rho = c^2/T^2$ $(c \geq 0)$ with c fixed, and $W_1(r)$ and $W_2(r)$ are independent k-dimensional standard Brownian motions.

The modified LM and qLL statistics for the homogeneous-slope model are given by

$$\widetilde{LM}_{homo} = \frac{1}{T}y'M_Z(I_T \otimes \hat{V}^{-1})D_X(LL' \otimes \hat{Q}^{-1})D_X'(I_T \otimes \hat{V}^{-1})M_Z y, \tag{2.18}$$

$$\widetilde{qLL}_{homo} = \sum_{\ell=1}^{k} \hat{v}_\ell'(G_{\bar{c}} - M_e)\hat{v}_\ell, \tag{2.19}$$

where $\hat{Q} = T^{-1}\sum_{t=1}^{T} X_t'\hat{V}^{-1}X_t$, $D_X = \text{diag}\{X_1, X_2, \cdots, X_T\}$, $M_Z = I_{NT} - Z(Z'(I_T \otimes \hat{V}^{-1})Z)^{-1}Z'(I_T \otimes \hat{V}^{-1})$, $Z = [F_N, X]$, $X = [X_1', X_2', \cdots, X_T']'$ and $\hat{v}_\ell = (I_T \otimes \iota_{k,\ell}'\hat{Q}^{-1/2})D_X'$.

The limiting distributions of $\widetilde{LM}_{homo}$ and $\widetilde{qLL}_{homo}$ are given by the following corollaries.

Corollary 2 *Under Assumptions A1, A2, B3, A4, A5, B6–B8 with $\rho = c^2/T^2$, the modified LM test statistic $\widetilde{LM}_{homo}$ weakly converges to*

$$\widetilde{LM}_{homo} \xrightarrow{d} \int_0^1 \tilde{V}_k(r;c)'\tilde{V}_k(r;c)dr$$

as $T \to \infty$ while N is fixed, where $\tilde{V}_k(r;c)$ is defined as $\tilde{V}_{kN}(r;c)$ in Corollary 1 with $W_1(r)$ and $W_2(r)$ being independent k-dimensional standard Brownian motions.

Corollary 3 *Under Assumptions A1, A2, B3, A4, A5, B6–B8 with* $\rho = c^2/T^2$*, the* $\widetilde{qLL}_{homo}$ *statistic weakly converges to*

$$\widetilde{qLL}_{homo} \xrightarrow{d} \sum_{\ell=1}^{k} R_\ell(c, \bar{c}), \tag{2.20}$$

where $R_\ell(c, \bar{c})$ *is defined as in Theorem 2 with* $W_m(r) = (W_{m,1}(r), W_{m,2}(r), \cdots, W_{m,k}(r))'$*,* $m = 1, 2$ *being independent* k*-dimensional standard Brownian motions.*

Note that the limiting distributions in Corollaries 2 and 3 are almost the same as those in Corollary 1 and Theorem 2, respectively; the only difference is the number of independent Brownian motions.

2.3 Choice of the Localizing Parameter and Calculation of Critical Values

In this section, we discuss how to choose the localizing parameter $\bar{c}$ for the qLL tests and the calculation of critical values. We first note that critical values for the null limiting distribution of the LM test statistic, which is the generalized Von Mises distribution, have already been tabulated in the literature; for example, Canova and Hansen (1995) tabulate critical values up to 12 degrees of freedom. However, in the heterogeneous-slope case, the degrees of freedom are kN as given in Theorem 1, which could be very large in practical analysis. Similarly, the null limiting distribution of the qLL test statistic in the heterogeneous-slope case is also the sum of kN independent random variables. In order to calculate critical values computationally efficiently for all practical values of kN, we use the numerical integration of the characteristic function. This is computationally much faster and more accurate than the simulation based method.

In general, if the test statistic S is a nonnegative statistic, then it is known that the distribution of S can be computed by Lévy's inversion theorem,

$$P(S \leq x) = \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{1 - e^{-i\theta x}}{i\theta} \phi(\theta)\right] d\theta,$$

where $\phi(\theta)$ is the characteristic function of S. We can apply this formula to our case because $\widetilde{LM}_{homo}$, $\widetilde{LM}_{hetero}$, $-\widetilde{qLL}_{homo}$, and $-\widetilde{qLL}_{hetero}$ are all nonnegative. Then, we need the characteristic functions of the limiting distributions of these test statistics, which are given

by

$$\begin{aligned}\phi_{LM}(\theta;J) &= E\left[\exp\left(i\theta\int_0^1 \tilde{V}_J(r;0)'\tilde{V}_J(r;0)dr\right)\right],\\ \phi_{qLL}(\theta;J) &= E\left[\exp\left(-i\theta\sum_{\ell=1}^{J} R_\ell(c,\bar{c})\right)\right],\end{aligned}$$

where J is a positive integer value ($J = k$ or kN in our theorems). Note that we need only the characteristic function of the null distribution for the LM test whereas those under the null and the alternative are required for the qLL test because we will make use of the asymptotic local power functions to determine the localizing parameter $\bar{c}$.

Theorem 3 *The characteristic functions of the LM and the point optimal tests are given by*

$$\begin{aligned}\phi_{LM}(\theta;J) &= \left[\frac{\sin\sqrt{2i\theta}}{\sqrt{2i\theta}}\right]^{-J/2},\\ \phi_{qLL}(\theta;J) &= \frac{\phi_{LM}\left(\frac{a_1+a_2}{2i};J\right)\phi_{LM}\left(\frac{a_1-a_2}{2i};J\right)}{\phi_{LM}\left(\frac{-c^2}{2i};J\right)},\end{aligned}$$

where $a_1 = \bar{c}^2(2i\theta-1)/2$ *and* $a_2 = (\bar{c}/2)\{(2i\theta-1)^2c^2 + 8i\bar{c}^2\theta\}^{1/2}$.

Although the explicit expression of $\phi_{qLL}(\theta;J)$ is complicated, it can be expressed in compact form by using $\phi_{LM}(\theta;J)$.

The critical values of the LM test statistic are obtained by numerical integration for $J = 1,\cdots,500$. Because tables for $J = 1,\cdots,500$ are too large and inconvenient, we derive the response surface of the critical values. Having considered various functions of J, we adopt the following regression:

$$cv_{LM}(p,J) = a_0 + a_1\sqrt{J} + a_2 J + a_3\frac{1}{\sqrt{J}} + a_4\frac{1}{J},$$

where $cv_{LM}(p,J)$ represents the percentiles of the LM test statistic for $p = 0.9$, 0.95 and 0.99 and J represents the degrees of freedom. The estimated coefficients are given in Table 2.1. The largest ratio of the residual to the actual critical value is 0.00056 in absolute value.

On the other hand, before obtaining the critical values of the qLL test statistic, we need to determine the localizing parameter $\bar{c}$. Elliott and Müller (2006) proposed to set $\bar{c}$ to 10 for $J = 1,\cdots,10$, but as we will see later, this value is optimal only for $J = 1$, and our preliminary simulations reveal that when $J = 10$, the qLL test with $\bar{c} = 10$ is less powerful than the LM test for a wide range of the alternative.

In this chapter, we determine the localizing parameter $\bar{c}$ following Juhl and Xiao (2003) and Kurozumi (2003), which are based on the idea of Cox and Hinkley (1974, p.102). We propose to choose $\bar{c}$ that maximizes the weighted average of power[3]:

$$\bar{c} = \arg\max \int_0^M P(-qLL > x; c)dc$$

where x is a critical value for a given significance level and M is chosen to be so large that $P(-qLL > x; M)$ is close to 1. In this chapter, we set the significance level to 0.05 and calculate the optimal values of $\bar{c}$ for $J = 1, \cdots, 500$. Next, we obtain the response surface of $\bar{c}$, and for the given value of J and $\bar{c}$ on the corresponding response surface, we calculate the critical values of the qLL test for $J = 1, \cdots, 500$ and again obtain the response surface of the critical values as a function of J and $\bar{c}$. The adopted regressions are as follows:

$$\bar{c} = a_0 + a_1\sqrt{J} + a_2 J + a_3\frac{1}{\sqrt{J}} + a_4\frac{1}{J} + a_5\frac{1}{J\sqrt{J}} + a_6\frac{1}{J^2}, \tag{2.21}$$

$$cv_{qLL}(p, J) = a_0 + a_1\sqrt{J} + a_2 J + a_3\frac{1}{\sqrt{J}} + a_4\frac{1}{J} + b_1\sqrt{\bar{c}} + b_2\bar{c} + b_3\frac{1}{\sqrt{\bar{c}}} + b_4\frac{1}{\bar{c}}, \tag{2.22}$$

where $cv_{qLL}(p, J)$ represents the percentiles of the qLL test statistic for $p = 0.9$, 0.95, and 0.99 and we used the same notation a_i in equations (2.21) and (2.22) for notational convenience but they take different values depending on the equations. The estimated coefficients are given in Table 2.1. The largest ratio of the residual to the true value of $\bar{c}$ is 0.0285 in absolute value, while the corresponding ratio for the critical values is 0.00007. In practice, we first obtain $\bar{c}$ from equation (2.21) and then obtain critical values by putting the estimated $\bar{c}$ into (2.22). We also note that the critical values obtained by (2.21) and (2.22) correspond to $-\widetilde{qLL}_{homo}$ and $-\widetilde{qLL}_{hetero}$.

2.4 Simulation Results

In this section, we investigate the finite sample properties of the tests proposed in this study by using the Monte Carlo experiment. We examine the sizes and powers of the LM, qLL and sup-Wald tests. The following is the data generating process we considered in the simulations:

(DGP1: Homogeneous-slope model, time-varying parameter, $k = 1$)

$$y_{it} = \alpha_i + x_{it}\beta_t + u_{it}, \qquad \beta_t = \beta_{t-1} + e_t$$

[3]Juhl and Xiao (2003) proposed to choose $\bar{c}$ such that $\int_0^M (\varphi(c, c) - \varphi(c, \bar{c}))dc$ is minimized, where $\varphi(c, \bar{c}) = P(qLL < x; c)$ is a power function of the qLL test and $\varphi(c, c)$ is a power envelope. We can easily see that this strategy is the same as the one considered in this study.

$$\text{where} \quad \beta_0 = 1 \quad \text{and} \quad e_t \sim i.i.d.N\left(0, \frac{c^2}{T^2}Q^{-1}\right) \quad \text{with} \quad Q = Var(X_t' V^{-1} u_t).$$

(DGP2: Homogeneous-slope model, one-time break, $k = 1$)

$$y_{it} = \alpha_i + x_{it}\beta_t + u_{it}, \qquad \beta_t = \begin{cases} 1 & \text{for } t \leq 0.5T, \\ 1 + c/\sqrt{T} & \text{for } t > 0.5T. \end{cases}$$

(DGP3: Heterogeneous-slope model, time-varying parameter, $k = 1$)

$$y_{it} = \alpha_i + x_{it}\beta_{it} + u_{it}, \qquad \beta_{it} = \beta_{i,t-1} + e_{it}$$

$$\text{where} \quad \beta_{i0} \sim i.i.d.N(1,1) \quad \text{and} \quad e_t \sim i.i.d.N\left(0, \frac{c^2}{T^2}Q^{-1}\right) \quad \text{with} \quad Q = Var(D_{xt}' V^{-1} u_t).$$

(DGP4: Heterogeneous-slope model, one-time break, $k = 1$)

$$y_{it} = \alpha_i + x_{it}\beta_{it} + u_{it}, \quad \beta_{it} = \begin{cases} \beta_{i0} & \text{for } t \leq 0.5T, \\ \beta_{i0} + c/\sqrt{T} & \text{for } t > 0.5T, \end{cases} \quad \text{where} \quad \beta_{i0} \sim i.i.d.N(1,1).$$

In each DGP, the regressor x_{it} follows a stationary AR(1) process:

$$x_{it} = \rho_i x_{i,t-1} + \xi_{it}, \quad \text{where} \quad \xi_{it} \sim i.i.d.N(0,1)$$

with $x_{i1} = \xi_{i1}/\sqrt{1-\rho_i^2}$. We set $\alpha_i \sim i.i.d.N(1,1)$ and $\rho_i \sim i.i.d.U(0.3, 0.7)$.

Also, u_t is generated by $u_t = V^{1/2}\tilde{u}_t$ with

$$V = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_N \end{bmatrix} \begin{bmatrix} 1 & \rho_v & \cdots & \rho_v \\ \rho_v & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_v \\ \rho_v & \cdots & \rho_v & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_N \end{bmatrix},$$

$\tilde{u}_t \sim i.i.d.N(0, I_N)$, $\sigma_i^2 \sim i.i.d.U(1,3)$ and $\rho_v = 0.6$.

Throughout the simulation, the values of σ_i^2 and ρ_i are fixed across replications. The number of replications is 2,000, and the nominal size is 0.05. We estimate V and Q under the assumption that u_{it} is heteroskedastic, cross-sectionally dependent, and serially uncorrelated.

In this simulation, we compare the sizes and powers of the LM test ($\widetilde{LM}_{homo}$ and $\widetilde{LM}_{hetero}$), the qLL test ($\widetilde{qLL}_{homo}$ and $\widetilde{qLL}_{hetero}$) and the sup-Wald test (sup-W_{homo} and sup-W_{hetero}). The sup-W_{homo} and sup-W_{hetero} tests are based on the homogeneous-slope and heterogeneous-slope models, respectively. We set the trimming parameter to 0.15 for sup-Wald tests. Note that the LM and qLL tests are optimal against the alternative of

time-varying parameter (DGP1, DGP3) whereas the sup-Wald test is constructed against the alternative of one-time break (DGP2, DGP4).

Table 2.2 shows the empirical sizes of the tests under DGP1 and DGP3[4]. When the true process has homogeneous slopes (DGP1), the empirical sizes of $\widetilde{LM}_{homo}$ and $\widetilde{qLL}_{homo}$ are close to the nominal one for all the cases. Because the homogeneous-slope model is a special case of the heterogeneous-slope one under the null hypothesis, we can also see that the empirical sizes of $\widetilde{LM}_{hetero}$ and $\widetilde{qLL}_{hetero}$ are close to 0.05, although they tend to be conservative. On the other hand, sup-W_{homo} and sup-W_{hetero} perform poorly, especially when N is large.

When the model has heterogeneous slopes (DGP3), the tests developed for the homogeneous-slope model severely suffer from size distortion. In fact, this tendency is explained theoretically by noting that the LM and qLL test statistics developed for the homogeneous-slope model can be shown to have the null limiting distributions that stochastically dominate those obtained in Corollary 1 (from above) and Theorem 2 (from below), which implies that the rejection frequencies of those tests tend to be greater than the significance level.[5] Thus, we have to be careful about using these tests if we need to take into account the possibility of the slope heterogeneity. On the other hand, Table 2.2 shows that the LM and qLL tests developed for the heterogeneous-slope model can control the empirical size.

Figure 2.1 shows the size-adjusted powers for the homogeneous-slope model with time-varying parameter. Under DGP1, $\widetilde{LM}_{homo}$ performs slightly better than $\widetilde{qLL}_{homo}$ when c is close to zero, but this relation is reversed when c takes values away from zero. Overall, sup-W_{homo} performs better than $\widetilde{LM}_{homo}$, but $\widetilde{qLL}_{homo}$ outperforms sup-W_{homo} when c is away from zero, except the case when $N = 50$ and $T = 100$. On the other hand, under DGP2 with one-time structural break, $\widetilde{LM}_{homo}$ has the highest power. In this case, the sup-W_{homo} performs slightly better than $\widetilde{qLL}_{homo}$.

For the heterogeneous-slope model, we can see from Figure 2.3 that there is no significant difference of power between $\widetilde{LM}_{hetero}$ and $\widetilde{qLL}_{hetero}$ tests, especially when N is large, under DGP3 with time-varying parameter. In this case, sup-W_{hetero} performs worse than $\widetilde{LM}_{hetero}$ and $\widetilde{qLL}_{hetero}$. Under DGP4 with one-time break, we can see that $\widetilde{LM}_{hetero}$ has the highest power and sup-W_{hetero} has the lowest power. As we can see from the figures, the LM test

[4]Under the null hypothesis, DGP2 and DGP4 are equivalent to DGP1 and DGP3, respectively, and thus we omit the results under DGP2 and DGP4.

[5]It can be shown that $\widetilde{LM}_{homo}$ and $-\widetilde{qLL}_{homo}$ do not diverge to infinity but are stochastically bounded even in the presence of heterogeneity in the slopes with constant parameters in the time series direction. We omit the proof to save space.

can effectively detect an abrupt break, whereas the qLL test has high power against smooth structural changes.

Noting that all the tests developed in the previous sections can be applied under the homogeneous-slope model, we can compare the powers of these tests. As expected, we can see from Figures 2.5 and 2.6 that the tests developed for the homogeneous-slope model are more powerful than those developed for the heterogeneous-slope model, because the true model has homogeneous slopes. On the other hand, as seen in Table 2.2, there is no meaning in comparing these powers for the heterogeneous-slope case because $\widetilde{LM}_{homo}$, $\widetilde{qLL}_{homo}$ and sup-W_{homo} severely suffer from size distortion.

Overall, we can see from Table 2.2 that the sup-Wald test is inappropriate for testing parameter constancy in cross-sectionally dependent panel data models or seemingly unrelated regression models with moderately-sized N. In such cases, the proposed LM and qLL tests can control the empirical size well.

From the simulation results above, we propose the following sequential testing procedure in practical analysis. First, we apply the tests based on the homogeneous-slope model ($\widetilde{LM}_{homo}$ and $\widetilde{qLL}_{homo}$). If we do not reject the null hypothesis, then we can state that the parameter β is constant across time. On the other hand, if we reject the null hypothesis, then there are two possibilities. One is that β varies across cross-sections, and the other is that β is time-varying. In this case, we should apply the tests based on heterogeneous-slope models ($\widetilde{LM}_{hetero}$ and $\widetilde{qLL}_{hetero}$). If the null hypothesis is rejected, then we conclude that β is time-varying, and if not, then β is heterogeneous but stable in the time series direction.

2.5 Conclusion

We have proposed the locally best invariant test based on Tanaka (1996) and the asymptotically point optimal test based on Elliott and Müller (2006) for parameter constancy in the time series direction in panel data models. The asymptotic critical values for both tests are obtained by numerical integration and the response surface regressions are conducted. By Monte Carlo simulations, we found that the tests based on the homogeneous-slope model perform poorly when the true model has heterogeneous slopes, while we can control the tests based on the heterogeneous-slope model for both the homogeneous- and heterogeneous-slope cases, although these tests may suffer from loss of power if the true model has homogeneous slopes.

When the errors are serially correlated, we need to consistently estimate the long-run

variance to preserve the optimality of our tests under the assumption of the local-to-zero variance. In this case, other tests such as those based on the fixed-b asymptotics by Kiefer and Vogelsang (2005) and the self-normalization based test by Shao and Zhang (2010) may better control the empirical size, although they sacrifice optimality. In this case, our optimality tests may still be useful because those asymptotic powers can be seen as a benchmark. The development of a test with good finite sample property is our ongoing research.

The tests developed in this study may be extended to various directions. For example, our tests in this chapter are applicable only in the case where T is greater than N, but tests with large N and moderately sized T are also useful in practical analysis. We may generalize the model by allowing for dynamics, or factor structures to model cross-section dependence. In addition, it would be useful to develop tests for parameter constancy in both time series and cross-sectional directions. These are left for future studies.

2.6 Appendix

Lemma 1

(i) The t-th block element of $M_Z u$ is given by

$$\tilde{u}_t - \tilde{D}_{xt}\left(\sum_{t=1}^{T}\tilde{D}_{xt}'V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^{T}\tilde{D}_{xt}'V^{-1}u_t, \tag{2.23}$$

where $\tilde{D}_{xt} = D_{xt} - T^{-1}\sum_{s=1}^{T} D_{xs}$ and $\tilde{u}_t = u_t - T^{-1}\sum_{s=1}^{T} u_s$.

(ii) The t-th block element of $M_Z v$ is given by

$$\tilde{v}_t - \tilde{D}_{xt}\left(\sum_{t=1}^{T}\tilde{D}_{xt}'V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^{T}\tilde{D}_{xt}'V^{-1}v_t, \tag{2.24}$$

where $\tilde{v}_t = v_t - T^{-1}\sum_{s=1}^{T} v_s$.

Proof of Lemma 1

(i) Let $\tilde{D} = \begin{bmatrix} I_N & 0 \\ -\bar{D}_x' & I_{kN} \end{bmatrix}$, where $\bar{D}_x = T^{-1}\sum_{t=1}^{T} D_{xt}$. Then,

$$\begin{aligned} M_Z &= I_{NT} - Z\left(Z'(I_T \otimes V^{-1})Z\right)^{-1} Z'(I_T \otimes V^{-1}) \\ &= I_{NT} - Z\tilde{D}'\left(\tilde{D}Z'(I_T \otimes V^{-1})Z\tilde{D}'\right)^{-1}\tilde{D}Z'(I_T \otimes V^{-1}). \end{aligned}$$

By simple algebra, we obtain

$$
\begin{aligned}
& Z\tilde{D}'\left(\tilde{D}Z'(I_T\otimes V^{-1})Z\tilde{D}'\right)^{-1}\tilde{D}Z'(I_T\otimes V^{-1})u \\
= & \begin{bmatrix} \bar{u}+\tilde{D}_{x1}\left(\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}u_t \\ \bar{u}+\tilde{D}_{x2}\left(\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}u_t \\ \vdots \\ \bar{u}+\tilde{D}_{xT}\left(\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}u_t \end{bmatrix}.
\end{aligned}
$$

Thus, the t-th block element of $M_Z u$ is given by (2.23).

(ii) This proof is the same as (i). ■

Proof of Theorem 1

First, we derive the limiting distribution of the $[Tr]$-th block of

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}(L'\otimes I_{kN})D'_{dx}(I_T\otimes V^{-1})M_Z y \\
= & \frac{1}{\sqrt{T}}(L'\otimes I_{kN})D'_{dx}(I_T\otimes V^{-1})M_Z u + \frac{1}{\sqrt{T}}(L'\otimes I_{kN})D'_{dx}(I_T\otimes V^{-1})M_Z v. \quad (2.25)
\end{aligned}
$$

Let us consider the first term in (2.25). By Lemma 1 (i), we obtain

$$
\begin{aligned}
& (L'\otimes I_{kN})D'_{dx}(I_T\otimes V^{-1})M_Z u \\
= & \begin{bmatrix} \sum_{t=1}^T D'_{xt}V^{-1}\tilde{u}_t - \sum_{t=1}^T D'_{xt}V^{-1}\tilde{D}_{xt}\left(\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}u_t \\ \sum_{t=1}^{T-1} D'_{xt}V^{-1}\tilde{u}_t - \sum_{t=1}^{T-1} D'_{xt}V^{-1}\tilde{D}_{xt}\left(\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}u_t \\ \vdots \\ D'_{x1}V^{-1}\tilde{u}_1 - D'_{x1}V^{-1}\tilde{D}_{x1}\left(\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^T \tilde{D}'_{xt}V^{-1}u_t \end{bmatrix}.
\end{aligned}
$$

Then, the $[Tr]$-th block of $(L'\otimes I_{kN})D'_{dx}(I_T\otimes V^{-1})M_Z u$ can be expressed as

$$
\begin{aligned}
& \sum_{t=[Tr]}^{T} D'_{xt}V^{-1}\tilde{u}_t - \sum_{t=[Tr]}^{T} D'_{xt}V^{-1}\tilde{D}_{xt}\left(\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}u_t \\
= & \sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{u}_t - \sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{D}_{xt}\left(\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}u_t \\
= & \left[\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}u_t - \sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{D}_{xt}\left(\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^{T} D'_{xt}V^{-1}u_t\right] \\
& - \left[\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1} - \sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{D}_{xt}\left(\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^{T} D'_{xt}V^{-1}\right]\bar{u} \\
= & C_T(r) - D_T(r), \quad \text{say.}
\end{aligned}
$$

First, consider the term $C_T(r)$. By Assumption A7, we can see that $T^{-1}\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{D}_{xt} \xrightarrow{p} r\tilde{Q}$ holds uniformly in r, where $\tilde{Q} = E(\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt})$. Therefore, by Assumption A8(a),

$$\begin{aligned}
\frac{1}{\sqrt{T}}C_T(r) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}u_t - \left[\frac{1}{T}\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{D}_{xt}\right. \\
&\qquad \left. \cdot \left(\frac{1}{T}\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} D'_{xt}V^{-1}u_t\right] + o_p(1) \\
&\xrightarrow{d} Q^{1/2}W_1(r) - (r\tilde{Q})\tilde{Q}^{-1}Q^{1/2}W_1(1) \\
&= Q^{1/2}\left(W_1(r) - rW_1(1)\right) = B_1(r). \qquad (2.26)
\end{aligned}$$

Also,

$$\begin{aligned}
\frac{1}{\sqrt{T}}D_T(r) &= \left[\frac{1}{T}\sum_{s=1}^{[Tr]-1} D'_{xs}V^{-1} - \left\{\frac{1}{T}\sum_{s=1}^{[Tr]-1} D'_{xs}V^{-1}\tilde{D}_{xs}\right.\right. \\
&\qquad \left.\left. \cdot \left(\frac{1}{T}\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} D'_{xt}V^{-1}\right\}\right]\left(\sqrt{T}\bar{u}\right) \\
&= o_p(1)\cdot O_p(1) = o_p(1) \quad \text{uniformly in } r. \qquad (2.27)
\end{aligned}$$

Similarly, the $[Tr]$-th block of $(L'\otimes I_{kN})D'_{dx}(I_T\otimes V^{-1})M_Z v$ is given by

$$\begin{aligned}
&\left[\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}v_t - \sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{D}_{xt}\left(\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^{T} D'_{xt}V^{-1}v_t\right] \\
&- \left[\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1} - \sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{D}_{xt}\left(\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\sum_{t=1}^{T} D'_{xt}V^{-1}\right]\bar{v} \\
= \; & E_T(r) - F_T(r).
\end{aligned}$$

By Assumption A8(b), we have

$$\frac{1}{\sqrt{T}}E_T(r) \xrightarrow{d} cQ\Sigma_e^{1/2}\left(\int_0^r W_2(s)ds - r\int_0^1 W_2(s)ds\right) = cB_2(r). \qquad (2.28)$$

Next, consider the term $F_T(r)$. Define $\varepsilon_t = (\rho\Sigma_e)^{-1/2}e_t$. Then ε_t is an i.i.d. sequence with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t\varepsilon'_t) = I_{kN}$. Since

$$\sqrt{T}\bar{v} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_t = \frac{c}{T^{3/2}}\sum_{t=1}^{T} D_{xt}\Sigma_e^{1/2}\sum_{s=1}^{t}\varepsilon_s = O_p(1),$$

where we used the relation $\rho = c^2/T^2$, we obtain the following result:

$$\begin{aligned}\frac{1}{\sqrt{T}}F_T(r) &= \left[\frac{1}{T}\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1} - \left\{\frac{1}{T}\sum_{t=1}^{[Tr]-1} D'_{xt}V^{-1}\tilde{D}_{xt}\right.\right. \\ &\qquad \left.\left. \cdot \left(\frac{1}{T}\sum_{t=1}^{T}\tilde{D}'_{xt}V^{-1}\tilde{D}_{xt}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}D'_{xt}V^{-1}\right\}\right]\left(\sqrt{T}\bar{v}\right) \\ &= o_p(1)\cdot O_p(1) = o_p(1) \quad \text{uniformly in } r. \end{aligned} \tag{2.29}$$

Thus, from (2.26)–(2.29), the $[Tr]$-th block of $T^{-1/2}(L'\otimes I_{kN})D'_{dx}(I_T\otimes V^{-1})M_Z y$ weakly converges to

$$\frac{1}{\sqrt{T}}\left(C_T(r) - D_T(r) + E_T(r) - F_T(r)\right) \xrightarrow{d} B_1(r) + cB_2(r).$$

Finally, by the continuous mapping theorem, we obtain

$$\begin{aligned} LM_{hetero} &= \frac{1}{T^2}y'M_Z(I_T\otimes V^{-1})D_{dx}(L\otimes I_{kN})(I_T\otimes\Sigma_e)(L'\otimes I_{kN})(I_T\otimes V^{-1})D'_{dx}M_Z y \\ &\xrightarrow{d} \int_0^1 \left(B_1(r)+cB_2(r)\right)'\Sigma_e\left(B_1(r)+cB_2(r)\right)dr \\ &= \int_0^1 V_{kN}(r;c)'V_{kN}(r;c)dr \end{aligned}$$

as $T\to\infty$.■

Proof of the consistency of $\hat{\sigma}_{ij}$

Let us stack equation (2.2) for $t = 1,\cdots,T$ with a fixed i such that

$$\begin{aligned} y_i &= Z_i\gamma_i + D_i(L\otimes I_k)e_i + u_i \\ &= Z_i\gamma_i + v_i + u_i, \end{aligned}$$

where $y_i = [y_{i1}, y_{i2},\cdots, y_{iT}]'$, $Z_i = [\iota_T, X_i]$ with $\iota_T = [1,1,\cdots,1]'$, $X_i = [x_{i1}, x_{i2},\cdots, x_{iT}]'$, $D_i = \text{diag}\{x'_{i1}, x'_{i2},\cdots, x'_{iT}\}$, $u_i = [u_{i1}, u_{i2},\cdots, u_{iT}]'$, $v_i = [v_{i1}, v_{i2},\cdots, v_{iT}]'$, $\gamma_i = [\alpha_i, \beta'_i]'$, $v_i = D_i(L\otimes I_k)e_i$ and $e_i = [e_{i1}, e_{i2},\cdots, e_{iT}]'$. Then, the OLS residual $\hat{w}_i$ can be expressed as

$$\hat{w}_i = M_i y_i = M_i(u_i + v_i),$$

where $M_i = I_T - Z_i(Z'_iZ_i)^{-1}Z'_i$. Hence, $\hat{\sigma}_{ij}$ can be rewritten as

$$\begin{aligned} \hat{\sigma}_{ij} &= \frac{1}{T}y'_iM_iM_jy_j \\ &= \frac{1}{T}u'_iM_iM_ju_j + \frac{1}{T}v'_iM_iM_jv_j + \frac{1}{T}u'_iM_iM_jv_j + \frac{1}{T}v'_iM_iM_ju_j \\ &= J_{1T} + J_{2T} + J_{3T} + J_{4T}, \quad \text{say.} \end{aligned} \tag{2.30}$$

First, we show that $J_{1T} \xrightarrow{p} \sigma_{ij}$. The term J_{1T} can be expressed as

$$\begin{aligned} J_{1T} &= \frac{1}{T}u_i'u_j - \frac{1}{T}u_i'Z_i\left(\frac{1}{T}Z_i'Z_i\right)^{-1}\frac{1}{T}Z_i'u_j - \frac{1}{T}u_i'Z_j\left(\frac{1}{T}Z_j'Z_j\right)^{-1}\frac{1}{T}Z_j'u_j \\ &\quad + \frac{1}{T}u_i'Z_i\left(\frac{1}{T}Z_i'Z_i\right)^{-1}\frac{1}{T}Z_i'Z_j\left(\frac{1}{T}Z_j'Z_j\right)^{-1}\frac{1}{T}Z_j'u_j. \end{aligned} \tag{2.31}$$

The first term in (2.31) converges in probability to σ_{ij} by the weak law of large numbers, while the other terms are $O_p(T^{-1})$. Therefore, we have $J_{1T} \xrightarrow{p} \sigma_{ij}$.

Next, J_{2T} and J_{3T} can be expressed as

$$\begin{aligned} J_{2T} &= \frac{1}{T}v_i'v_j - \frac{1}{T}v_i'Z_i\left(\frac{1}{T}Z_i'Z_i\right)^{-1}\frac{1}{T}Z_i'v_j - \frac{1}{T}v_i'Z_j\left(\frac{1}{T}Z_j'Z_j\right)^{-1}\frac{1}{T}Z_j'v_j \\ &\quad + \frac{1}{T}v_i'Z_i\left(\frac{1}{T}Z_i'Z_i\right)^{-1}\frac{1}{T}Z_i'Z_j\left(\frac{1}{T}Z_j'Z_j\right)^{-1}\frac{1}{T}Z_j'v_j, \end{aligned} \tag{2.32}$$

$$\begin{aligned} J_{3T} &= \frac{1}{T}u_i'v_j - \frac{1}{T}u_i'Z_i\left(\frac{1}{T}Z_i'Z_i\right)^{-1}\frac{1}{T}Z_i'v_j - \frac{1}{T}u_i'Z_j\left(\frac{1}{T}Z_j'Z_j\right)^{-1}\frac{1}{T}Z_j'v_j \\ &\quad + \frac{1}{T}u_i'Z_i\left(\frac{1}{T}Z_i'Z_i\right)^{-1}\frac{1}{T}Z_i'Z_j\left(\frac{1}{T}Z_j'Z_j\right)^{-1}\frac{1}{T}Z_j'v_j. \end{aligned} \tag{2.33}$$

Since

$$\begin{aligned} \left|\frac{1}{T}v_i'v_j\right| &= \left|\frac{1}{T}\sum_{t=1}^{T}\left\{\left(\sum_{s=1}^{t}e_{is}\right)'x_{it}x_{jt}'\left(\sum_{s=1}^{t}e_{is}\right)\right\}\right| \\ &\le \left(\sup_t\left\|\sum_{s=1}^{t}e_{is}\right\|\right)^2\cdot\left\|\frac{1}{T}\sum_{t=1}^{T}x_{it}x_{jt}'\right\| \\ &= O_p(T^{-1})\cdot O_p(1) = O_p(T^{-1}), \end{aligned}$$

$\left\|T^{-1}v_i'Z_j\right\| = O_p(T^{-1/2})$ and $\left|T^{-1}v_i'u_j\right| = O_p(T^{-1})$, we have $J_{2T} \xrightarrow{p} 0$ and $J_{3T} \xrightarrow{p} 0$. We can prove $J_{4T} \xrightarrow{p} 0$ similarly.

By using these results above, we have $\hat{\sigma}_{ij} = J_{1T} + J_{2T} + J_{3T} + J_{4T} \xrightarrow{p} \sigma_{ij}$. ■

Lemma 2 *Let $v = [v_1, v_2, \cdots, v_T]'$ be a $T \times 1$ random vector such that $T^{-1/2}\sum_{t=1}^{[Tr]} v_t \xrightarrow{d} W_1(r) + c\int_0^r W_2(s)ds$, where $W_1(r)$ and $W_2(r)$ are independent scalar standard Brownian motions. Then,*

$$\begin{aligned} v'(G_{\bar{c}} - M_e)v \xrightarrow{d} &\left[-\bar{c}L(1)^2 - \bar{c}^2\int_0^1 L(s)^2ds - \frac{2\bar{c}}{1-e^{2\bar{c}}}\left(e^{-\bar{c}}L(1) + \bar{c}\int_0^1 e^{-\bar{c}s}L(s)ds\right)^2\right. \\ &\left. + \left(L(1) + \bar{c}\int_0^1 L(s)ds\right)^2\right], \end{aligned}$$

$$
\begin{aligned}
\textit{where} \qquad L(r) &= J(r) + cK(r), \\
J(r) &= W_1(r) - \bar{c}\int_0^r e^{-\bar{c}(r-s)} W_1(s)ds, \\
K(r) &= \int_0^r W_2(s)ds - \bar{c}\int_0^r e^{-\bar{c}(r-s)} \left(\int_0^s W_2(\lambda)d\lambda \right) ds.
\end{aligned}
$$

Proof of Lemma 2

Following Lemma 6 in Elliott and Müller (2006), we have

$$
\begin{aligned}
& v'(G_{\bar{c}} - M_e)v \\
= \;& v'(H_{\bar{c}}^{-1} - I_T)v - v'H_{\bar{c}}^{-1}\iota(\iota'H_{\bar{c}}^{-1}\iota)^{-1}\iota'H_{\bar{c}}^{-1}v + (T^{-1/2}\iota'v)^2 \\
= \;& (r_{\bar{c}} - 1)B_T^2 - (1 - r_{\bar{c}})^2 B_{-1}'B_{-1} + (T^{-1}\iota'H_{\bar{c}}^{-1}\iota)^{-1} \left\{ T(1 - r_{\bar{c}})T^{-3/2}\sum_{t=1}^{T-1} r_{\bar{c}}^t B_t + r_{\bar{c}}^t T^{-1/2} B_T \right\}^2,
\end{aligned}
$$

where $B = [B_1, B_2, \cdots, B_T]'$, $B_{-1} = [0, B_1, \cdots, B_{T-1}]'$, and $B_t = \sum_{s=1}^t r_{\bar{c}}^{t-s} v_s$.

By the following joint convergence

$$
\begin{aligned}
\frac{1}{\sqrt{T}} B_T &\xrightarrow{d} L(1), && (2.34) \\
\frac{1}{T^2}\sum_{t=1}^{T} B_{t-1}^2 &\xrightarrow{d} \int_0^1 L(s)^2 ds, && (2.35) \\
\frac{1}{T^{3/2}}\sum_{t=1}^{T-1} r_{\bar{c}}^t B_t &\xrightarrow{d} \int_0^1 e^{-\bar{c}s} L(s)ds, && (2.36) \\
\frac{1}{T^{3/2}}\sum_{t=1}^{T} B_{t-1} &\xrightarrow{d} \int_0^1 L(s)ds, && (2.37) \\
r_{\bar{c}}^T &\longrightarrow e^{-\bar{c}}, && (2.38)
\end{aligned}
$$

and the continuous mapping theorem, we obtain the desired result.

To prove (2.34), we first decompose $v_t = u_{1t} + c\, u_{2t}$ such that $T^{-1/2}\sum_{t=1}^{[Tr]} u_{1t} \xrightarrow{d} W_1(r)$ and $T^{-1/2}\sum_{t=1}^{[Tr]} u_{2t} \xrightarrow{d} \int_0^r W_2(s)ds$ hold jointly. Then,

$$
\begin{aligned}
\frac{1}{\sqrt{T}} B_{[Tr]} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} r_{\bar{c}}^{T-t} u_{1t} + c \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} r_{\bar{c}}^{T-t} u_{2t} \\
&= A_{1T}(r) + c \cdot A_{2T}(r), \quad \text{say.}
\end{aligned}
$$

Since $r_{\bar{c}} = 1 - \bar{c}/T + o(T^{-1})$, we have

$$\begin{aligned}
A_{1T}(r) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} r_{\bar{c}}^{T-t} u_{1t} \\
&= r_{\bar{c}}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_{1t} - \frac{\bar{c}}{T^{3/2}} \sum_{t=1}^{[Tr]} r_{\bar{c}}^{[Tr]-t-1} \left(\sum_{s=1}^{t} u_{1s} \right) + o_p(1) \\
&\xrightarrow{d} W_1(r) - \bar{c} \int_0^r e^{-\bar{c}(r-s)} W_1(s) ds, \\
A_{2T}(r) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} r_{\bar{c}}^{T-t} u_{2t} \\
&= r_{\bar{c}}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_{2t} - \frac{\bar{c}}{T^{3/2}} \sum_{t=1}^{[Tr]} r_{\bar{c}}^{[Tr]-t-1} \left(\sum_{s=1}^{t} u_{2s} \right) + o_p(1) \\
&\xrightarrow{d} \int_0^r W_2(s) ds - \bar{c} \int_0^r e^{-\bar{c}(r-s)} \left(\int_0^s W_2(\lambda) d\lambda \right) ds.
\end{aligned}$$

Therefore, $T^{-1/2} B_{[Tr]} \xrightarrow{d} L(r)$ holds. We obtain the result (2.34) by substituting $r = 1$.

(2.35)–(2.37) can also be proved by the continuous mapping theorem. (2.38) is obvious. ■

Proof of Theorem 2

This can be proved by following Elliott and Müller (2006). By Assumption A8, the sum of the first $[Tr]$ blocks of $T^{-1/2}(I_T \otimes Q^{-1/2}) D'_{dx}(I_T \otimes V^{-1}) M_Z y$ converges in distribution to

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \left[(I_T \otimes Q^{-1/2}) D'_{dx} (I_T \otimes V^{-1}) M_Z y \right]_{[t]} \\
&\xrightarrow{d} W_1(r) - r W_1(1) + c \left(\int_0^r W_2(s) ds - r \int_0^1 W_2(s) ds \right).
\end{aligned}$$

Next, for any choice of scalar q_ℓ, we have

$$\hat{v}'_\ell (G_{\bar{c}} - M_e) \hat{v}_\ell = (\hat{v}_\ell + q_\ell \iota_T)' (G_{\bar{c}} - M_e)(\hat{v}_\ell + q_\ell \iota_T).$$

Here we set $q_\ell = T^{-1} \iota_{kN,\ell} Q^{-1/2} \sum_{t=1}^{T} D_{xt} V^{-1} (u_t + v_t)$ (so that $\sqrt{T} q_\ell \xrightarrow{d} W_{1,\ell}(r) + c \int_0^r W_{2,\ell}(s) ds$). Then we have

$$\frac{1}{\sqrt{T}} (\iota'_{[Tr]}, 0'_{T-[Tr]})(\hat{v}_\ell + q_\ell \iota_T) \xrightarrow{d} W_{1,\ell}(r) + c \int_0^r W_{2,\ell}(s) ds,$$

where 0_k is a $k \times 1$ vector of zeros. By Lemma 2, we obtain desired the result. ■

Proof of Theorem 3

Since the limiting distributions consist of the sum of the functionals of J independent Brownian motions for both the test statistics ($J = k$ and kN), it is sufficient to show that

$$E\left[\exp\left\{i\theta\int_0^1 (W_1(r) - rW_1(1))^2\,dr\right\}\right] = \phi_{LM}(\theta, 1) = \left[\frac{\sin\sqrt{2i\theta}}{\sqrt{2i\theta}}\right]^{-1/2}, \tag{2.39}$$

$$E\left[\exp\left(-i\theta R_1(c, \bar{c})\right)\right] = \frac{\phi_{LM}\left(\dfrac{a_1 + a_2}{2i}; 1\right)\phi_{LM}\left(\dfrac{a_1 - a_2}{2i}; 1\right)}{\phi_{LM}\left(\dfrac{-c^2}{2i}; 1\right)}, \tag{2.40}$$

but because the relation (2.39) is derived by (4.13) in Tanaka (1996), we focus on (2.40). In the following, we assume normality in u_t and e_t without loss of generality because weak convergences are established by the invariance principle.

Because (2.40) corresponds to the case of $N = 1$ and $k = 1$ in the heterogeneous-slope case ($J = 1$), we consider not a panel data model but a simple time series model with only a constant as a regressor ($x_t = 1$):

$$y_t = \alpha + \beta_t + u_t, \quad \beta_t = \beta_{t-1} + e_t, \quad t = 1, \cdots, T, \tag{2.41}$$

where we set $\beta_0 = 0$, $E[u_t^2] = V = 1$ and $E[e_t^2] = \rho\Sigma_e = \rho = 1 - c^2/T^2$ ($\Sigma_e = 1$) without loss of generality. In this case, because $y|\alpha \sim N(Z\alpha, \Sigma(\rho))$ where $Z = [1, 1, \cdots, 1]'$ and $\Sigma(\rho) = I_T + \rho LL'$ as $D_{dx} = \mathrm{diag}\{1, 1, \cdots, 1\} = I_T$, the LM test statistic (2.8) becomes

$$LM = \frac{1}{T^2} y' M_Z LL' M_Z y, \tag{2.42}$$

where $M_Z = I_T - Z(Z'Z)^{-1}Z'$. On the other hand, the Neyman-Pearson lemma tells us that the exact point optimal test rejects the null hypothesis when

$$P_T(\bar{\rho}) = y'(M_Z - \tilde{M}_Z'\Sigma(\bar{\rho})^{-1}\tilde{M}_Z)y \tag{2.43}$$

takes large values, where $\tilde{M}_Z = I_T - Z(Z'\Sigma(\bar{\rho})^{-1}Z)^{-1}Z'\Sigma(\bar{\rho})^{-1}$. In the following, we first show that $P_T(\bar{\rho})$ is numerically equivalent to $-qLL$ as long as this simple model is concerned and then show that the limiting characteristic function of $P_T(\bar{\rho})$ is expressed as the right hand side of (2.40), using the characteristic function of (2.42).

Note that for model (2.41), the qLL test statistic (2.13) becomes

$$-qLL = y' M_Z \left(M_Z - G_{\bar{c}}\right) M_Z y$$

because $Q = 1$, $\iota_{k,\ell} = \iota_{1,1} = 1$ and $M_e = M_Z$ in this case. On the other hand, since $\tilde{M}_Z'\Sigma(\bar{\rho})^{-1}\tilde{M}_Z = M_Z(\Sigma(\bar{\rho})^{-1}\tilde{M}_Z)M_Z$ from direct calculation, (2.43) becomes

$$P_T(\bar{\rho}) = y'M_Z(M_Z - \Sigma(\bar{\rho})^{-1}\tilde{M}_Z)M_Z y. \tag{2.44}$$

Thus, $P_T(\bar{\rho})$ is shown to be equal to $-qLL$ if we prove that $\Sigma(\bar{\rho})^{-1}\tilde{M}_Z = G_{\bar{c}}$. Let B_e be a $T \times (T-1)$ matrix such that $B_e'B_e = I_{T-1}$ and $B_e'Z = 0$. Then, we have

$$\begin{aligned}
\Sigma(\bar{\rho})\tilde{M}_Z &= \Sigma(\bar{\rho})^{-1} - \Sigma(\bar{\rho})^{-1}Z(Z'\Sigma(\bar{\rho})^{-1}Z)^{-1}Z'\Sigma(\bar{\rho})^{-1} \\
&= B_e(B_e'\Sigma(\bar{\rho})B_e)^{-1}B_e \\
&= G_{\bar{c}},
\end{aligned}$$

where the second equality holds by Lemma 9.1 in Tanaka (1996) while the last equality is proved by Lemma 4(ii) of Elliott and Müller (2006). Therefore, we can see that $P_T(\bar{\rho}) = -qLL$, which implies that $P_T(\bar{\rho}) \xrightarrow{d} -R_1(c, \bar{c})$ under the local alternative by Theorem 2.

Next, we derive the characteristic function of $P_T(\bar{\rho})$ by following Kurozumi (2003). Let H be a $T \times (T-1)$ matrix such that $H'H = I_{T-1}$, $HH' = M_Z$ and $H'LL'H$ is a diagonal matrix given by $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_{T-1}\}$, the existence of which is proved by Patterson and Thompson (1971). Because it can be shown that $\Sigma(\bar{\rho})\tilde{M}_Z$ and $H(I_{T-1}+\bar{\rho}\Lambda)^{-1}H'$ are the Moore-Penrose (MP) inverse of $M_Z\Sigma(\bar{\rho})M_Z$, we have $\Sigma(\bar{\rho})\tilde{M}_Z = H(I_{T-1}+\bar{\rho}\Lambda)^{-1}H'$ because of the uniqueness of the MP inverse. From this relation, (2.44) becomes

$$\begin{aligned}
P_T(\bar{\rho}) &= y'HH'\left[HH' - H(I_{T-1}+\bar{\rho}\Lambda)^{-1}H'\right]HH'y \\
&= y'H\left[I_{T-1} - (I_{T-1}+\bar{\rho}\Lambda)^{-1}\right]H'y \\
&= \sum_{j=1}^{T-1}\left(1 - \frac{1}{1+\bar{\rho}\lambda_j}\right)y_j^{*2},
\end{aligned} \tag{2.45}$$

where y_j^* is the j-th element of $y^* = H'y$. Since $y = Z\alpha + u + Le$ in our simple case, we have

$$y^* = H'u + H'Le \sim N(0, I_{T-1} + \rho\Lambda)$$

by using $H'Z = 0$, $H'H = I_{T-1}$ and $H'LL'H = \Lambda$. As a result, (2.45) becomes

$$P_T(\bar{\rho}) = \sum_{j=1}^{T-1}\left(1 - \frac{1}{1+\bar{\rho}\lambda_j}\right)(1+\rho\lambda_j)u_j^{*2} = \sum_{j=1}^{T-1}\left(1 - \frac{1}{1+\bar{c}^2\lambda_j^*}\right)(1+c^2\lambda_j^*)u_j^{*2}, \tag{2.46}$$

where $u_j^* \sim i.i.d.N(0,1)$ and $\lambda_j^* = \lambda_j/T^2$, because $\bar{\rho} = \bar{c}^2/T^2$ and $\rho = c^2/T^2$.

Similarly, we can see that

$$\begin{aligned} LM &= \frac{1}{T^2} y' HH' LL' HH' y \\ &= \frac{1}{T^2} y^* \Lambda y^* \\ &= \sum_{j=1}^{T-1} \lambda_j^* (1 + c^2 \lambda_j^*) u_j^{*2}, \end{aligned}$$

and thus, because $u_j^* \sim i.i.d.N(0,1)$, the characteristic function of the LM statistic under the null hypothesis ($c = 0$) is given by

$$\begin{aligned} \phi_{LM,T}(\theta, 1) &= E\left[\exp\left(i\theta \sum_{j=1}^{T-1} \lambda_j^* u_j^{*2}\right)\right] \\ &= \prod_{j=1}^{T-1} \left(1 - 2i\theta\lambda_j^*\right)^{1/2}. \end{aligned}$$

Because the limiting characteristic function of LM under the null hypothesis is given by (2.39), we can see that $\phi_{LM,T}(\theta, 1) \to \phi_{LM}(\theta, 1)$ by the continuity theorem. That is,

$$\prod_{j=1}^{T-1} \left(1 - 2i\theta\lambda_j^*\right)^{1/2} \to \phi_{LM}(\theta, 1). \tag{2.47}$$

On the other hand, from (2.46), the characteristic function of $P(\bar{c})$ becomes

$$\begin{aligned} E\left[\exp\left\{i\theta P(\bar{c})\right\}\right] &= \prod_{j=1}^{T-1} \left[1 - 2i\theta\left(1 - \frac{1}{1 + \bar{c}^2\lambda_j^*}\right)(1 + c^2\lambda_j^*)\right]^{-1/2} \\ &= \frac{\prod_{j=1}^{T-1} \left[1 - (a_1 + a_2)\lambda_j^*\right]^{-1/2} \prod_{j=1}^{T-1} \left[1 - (a_1 - a_2)\lambda_j^*\right]^{-1/2}}{\prod_{j=1}^{T-1} \left[1 - (-c^2)\lambda_j^*\right]^{-1/2}} \\ &= \frac{\phi_{LM,T}\left(\frac{a_1 + a_2}{2i}, 1\right) \phi_{LM,T}\left(\frac{a_1 - a_2}{2i}, 1\right)}{\phi_{LM,T}\left(\frac{-c^2}{2i}, 1\right)} \\ &\to (2.40), \end{aligned}$$

where the last convergence holds because of (2.47).■

Table 2.1: Response Surface Regressions

	a_0	a_1	a_2	a_3	a_4				
$cv_{LM}(0.90, J)$	0.041741	0.190973	0.166668	-0.056545	0.004321	-	-	-	-
$cv_{LM}(0.95, J)$	0.109538	0.245113	0.166669	-0.056155	-0.003962	-	-	-	-
$cv_{LM}(0.99, J)$	0.283569	0.346561	0.166672	-0.028624	-0.025045	-	-	-	-
	a_0	a_1	a_2	a_3	a_4	a_5	a_6		
$\bar{c}$	1.125146	-0.025592	0.000385	8.888338	-4.209289	12.564858	-8.443815	-	-
	a_0	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4
$cv_{qLL}(0.90, J)$	-449.402036	10.508070	-0.022087	88.083260	-38.644159	78.964053	-6.911589	893.915132	-669.262679
$cv_{qLL}(0.95, J)$	-444.412521	10.561831	-0.022082	93.029330	-41.234587	75.012327	-6.354824	898.031000	-675.196055
$cv_{qLL}(0.99, J)$	-431.159336	10.656991	-0.021988	105.289824	-47.347933	65.627234	-5.114713	902.719481	-685.374079

Table 2.2: Empirical Size of the Tests under DGP1 and DGP3

	$\widetilde{LM}_{homo}$			$\widetilde{qLL}_{homo}$			sup-W_{homo}		
DGP1 (homogeneous-slope)	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$
$T = 100$	0.053	0.051	0.067	0.049	0.050	0.084	0.091	0.222	0.615
$T = 200$	0.051	0.054	0.052	0.048	0.052	0.048	0.068	0.119	0.222
DGP3 (heterogeneous-slope)	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$
$T = 100$	0.736	0.889	0.995	0.954	0.999	1.000	0.849	0.968	0.999
$T = 200$	0.734	0.878	0.940	0.965	1.000	1.000	0.862	0.962	0.990
	$\widetilde{LM}_{hetero}$			$\widetilde{qLL}_{hetero}$			sup-W_{hetero}		
DGP1 (homogeneous-slope)	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$
$T = 100$	0.039	0.027	0.013	0.037	0.022	0.011	0.164	0.854	1.000
$T = 200$	0.038	0.036	0.025	0.035	0.033	0.023	0.092	0.379	0.985
DGP3 (heterogeneous-slope)	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$	$N = 10$	$N = 25$	$N = 50$
$T = 100$	0.033	0.022	0.022	0.031	0.022	0.018	0.160	0.836	1.000
$T = 200$	0.033	0.042	0.026	0.035	0.041	0.025	0.083	0.376	0.982

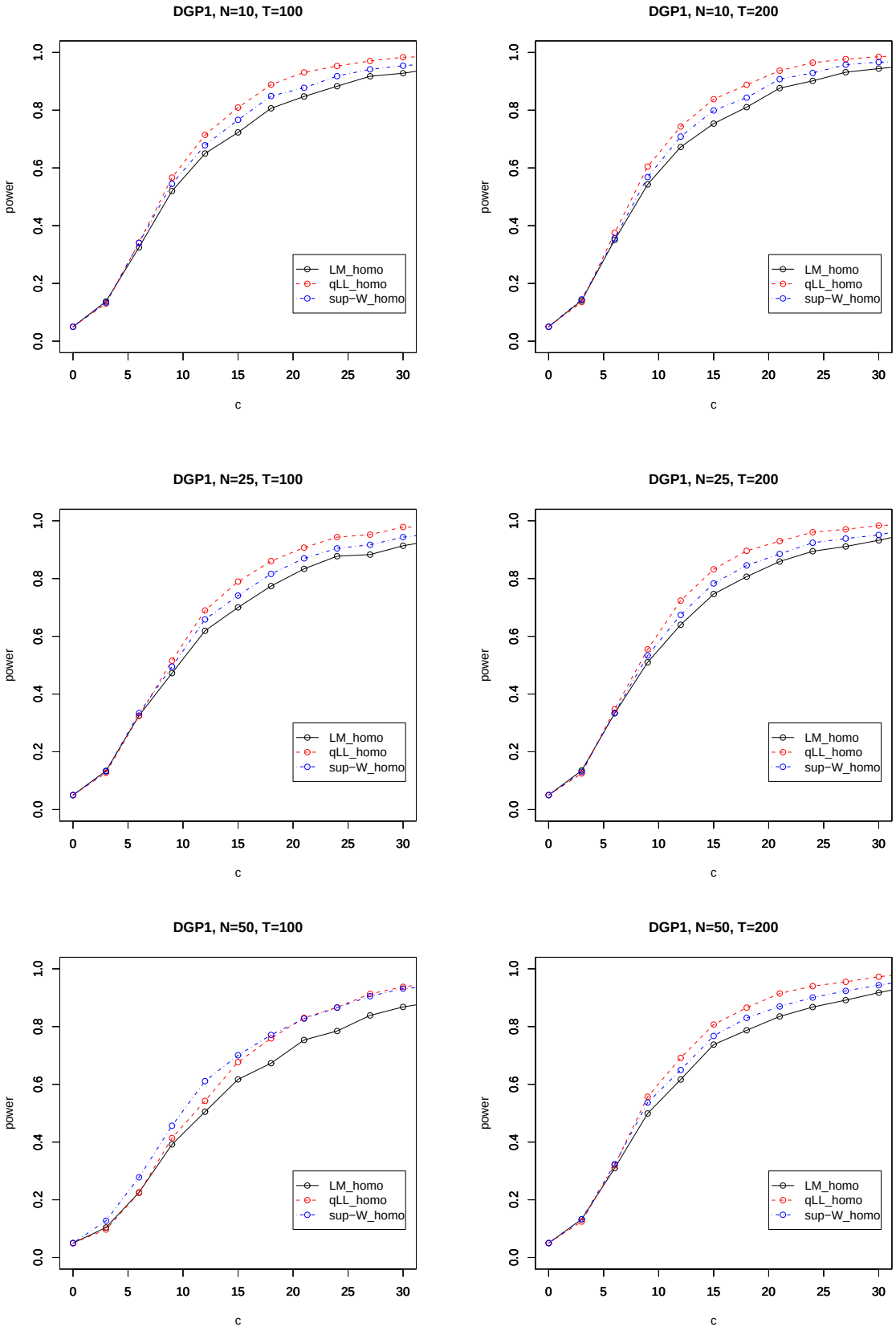

Figure 2.1: Size-adjusted power of the $\widetilde{LM}_{homo}$, $\widetilde{qLL}_{homo}$ and sup-W_{homo} tests under DGP1

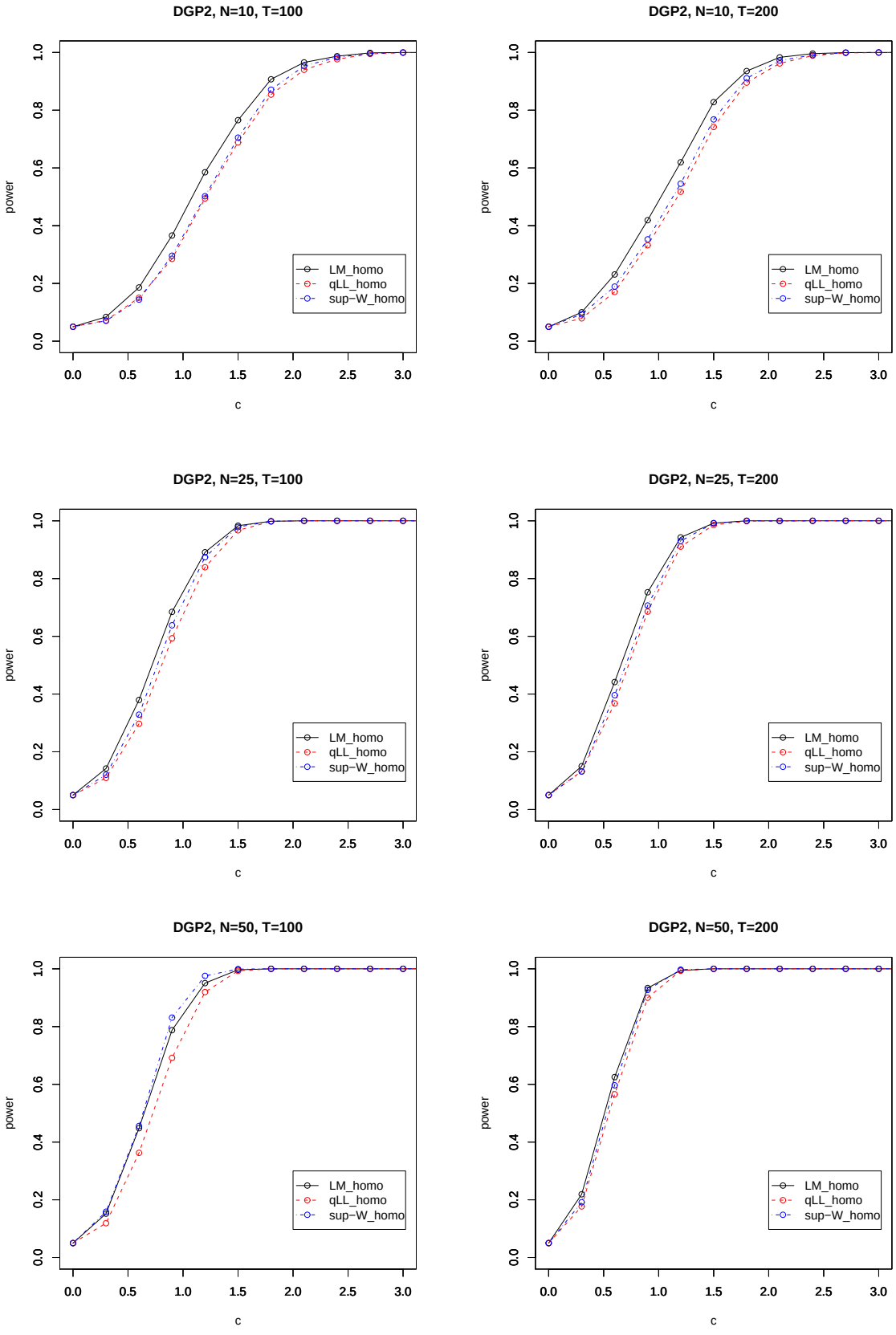

Figure 2.2: Size-adjusted power of the $\widetilde{LM}_{homo}$, $\widetilde{qLL}_{homo}$ and sup-W_{homo} tests under DGP2

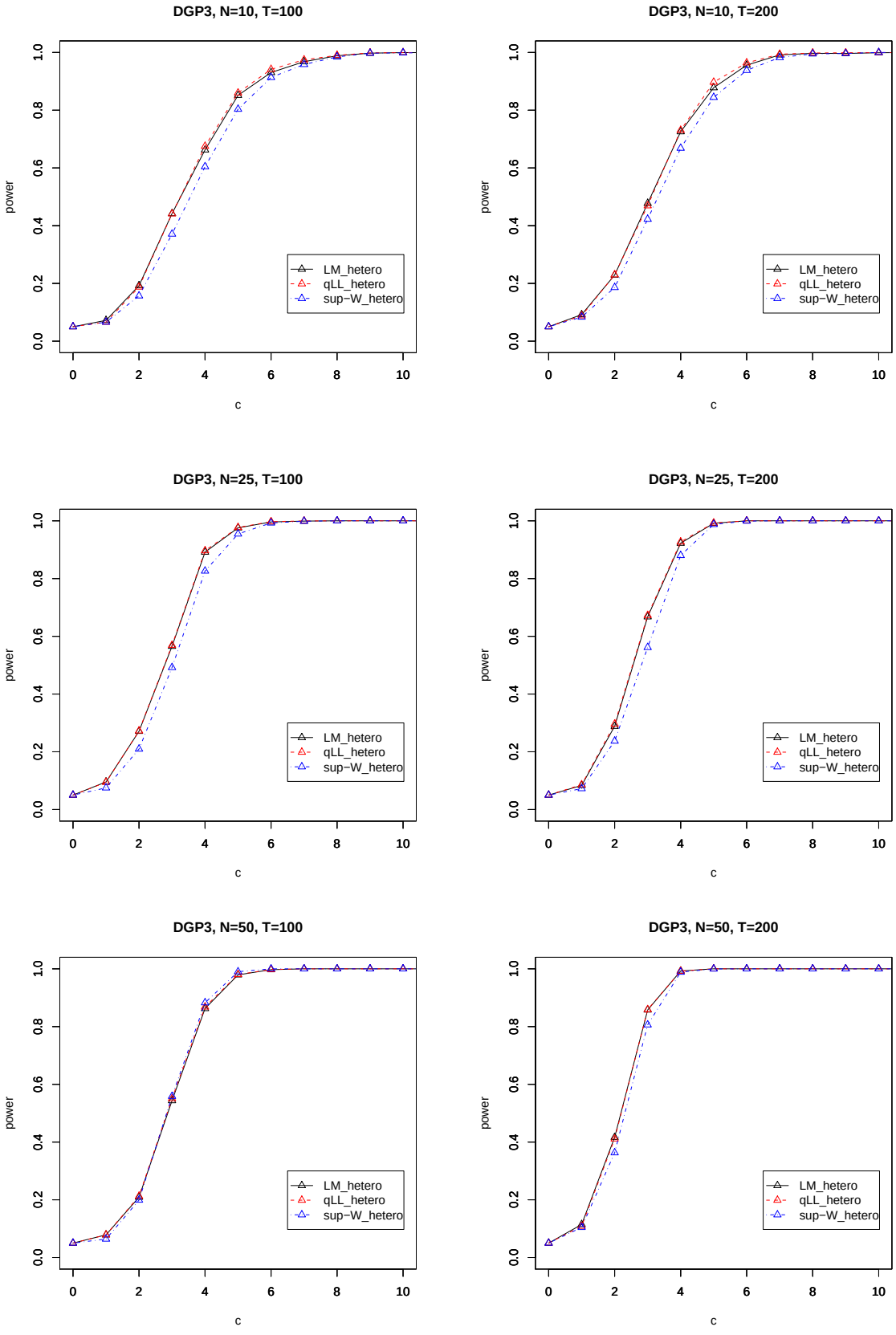

Figure 2.3: Size-adjusted power of the $\widetilde{LM}_{hetero}$, $\widetilde{qLL}_{hetero}$ and sup-W_{hetero} tests under DGP3

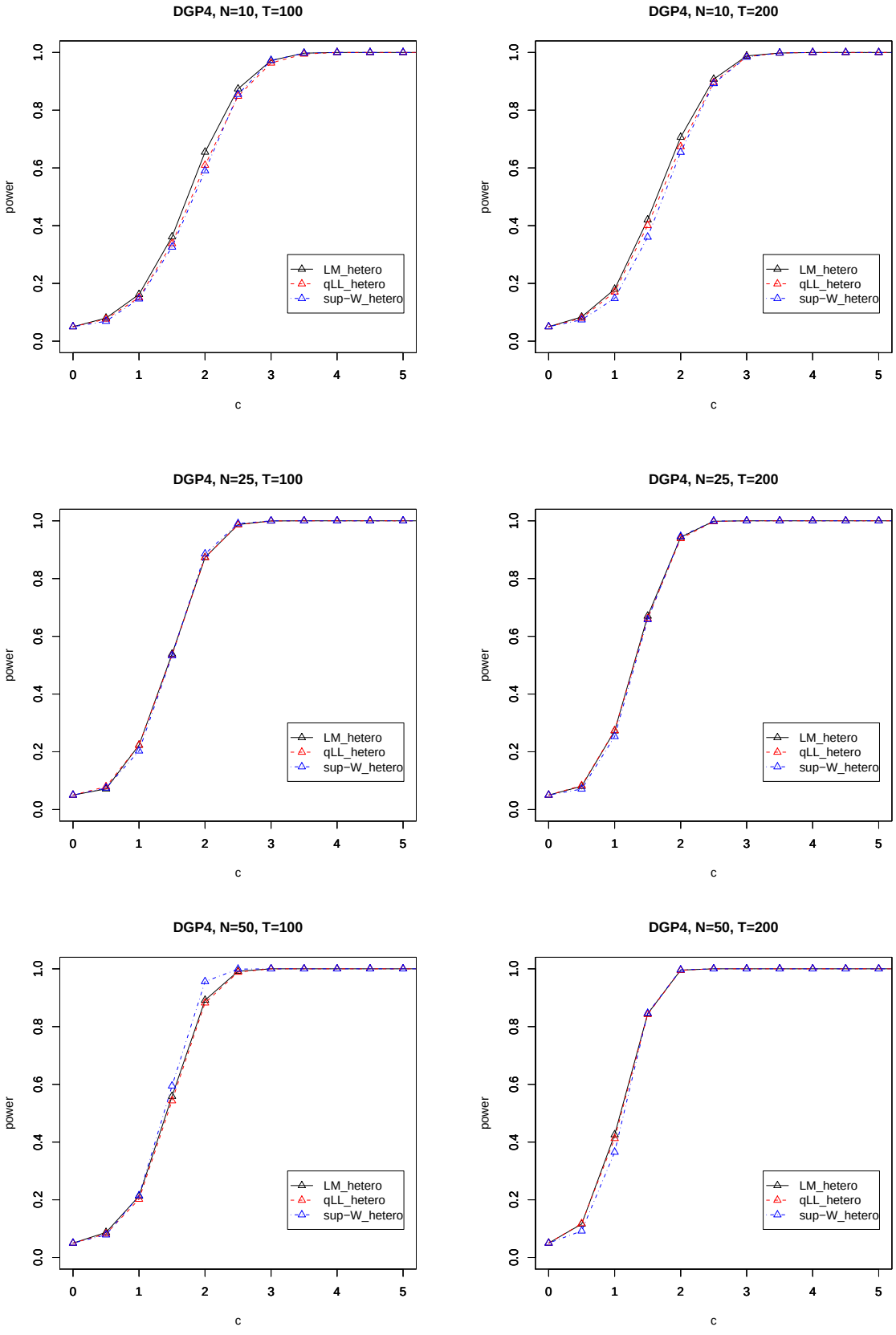

Figure 2.4: Size-adjusted power of the $\widetilde{LM}_{hetero}$, $\widetilde{qLL}_{hetero}$ and sup-W_{hetero} tests under DGP4

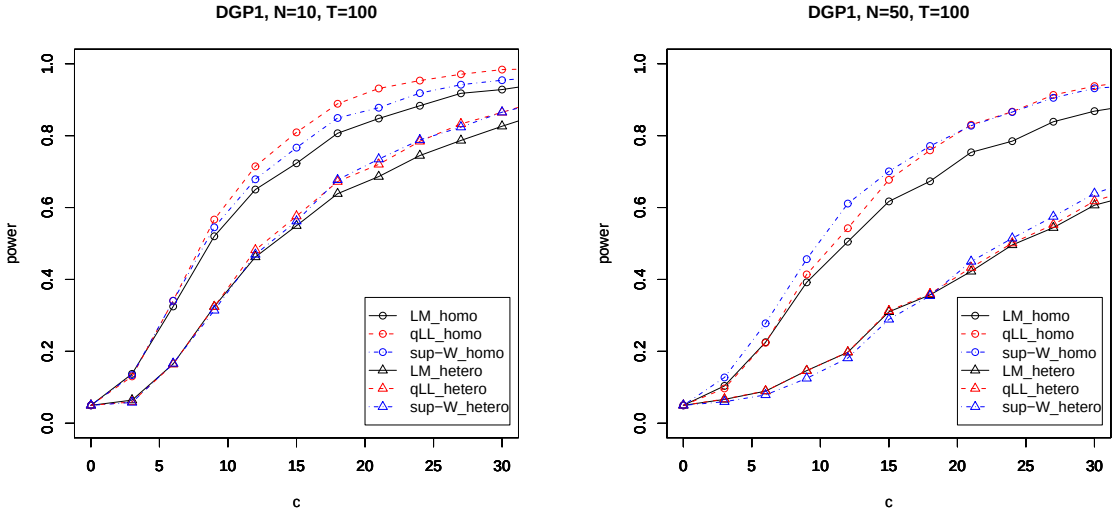

Figure 2.5: Size-adjusted power of the $\widetilde{LM}_{homo}$, $\widetilde{qLL}_{homo}$, sup-W_{homo}, $\widetilde{LM}_{hetero}$, $\widetilde{qLL}_{hetero}$ and sup-W_{hetero} tests under DGP1

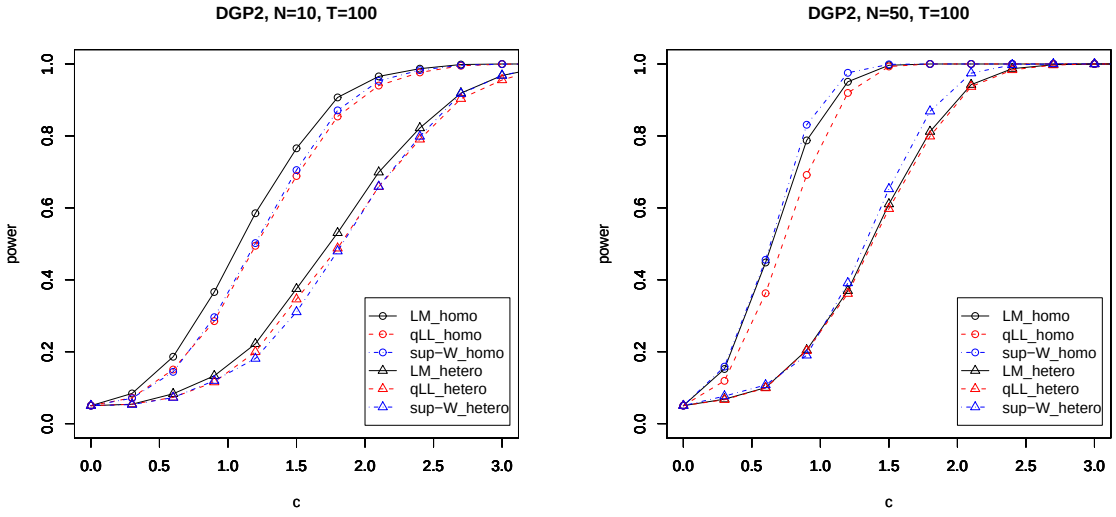

Figure 2.6: Size-adjusted power of the $\widetilde{LM}_{homo}$, $\widetilde{qLL}_{homo}$, sup-W_{homo}, $\widetilde{LM}_{hetero}$, $\widetilde{qLL}_{hetero}$ and sup-W_{hetero} tests under DGP2

Chapter 3

Bias Correction of the Long-Run Variance Estimator for Time Series Models with Structural Breaks

In this chapter, we derive the first-order bias of the long-run variance estimator in the presence of multiple structural breaks, and propose a bias-corrected long-run variance estimator. Simulation results show that the proposed long-run variance estimator has good finite sample property.

3.1 Introduction

Estimation of the long-run variance is important when we make statistical inferences in time series models with serially correlated errors. For example, in order to construct confidence intervals of parameters, we need to use a long-run variance estimator. Furthermore, when we apply hypothesis tests, we have to estimate the long-run variance for the scale adjustment. Therefore, we need to use a precise long-run variance estimator to improve the accuracy of statistical inferences.

From this point of view, the bias of the long-run variance estimator has been investigated in the literature. For example, den Haan and Levin (1997) derived the asymptotic bias of the kernel estimator and the vector autoregressive spectral density estimator. Velasco and Robinson (1999) obtained the Edgeworth expansion of the nonparametric kernel estimator. However, the existing methods do not consider the cases where structural breaks are present.

In this chapter, we derive the bias of the autoregressive spectral density estimator, taking

structural breaks into account, and we propose a bias-corrected long-run variance estimator. We find that, as the number of structural breaks increases, the downward bias of the long-run variance estimator gets larger. We find through simulations that the bias-corrected long-run variance estimator has much less bias than the one without bias correction. Also, the mean squared error of the bias-corrected estimator is comparable to that of other estimators.

This chapter is organized as follows. In Section 3.2, we introduce the model and assumptions, and the first-order bias of the long-run variance estimator is derived in Section 3.3. The bias correction method is explained in Section 3.4. In Section 3.5 we extend the results to the case with the infinite-order autoregressive errors. Section 3.6 gives the simulation results, and Section 3.7 concludes the chapter. Mathematical proofs are delegated to the appendix.

3.2 Model and Assumptions

Let us consider the following model with multiple shifts in mean:

$$y_t = \begin{cases} \mu_1 + u_t & \text{for } t = 1, \cdots, T_1, \\ \mu_2 + u_t & \text{for } t = T_1 + 1, \cdots, T_2, \\ & \vdots \\ \mu_{m+1} + u_t & \text{for } t = T_m + 1, \cdots, T, \end{cases} \tag{3.1}$$

where T_ℓ $(\ell = 1, \cdots, m)$ are the break dates and m is the number of structural breaks. We assume that u_t is a zero-mean stationary process and that the break dates are known. We need to note that, when the break dates are unknown, we need to estimate them.[1]

We are interested in estimating the long-run variance of u_t defined by $\omega = \sum_{\ell=-\infty}^{\infty} E(u_t u_{t-\ell})$.

[1]For example, we may estimate the break dates by minimizing the sum of squared residuals.

Here we use the following residuals in order to estimate ω:

$$\begin{aligned}\hat{u}_t &= \begin{cases} y_t - \bar{y}_1 & \text{for } t = 1, \cdots, T_1, \\ y_t - \bar{y}_2 & \text{for } t = T_1 + 1, \cdots, T_2, \\ & \vdots \\ y_t - \bar{y}_{m+1} & \text{for } t = T_m + 1, \cdots, T, \end{cases} \\ &= \begin{cases} u_t - \bar{u}_1 & \text{for } t = 1, \cdots, T_1, \\ u_t - \bar{u}_2 & \text{for } t = T_1 + 1, \cdots, T_2, \\ & \vdots \\ u_t - \bar{u}_{m+1} & \text{for } t = T_m + 1, \cdots, T, \end{cases} \end{aligned} \tag{3.2}$$

where $\bar{y}_\ell = (T_\ell - T_{\ell-1})^{-1} \sum_{t=T_{\ell-1}+1}^{T_\ell} y_t$, $\bar{u}_\ell = (T_\ell - T_{\ell-1})^{-1} \sum_{t=T_{\ell-1}+1}^{T_\ell} u_t$, $T_0 = 0$ and $T_{m+1} = T$.

One of the commonly used methods to estimate the long-run variance is the kernel estimator given by

$$\hat{\omega}_{kernel} = \hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} k\left(\frac{j}{m}\right) \hat{\gamma}_j, \tag{3.3}$$

where $k(\cdot)$ is the kernel function, m is the bandwidth, and $\hat{\gamma}_j$ is the estimator of the j-th autocovariance of u_t, which is defined by $\hat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}$.

Also, the autoregressive spectral density estimator of ω based on the AR(p) model is given by

$$\hat{\omega}_{AR} = \frac{\hat{\sigma}_\varepsilon^2}{\left(1 - \sum_{j=1}^{p} \hat{\phi}_j\right)^2}, \tag{3.4}$$

where $\hat{u}_t = \sum_{j=1}^{p} \hat{\phi}_j \hat{u}_{t-j} + \hat{\varepsilon}_t$ with $\hat{\phi}_j$ $(j = 1, \cdots, p)$ being the OLS estimator, and $\hat{\sigma}_\varepsilon^2 = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t^2$. In this chapter, we derive the bias of the autoregressive spectral density estimator given by (3.4).

In order to derive the bias term, we make the following assumptions when $p \geq 1$:

Assumption 1 *$\{u_t\}$ follows a zero-mean stationary AR(p) process: $u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + \varepsilon_t$, where $1 - \sum_{j=1}^{p} \phi_j z^j \neq 0$ for $|z| \leq 1$, and $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$ and $E(\varepsilon_t^3 | \mathcal{F}_{t-1}) = \kappa_3$.*

Assumption 2 *$\lim_{T\to\infty} T_i/T = \lambda_i$ and $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_m < \lambda_{m+1} = 1$.*

When $p = 0$, we use the following Assumption 1', instead of Assumption 1.

Assumption 1' *$u_t = \varepsilon_t$ for all t, where $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\varepsilon^2$.*

Assumptions 1 and 1' exclude the case where $\{u_t\}$ is a unit root process. Assumption 2 is standard for structural break models.

3.3 Derivation of the Bias

In this section, we derive the bias of the the long-run variance estimator up to $O(T^{-1})$, under the assumption that $\{u_t\}$ follows a stationary AR(p) process. The case with infinite-order autoregressive errors will be discussed later. Throughout this chapter, we define the bias as the expectation up to $O(T^{-1})$, ignoring the $o_p(T^{-1})$ terms.[2]

3.3.1 Bias of the OLS estimator of the autoregressive coefficients

First, we derive the bias of the OLS estimator of ϕ_j $(j = 1, \cdots, p)$ for $p \geq 1$, which is given by

$$\hat{\phi} = \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} & \cdots & \hat{r}_{1p} \\ \hat{r}_{21} & \hat{r}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{r}_{p-1,p} \\ \hat{r}_{p1} & \cdots & \hat{r}_{p,p-1} & \hat{r}_{pp} \end{bmatrix}^{-1} \begin{bmatrix} \hat{r}_{10} \\ \hat{r}_{20} \\ \vdots \\ \hat{r}_{p0} \end{bmatrix},$$

where $\hat{r}_{ij} = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{u}_{t-i}\hat{u}_{t-j}$.

In order to derive the bias of $\hat{\phi}$, we define the following three $(p+1) \times (p+1)$ matrices, based on Stine and Shaman (1989) and Patterson (2000):

$$\begin{aligned} B_{1p} &= \text{diag}\{0, 1, \cdots, p\}, \\ B_{2p} &= \begin{cases} \left[-e_0, -e_1, \cdots, -e_{\frac{p}{2}-1}, 0_{(p+1)\times 1}, e_{\frac{p}{2}-1}, \cdots, e_1, e_0\right] & \text{when } p \text{ is even}, \\ \left[-d_1, -d_2, \cdots, -d_{\frac{p-1}{2}}, 0_{(p+1)\times 1}, d_{\frac{p-1}{2}}, \cdots, d_1, d_0\right] & \text{when } p \text{ is odd}, \end{cases} \\ (B_{3p})_{ij} &= \begin{cases} -1 & \text{for } j < i \leq p - j + 2, \\ 1 & \text{for } p - j + 2 < i \leq j, \\ 0 & \text{otherwise}, \end{cases} \end{aligned}$$

[2] If we need to evaluate the expectation without ignoring the $o_p(T^{-1})$ terms, we have to make additional assumptions about the existence of higher-order moments.

where $0_{(p+1)\times 1}$ is a $(p+1)\times 1$ vector of zeros, e_j is a $(p+1)\times 1$ vector with ones in rows $j+3, j+5, \cdots, p+1-j$ and zeros elsewhere, and d_j is a $(p+1)\times 1$ vector with ones in rows $j+2, j+4, \cdots, p+1-j$ and zeros elsewhere. For example, $d_0 = [0,1,0,1]'$ and $d_1 = [0,0,1,0]'$ for $p=3$, while $e_0 = [0,0,1,0,1]'$ and $e_1 = [0,0,0,1,0]'$ for $p=4$.

Let $D_p^{(m)} = B_{1p} + B_{2p} + (m+1)B_{3p}$, and we divide $D_p^{(m)}$ into four blocks as follows:

$$D_p^{(m)} = \left[\begin{array}{c|c} 0_{1\times 1} & 0_{1\times p} \\ \hline -K_p^{(m)} & B_p^{(m)} \end{array}\right]. \tag{3.5}$$

where $K_p^{(m)}$ and $B_p^{(m)}$ are $p\times 1$ and $p\times p$, respectively, that is, $K_p^{(m)}$ is (-1) times the $p\times 1$ lower-left block element of $D_p^{(m)}$, and $B_p^{(m)}$ is the $p\times p$ lower-right block element of $D_p^{(m)}$. The values of $K_p^{(m)}$ and $B_p^{(m)}$ for $p=1,\cdots,5$ are given in Table 3.1.

The following theorem gives the bias of the OLS estimator $\hat{\phi}$.

Theorem 1 *Under Assumptions 1 and 2, the expectation of the OLS estimator $\hat{\phi}$ up to $O(T^{-1})$ is given by*

$$E(\hat{\phi}) = \phi - \frac{1}{T-p}\left(K_p^{(m)} + B_p^{(m)}\phi\right) + o(T^{-1}), \tag{3.6}$$

where $\phi = [\phi_1, \cdots, \phi_p]'$.

Remark 1 *The expectation of the OLS estimator without structural breaks can be obtained by letting $m=0$ in equation* (3.6).

Remark 2 *The first-order bias of the OLS estimator does not depend on the maintained break fractions λ_i $(i=1,\cdots,m)$.*

Remark 3 *When $p=1$, by Theorem 1, the expectation of the OLS estimator with m structural breaks in mean reduces to*

$$E(\hat{\phi}_1) = \phi_1 - \frac{1}{T-1}\{(m+1) + (m+3)\phi_1\} + o(T^{-1}).$$

Hence, when $\phi_1 > 0$, we can see that the downward bias of the OLS estimator gets larger as the number of structural breaks increases, which also leads to a downward bias in (3.4).

3.3.2 Bias of the long-run variance estimator

Next, we derive the bias of $\hat{\omega}_{AR}$, which is given by

$$\hat{\omega}_{AR} = \begin{cases} \dfrac{\hat{\sigma}_\varepsilon^2}{\left(1 - \sum_{j=1}^p \hat{\phi}_j\right)^2} & \text{for } p \geq 1, \\ \hat{\sigma}_\varepsilon^2 & \text{for } p = 0. \end{cases} \tag{3.7}$$

It is known that, when random variables X and Y satisfy $X - E(X) = O_p(T^{-1/2})$, $Y - E(Y) = O_p(T^{-1/2})$, $E(X) \neq 0$, and $E(Y) \neq 0$, the following relation holds:

$$E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)}\left[1 - \frac{Cov(X,Y)}{E(X)E(Y)} + \frac{Var(Y)}{\{E(Y)\}^2}\right] + o(T^{-1}), \tag{3.8}$$

which can be obtained by the Taylor expansion of $f(x,y) = x/y$ around $(x,y) = (E(X), E(Y))$, and by taking expectations, ignoring the $o_p(T^{-1})$ terms; see Mood, Graybill, and Boes (1974, p.181).

Therefore, in order to derive the bias of (3.7) up to $O(T^{-1})$, we need to obtain $E[(1 - \sum_{j=1}^p \hat{\phi}_j)^2]$, $E[\hat{\sigma}_\varepsilon^2]$, $Var[(1 - \sum_{j=1}^p \hat{\phi}_j)^2]$, and $Cov[\hat{\sigma}_\varepsilon^2, (1 - \sum_{j=1}^p \hat{\phi}_j)^2]$ for $p \geq 1$. Note that we only need $E[\hat{\sigma}_\varepsilon^2]$ when $p = 0$.

The next lemma gives the results for $p \geq 1$:

Lemma 1 *Under Assumptions 1 and 2, the following relations hold:*

(a) $E\left[\left(1 - \sum_{j=1}^p \hat{\phi}_j\right)^2\right] = (1 - \iota'\phi)^2 + \frac{1}{T-p}\left\{2(1 - \iota'\phi)\iota'\left(K_p^{(m)} + B_p^{(m)}\phi\right) + \sigma_\varepsilon^2 \iota' R^{-1}\iota\right\} + o(T^{-1}),$

(b) $E\left[\hat{\sigma}_\varepsilon^2\right] = \sigma_\varepsilon^2 - \frac{p+m+1}{T-p}\sigma_\varepsilon^2 + o(T^{-1}),$

(c) $Var\left[\left(1 - \sum_{j=1}^p \hat{\phi}_j\right)^2\right] = \frac{4}{T-p}(1 - \iota'\phi)^2\sigma_\varepsilon^2\iota' R^{-1}\iota + o(T^{-1}),$

(d) $Cov\left[\hat{\sigma}_\varepsilon^2, \left(1 - \sum_{j=1}^p \hat{\phi}_j\right)^2\right] = o(T^{-1}),$

where R is a $p \times p$ matrix whose (i,j) element is given by $\gamma_{|i-j|} = E(u_t u_{t-|i-j|})$, and ι is a $p \times 1$ vector of ones.

By (3.8) and Lemma 1, we obtain the first-order bias of the long-run variance estimator for $p \geq 1$:

Theorem 2 *Under Assumptions 1 and 2, the expectation of $\hat{\omega}_{AR}$ up to $O(T^{-1})$ is given by*

$$E(\hat{\omega}_{AR}) = \frac{\sigma_\varepsilon^2}{(1 - \iota'\phi)^2} + \frac{1}{T-p} \cdot \frac{\sigma_\varepsilon^2}{(1 - \iota'\phi)^3}\left\{\frac{3}{1 - \iota'\phi}\sigma_\varepsilon^2 \iota' R^{-1}\iota - (p+m+1)(1 - \iota'\phi) \right.$$
$$\left. -2 \cdot \iota'\left(K_p^{(m)} + B_p^{(m)}\phi\right)\right\} + o(T^{-1}).$$

Remark 4 *When $p = 1$, the expectation of $\hat{\omega}_{AR}$ is given by*

$$E(\hat{\omega}_{AR}) = \frac{\sigma_\varepsilon^2}{(1 - \phi_1)^2} - \frac{1}{T-1} \cdot \frac{\sigma_\varepsilon^2}{(1 - \phi_1)^3}\{(3m+1) + (m+1)\phi_1\} + o(T^{-1}). \tag{3.9}$$

Therefore, we can see that the first-order bias of $\hat{\omega}_{AR}$ gets larger as the number of structural breaks increases.

Similarly, when $p = 0$, we obtain the following theorem:

Theorem 2' *Under Assumptions 1' and 2, the expectation of $\hat{\omega}_{AR}$ up to $O(T^{-1})$ is given by*

$$E(\hat{\omega}_{AR}) = \sigma_\varepsilon^2 - \frac{m+1}{T}\sigma_\varepsilon^2 + o(T^{-1})$$

Remark 5 *The first-order bias of $\hat{\omega}_{AR}$ does not depend on the maintained break fractions λ_i $(i = 1, \cdots, m)$.*

3.4 Bias-Corrected Long-Run Variance Estimator

In this section, we propose the correction of the bias of (3.7) using Theorems 2 and 2'.

Since the first-order bias of (3.7) is given by Theorems 2 and 2', the bias-corrected estimator of ω is given by

$$\hat{\omega}_{AR,BC} = \hat{\omega}_{AR} - \hat{b}, \tag{3.10}$$

where

$$\hat{b} = \begin{cases} \dfrac{1}{T-p} \cdot \dfrac{\hat{\sigma}_\varepsilon^2}{(1-\iota'\hat{\phi})^3} \left\{ \dfrac{3}{1-\iota'\hat{\phi}} \hat{\sigma}_\varepsilon^2 \iota' \hat{R}^{-1} \iota - (p+m+1)(1-\iota'\hat{\phi}) \right. \\ \qquad\qquad \left. -2 \cdot \iota' \left(K_p^{(m)} + B_p^{(m)} \hat{\phi} \right) \right\} & \text{for } p \geq 1, \\ -\dfrac{m+1}{T}\, \hat{\sigma}_\varepsilon^2 & \text{for } p = 0, \end{cases}$$

and $\hat{\phi}$, $\hat{\sigma}_\varepsilon^2 = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t^2$, and $\hat{\gamma}_{ij}$ for the (i,j) element of $\hat{R}$ are the least squares estimators of ϕ, σ_ε^2, and γ_{ij}, respectively.

For example, when $p = 1$, the correcting term is given by

$$\hat{b} = -\frac{1}{T-1} \cdot \frac{\hat{\sigma}_\varepsilon^2}{(1-\hat{\phi}_1)^3} \left\{ (3m+1) + (m+1)\hat{\phi}_1 \right\}.$$

3.5 Extension to the Model with AR(∞) Errors

In this section, we consider the case where the error term u_t is generated by a stationary AR(∞) process. In this case, we make the following assumption:

Assumption 1" *$u_t = \sum_{j=1}^{\infty} \phi_j u_{t-j} + \varepsilon_t$, where $1 - \sum_{j=1}^{\infty} \phi_j z^j \neq 0$ for $|z| \leq 1$, $\sum_{j=1}^{\infty} |\phi_j| < \infty$, and $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$ and $E(\varepsilon_t^3 | \mathcal{F}_{t-1}) = \kappa_3$.*

Since the error term is an infinite order AR process, we consider estimating the long-run variance by the autoregressive spectral density estimator based on the $AR(p_T)$ model, where p_T diverges to infinity at an appropriate rate. The following assumption is concerned with the choice of the lag truncation point p_T:

Assumption L

$$\begin{aligned} &(a) \quad p_T \to \infty \quad \textit{and} \quad p_T^4/T \to 0 \quad \textit{as} \quad T \to \infty. \\ &(b) \quad \textstyle\sum_{j=p_T+1}^{\infty} |\phi_j| = o(p_T/T) \qquad \textit{as} \quad T \to \infty. \end{aligned} \tag{3.11}$$

Assumption L(a) gives the upper bound of the divergence rate of p_T. This rate guarantees the consistency of the autoregressive spectral density estimator as proved by Berk (1974) and den Haan and Levin (1998), although condition (3.11) is stronger than theirs. Assumption L(b) imposes the lower bound of p_T. This assumption is also related with the higher order summability of $\{\phi_j\}$. For example, when $\sum_{j=0}^{\infty} j^{3+\alpha}|\phi_j| < \infty$ holds and p_T is greater than $O(T^{1/(4+\alpha)})$ for some $\alpha > 0$, Assumption L(b) is satisfied. Note that this assumption is satisfied if u_t follows a stationary finite-order ARMA process and $p_T = O(T^{\delta})$ for some $\delta > 0$, because $|\phi_j|$ declines geometrically to zero.

The following theorem gives the bias of the autoregressive spectral density estimator up to $O(p_T/T)$:

Theorem 2" *Under Assumptions 1", 2, and L, the expectation of $\hat{\omega}_{AR}$ up to $O(p_T/T)$ is given by*

$$E(\hat{\omega}_{AR}) = \frac{\sigma_\varepsilon^2}{(1-\iota'\phi)^2} + \frac{1}{T-p_T} \cdot \frac{\sigma_\varepsilon^2}{(1-\iota'\phi)^3} \left\{ \frac{3}{1-\iota'\phi}\sigma_\varepsilon^2 \iota' R^{-1} \iota - (p_T+m+1)(1-\iota'\phi) \right. \\ \left. -2 \cdot \iota' \left(K_{p_T}^{(m)} + B_{p_T}^{(m)} \phi \right) \right\} + o\left(\frac{p_T}{T}\right).$$

where

$$\phi = \begin{bmatrix} \phi_{p_T,1} \\ \phi_{p_T,2} \\ \vdots \\ \phi_{p_T,p_T} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p_T-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{p_T-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{p_T} \end{bmatrix}. \tag{3.12}$$

This first-order bias is exactly the same as the one in Theorem 2. Therefore, we can implement the bias correction exactly in the same way as explained in Section 3.4.

3.6 Simulation Results

In this section, we perform simulations to investigate the finite sample performance of the long-run variance estimators. We consider the following data generating processes:
(DGP1: 1 break)

$$y_t = \begin{cases} u_t & \text{for } t = 1, \cdots, 0.5T, \\ \delta + u_t & \text{for } t = 0.5T + 1, \cdots, T. \end{cases} \tag{3.13}$$

(DGP2: 2 breaks)

$$y_t = \begin{cases} u_t & \text{for } t = 1, \cdots, 0.3T, \\ \delta + u_t & \text{for } t = 0.3T + 1, \cdots, 0.7T, \\ 2\delta + u_t & \text{for } t = 0.7T + 1, \cdots, T, \end{cases} \tag{3.14}$$

with $\delta = 1,\ 2$. In order to obtain the long-run variance estimators, we estimate the break dates by minimizing the sum of squared residuals under the assumption that the number of breaks is known. For example, when $m = 1$, we estimate the break date by $\hat{T}_1 = \arg\min_{T_1 \in [\varepsilon T, (1-\varepsilon)T]} SSR(T_1)$, where $SSR(T_1)$ is the sum of squared residuals with the break date T_1. When $m = 2$, we use $(\hat{T}_1, \hat{T}_2) = \arg\min_{T_1 \in [\varepsilon T, (1-2\varepsilon)T], T_2 \in [T_1 + \varepsilon T, (1-\varepsilon)T]} SSR(T_1, T_2)$, where $SSR(T_1, T_2)$ is the sum of squared residuals with break dates (T_1, T_2). In both cases, we set the trimming parameter $\varepsilon = 0.15$. For comparison, we also consider the cases where the break dates are known.

The error term u_t follows the following processes:

$$\begin{cases} \text{AR}(1) : u_t = \phi u_{t-1} + \varepsilon_t, & \varepsilon_t \sim i.i.d.\ N\left(0,\ (1-\phi)^2\right), \\ \text{AR}(2) : u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t, & \varepsilon_t \sim i.i.d.\ N\left(0,\ (1-\phi_1-\phi_2)^2\right), \\ \text{MA}(1) : u_t = \varepsilon_t + \theta \varepsilon_{t-1}, & \varepsilon_t \sim i.i.d.\ N\left(0,\ \frac{1}{(1+\theta)^2}\right), \end{cases}$$

where the variance of ε_t is selected so that $\omega = 1$.

In this section, we compare the biases and mean squared errors (MSEs) of the following estimators:

(i): $\hat{\omega}_{kernel}$: the kernel estimator given by (3.3).

(ii): $\hat{\omega}_{AR}$: the autoregressive spectral density estimator given by (3.4).

(iii): $\hat{\omega}_{AR,BC}$: the bias-corrected estimator given by (3.10).

For the kernel estimator (3.3), we use the quadratic spectral kernel with the bandwidth parameter selected by Andrews' (1991) rule. When we implement the AR(p) regression to

obtain the autoregressive spectral density estimator, we select the lag length p by the Bayesian Information Criterion (BIC) with the maximum lag length 5.

Table 3.2 gives the results with AR(1) errors under DGP1 (one-time break). As we can see from the table, $\hat{\omega}_{kernel}$ has large downward bias in all cases. $\hat{\omega}_{AR}$ has less bias than $\hat{\omega}_{kernel}$, except for the case where $T = 100$ and $\phi = 0.2$.[3] The bias-corrected estimator ($\hat{\omega}_{AR,BC}$) has smaller bias, compared with the other estimators. We can also see that the bias becomes smaller as the magnitude of the break increases because the break date is more precisely estimated, so that the results are quite similar to those when the break date is known. In terms of the MSE, we can see that the MSE of the bias-corrected estimator is comparable to that of the other estimators in all cases.

Table 3.3 shows the results under DGP2 (2 breaks). In this case, the downward bias of the long-run variance gets larger, compared to the case under DGP1, but we can see that the relative performance of the estimators is quite similar.

Tables 3.4–3.7 show the results with AR(2) errors. We can see that the bias-corrected estimator has better finite sample property in most cases. Tables 3.8 and 3.9 give the results with MA(1) errors. In this case, when $\theta < 0$, all estimators have large upward bias. When $\theta \geq 0$, the bias-corrected long-run variance estimator has less bias than the other estimators.

Overall, we can see from the simulation results that the bias-corrected long-run variance estimator has less bias than other estimators in most cases, so that our proposed estimator has good finite sample performance.

3.7 Conclusion

We have derived the first-order bias of the long-run variance estimator, taking structural breaks into account, and proposed a bias-corrected long-run variance estimator. By Monte Carlo simulations, we have found that our proposed long-run variance estimator has good finite sample properties. Our proposed method is useful, for example, when we construct confidence intervals of the mean of a time series with structural breaks.

In this chapter, we focused only on estimation of the long-run variance, but our method can also be applied to testing for shifts in mean of a time series. This topic will be investigated in the next chapter.

[3] When $T = 100$ and $\phi = 0.2$, the performance of $\hat{\omega}_{AR}$ gets worse because the lag length selected by the BIC is too short.

3.8 Appendix A: Proofs of Theorem 1 and Some Related Lemmas

Lemma 2 *Under Assumptions 1 and 2,*

$$E(\hat{\phi}) = \phi + R^{-1}E\left[\hat{r} - r - (\hat{R} - R)\phi\right] - R^{-1}E\left[(\hat{R} - R)R^{-1}\left\{\hat{r} - r - (\hat{R} - R)\phi\right\}\right] + o(T^{-1}), \tag{3.15}$$

where $\hat{R}$ *and* R *are* $p \times p$ *matrices such that* $(\hat{R})_{ij} = \hat{r}_{ij}$, $(R)_{ij} = r_{ij}$, $\hat{r} = [\hat{r}_{01}, \cdots, \hat{r}_{0p}]'$, $r = [r_{01}, \cdots, r_{0p}]'$, $\hat{r}_{ij} = (T-p)^{-1}\sum_{t=p+1}^{T}\hat{u}_{t-i}\hat{u}_{t-j}$, *and* $r_{ij} = E(u_{t-i}u_{t-j})$.

Proof of Lemma 2

Since $\hat{R}^{-1}$ can be expressed as

$$\hat{R}^{-1} = R^{-1} - R^{-1}(\hat{R} - R)\hat{R}^{-1}, \tag{3.16}$$

we obtain

$$\begin{aligned}\hat{R}^{-1} &= R^{-1} - R^{-1}(\hat{R} - R)R^{-1} + R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)R^{-1} \\ &\quad -R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)\hat{R}^{-1},\end{aligned} \tag{3.17}$$

by recursively using relation (3.16). Therefore, since $\phi = R^{-1}r$, $\hat{r} - r = O_p(T^{-1/2})$, and $\hat{R} - R = O_p(T^{-1/2})$, we have

$$\begin{aligned}\hat{\phi} &= \hat{R}^{-1}\hat{r} \\ &= R^{-1}r + R^{-1}(\hat{r} - r) - R^{-1}(\hat{R} - R)R^{-1}r - R^{-1}(\hat{R} - R)R^{-1}(\hat{r} - r) \\ &\quad +R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)R^{-1}r + o_p(T^{-1}) \\ &= \phi + R^{-1}\left\{\hat{r} - r - (\hat{R} - R)\phi\right\} + o_p(T^{-1}) \\ &= -R^{-1}(\hat{R} - R)R^{-1}\left\{\hat{r} - r - (\hat{R} - R)\phi\right\} + o_p(T^{-1}).\end{aligned} \tag{3.18}$$

By ignoring the $o_p(T^{-1})$ term and taking expectation of the rest of (3.18), we obtain (3.15). ■

Lemma 3 *Under Assumptions 1 and 2,*

$$E(\hat{r}_{ij} - r_{ij}) = -\frac{m+1}{T-p}\omega + o(T^{-1}),$$

where $\omega = \sigma_\varepsilon^2/(1 - \sum_{j=1}^{p}\phi_j)^2$.

Proof of Lemma 3

Without loss of generality, we assume $i \leq j$. From (3.2), we have

$$\begin{aligned}
\hat{r}_{ij} &= \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_{t-i}\hat{u}_{t-j} \\
&= \frac{T_1-p+i}{T-p}\left\{\frac{1}{T_1-p+i}\sum_{t=p+1}^{T_1+i}(u_{t-i}-\bar{u}_1)(u_{t-j}-\bar{u}_1)\right\} \\
&\quad +\sum_{\ell=1}^{m}\left\{\frac{1}{T-p}\sum_{t=T_\ell+i+1}^{T_\ell+j}(u_{t-i}-\bar{u}_{\ell+1})(u_{t-j}-\bar{u}_\ell)\right\} \\
&\quad +\sum_{\ell=2}^{m}\left\{\frac{T_\ell-T_{\ell-1}+i-j}{T-p}\cdot\frac{1}{T_\ell-T_{\ell-1}+i-j}\sum_{t=T_{\ell-1}+j+1}^{T_\ell+i}(u_{t-i}-\bar{u}_\ell)(u_{t-j}-\bar{u}_\ell)\right\} \\
&\quad +\frac{T-T_m-j}{T-p}\left\{\frac{1}{T-T_m-j}\sum_{t=T_m+j+1}^{T}(u_{t-i}-\bar{u}_{m+1})(u_{t-j}-\bar{u}_{m+1})\right\}.
\end{aligned}$$

Note that the second term in the last equation does not appear when $i=j$, and the third term does not appear when $m=1$. Therefore,

$$\begin{aligned}
E(\hat{r}_{ij}) &= \frac{T_1-p+i}{T-p}\left\{r_{ij}-\frac{1}{\lambda_1(T-p)}\omega+o(T^{-1})\right\}+\sum_{\ell=1}^{m}\left\{\frac{j-i}{T-p}r_{ij}+o(T^{-1})\right\} \\
&\quad +\sum_{\ell=2}^{m}\left[\frac{T_\ell-T_{\ell-1}+j-i}{T-p}\left\{r_{ij}-\frac{1}{(\lambda_\ell-\lambda_{\ell-1})(T-p)}\omega+o(T^{-1})\right\}\right] \\
&\quad +\frac{T-T_m-j}{T-p}\left\{r_{ij}-\frac{1}{(1-\lambda_m)(T-p)}\omega+o(T^{-1})\right\} \\
&= r_{ij}-\sum_{\ell=1}^{m+1}(\lambda_\ell-\lambda_{\ell-1})\cdot\frac{1}{(\lambda_\ell-\lambda_{\ell-1})(T-p)}\omega+o(T^{-1}) \\
&= r_{ij}-\frac{m+1}{T-p}\omega+o(T^{-1}).\blacksquare
\end{aligned}$$

Lemma 4 *Under Assumptions 1 and 2,*

$$Cov(\hat{r}_{ij},\hat{r}_{i'j'}) = Cov(\tilde{r}_{ij},\tilde{r}_{i'j'})+O(T^{-3/2}),$$

where $\tilde{r}_{ij}=(T-p)^{-1}\sum_{t=p+1}^{T}u_{t-i}u_{t-j}$.

Proof of Lemma 4

Without loss of generality, we assume $i \leq j$ and $i' \leq j'$. We can see that $\hat{r}_{ij}$ can be expressed as

$$\begin{aligned}
\hat{r}_{ij} &= \frac{1}{T-p}\left[\sum_{t=p+1}^{T_1+i}(u_{t-i}-\bar{u}_1)(u_{t-j}-\bar{u}_1)+\sum_{\ell=1}^{m}\left\{\sum_{t=T_\ell+j+1}^{T_\ell+i}(u_{t-i}-\bar{u}_{\ell+1})(u_{t-j}-\bar{u}_\ell)\right\}\right. \\
&\qquad \left.+\sum_{m=2}^{m}\left\{\sum_{t=T_{\ell-1}+j+1}^{T_\ell+i}(u_{t-i}-\bar{u}_\ell)(u_{t-j}-\bar{u}_\ell)\right\}+\sum_{t=T_m+j+1}^{T}(u_{t-i}-\bar{u}_{m+1})(u_{t-j}-\bar{u}_{m+1})\right] \\
&= \frac{1}{T-p}\left[\sum_{t=p+1}^{T_1+i}u_{t-i}u_{t-j}-\bar{u}_1\left(\sum_{t=p+1}^{T_1+i}u_{t-i}\right)-\bar{u}_1\left(\sum_{t=p+1}^{T_1+i}u_{t-j}\right)+(T_1-p+i)\bar{u}_1^2\right. \\
&\qquad +\sum_{\ell=1}^{m}\left\{\sum_{t=T_\ell+i+1}^{T_\ell+j}u_{t-i}u_{t-j}-\bar{u}_\ell\left(\sum_{t=T_\ell+i+1}^{T_\ell+j}u_{t-i}\right)-\bar{u}_{\ell+1}\left(\sum_{t=T_\ell+i+1}^{T_\ell+j}u_{t-j}\right)+(j-1)\bar{u}_\ell\bar{u}_{\ell+1}\right\} \\
&\qquad +\sum_{\ell=2}^{m}\left\{\sum_{t=T_{\ell-1}+j+1}^{T_\ell+i}u_{t-i}u_{t-j}-\bar{u}_\ell\left(\sum_{t=T_{\ell-1}+j+1}^{T_\ell+i}u_{t-i}\right)-\bar{u}_\ell\left(\sum_{t=T_{\ell-1}+j+1}^{T_\ell+i}u_{t-j}\right)\right. \\
&\qquad\qquad \left.\left.+(T_\ell-T_{\ell-1}+i-j)\bar{u}_\ell^2\right\}\right. \\
&= \left.\sum_{t=T_m+j+1}^{T}u_{t-i}u_{t-j}-\bar{u}_{m+1}\left(\sum_{t=T_m+j+1}^{T}u_{t-i}\right)-\bar{u}_{m+1}\left(\sum_{t=T_m+j+1}^{T}u_{t-j}\right)+(T-T_m-j)\bar{u}_{m+1}^2\right] \\
&= (\tilde{r}_{ij,1}-c_{ij,1}-c_{ij,2}+c_{ij,3})+(\tilde{r}_{ij,2}-c_{ij,4}-c_{ij,5}+c_{ij,6}) \\
&\qquad +(\tilde{r}_{ij,3}-c_{ij,7}-c_{ij,8}+c_{ij,9})+(\tilde{r}_{ij,4}-c_{ij,10}-c_{ij,11}+c_{ij,12}), \quad \text{say,} \\
&= \tilde{r}_{ij}+c_{ij},
\end{aligned}$$

where $c_{ij}=\sum_{n=1}^{12}c_{ij,n}$. Note that $\tilde{r}_{ij,2}$, $c_{ij,4}$, $c_{ij,5}$, and $c_{ij,6}$ do not appear when $i=j$, and that $\tilde{r}_{ij,3}$, $c_{ij,7}$, $c_{ij,8}$, and $c_{ij,9}$ do not appear when $m=1$.

Therefore,

$$\begin{aligned}
Cov(\hat{r}_{ij},\hat{r}_{i'j'}) &= Cov(\tilde{r}_{ij},\tilde{r}_{i'j'})+Cov(c_{ij},\tilde{r}_{i'j'})+Cov(\tilde{r}_{ij},c_{i'j'})+Cov(c_{ij},c_{i'j'}) \\
&= Cov(\tilde{r}_{ij},\tilde{r}_{i'j'})+d_1+d_2+d_3, \quad \text{say.}
\end{aligned}$$

First, let us consider d_1, which can be expressed as $d_1=Cov(c_{ij},\tilde{r}_{i'j'})=\sum_{n=1}^{12}Cov(c_{ij,n},\tilde{r}_{i'j'})$. For $n=1$, by Cauchy-Schwarz inequality,

$$\begin{aligned}
\left|Cov(c_{ij,1},\tilde{r}_{i'j'})\right| &\leq \left(Var\left[\left(\frac{1}{T-p}\sum_{t=p+1}^{T_1+i}u_{t-i}\right)\bar{u}_1\right]Var\left[\frac{1}{T-p}\sum_{t=p+1}^{T}u_{t-i'}u_{t-j'}\right]\right)^{1/2} \\
&= d_{11}^{1/2}d_{12}^{1/2}, \quad \text{say.}
\end{aligned}$$

Since

$$\begin{aligned} d_{11} &= Var\left[\frac{T_1}{T-p}\left(\bar{u}_1 - \frac{1}{T_1}\sum_{t=1}^{p-i} u_t\right)\bar{u}_1\right] \\ &\leq 2\left[Var\left(\frac{T_1}{T-p}\bar{u}_1^2\right) + Var\left(\left(\frac{1}{T-p}\sum_{t=1}^{p-i} u_t\right)\bar{u}_1\right)\right] = O(T^{-2}) \end{aligned}$$

and $d_{12} = O(T^{-1})$, we obtain $Cov(c_{ij,1}, \tilde{r}_{i'j'}) = O(T^{-3/2})$. Similarly, for $n = 2, \cdots, 12$, $Cov(c_{ij,n}, \tilde{r}_{i'j'})$ can be shown to be $O(T^{-3/2})$. Therefore, we have $d_1 = O(T^{-3/2})$. In the same way, $d_2 = O(T^{-3/2})$ can be proved.

Then, consider the term d_3. Since $d_3 = Cov(c_{ij}, c_{i'j'}) = \sum_{n_1=1}^{9}\sum_{n_2=1}^{9} Cov(c_{ij,n_1}, c_{i'j',n_2})$ and

$$\begin{aligned} \left|Cov(c_{ij,n_1}, c_{i'j',n_2})\right| &\leq \left(Var(c_{ij,n_1})Var(c_{i'j',n_2})\right)^{1/2} \\ &= \left(O(T^{-2})O(T^{-2})\right)^{1/2} = O(T^{-2}), \end{aligned}$$

d_3 is of order T^{-2}. Therefore, we conclude that $Cov(\hat{r}_{ij}, \hat{r}_{i'j'}) = Cov(\tilde{r}_{ij}, \tilde{r}_{i'j'}) + O(T^{-3/2})$. ■

Proof of Theorem 1

By Lemma 2,

$$\begin{aligned} E(\hat{\phi}) &= \phi + R^{-1}E\left[\hat{r} - r - (\hat{R} - R)\phi\right] - R^{-1}E\left[(\hat{R} - R)R^{-1}\left\{\hat{r} - r - (\hat{R} - R)\phi\right\}\right] + o(T^{-1}) \\ &= \phi + (A) - (B) + o(T^{-1}), \quad \text{say.} \end{aligned}$$

Since $E(\hat{r} - r) = -(T-p)^{-1} \cdot (m+1)\omega\iota + o(T^{-1})$ and $E(\hat{R} - R) = -(T-p)^{-1} \cdot (m+1)\omega\iota\iota' + o(T^{-1})$ by Lemma 3, where ι is a $p \times 1$ vector of ones, we obtain

$$\begin{aligned} (A) &= R^{-1}E\left[\hat{r} - r - (\hat{R} - R)\phi\right] \qquad (3.19) \\ &= R^{-1}\left[-\frac{m+1}{T-p}\omega\iota + \frac{m+1}{T-p}\omega\iota\iota'\phi + o(T^{-1})\right] \\ &= R^{-1}\left[-\frac{m+1}{T-p} \cdot \frac{\sigma_\varepsilon^2}{\left(1 - \sum_{j=1}^{p}\phi_j\right)^2} \cdot \iota\left(1 - \sum_{j=1}^{p}\phi_j\right) + o(T^{-1})\right] \\ &= -\frac{m+1}{T-p} \cdot \frac{\sigma_\varepsilon^2}{1 - \sum_{j=1}^{p}\phi_j} R^{-1}\iota + o(T^{-1}). \qquad (3.20) \end{aligned}$$

Note that the first-order bias of (A) is equal to $\{-(m+1)\}$ times equation (3.6) in Shaman and Stine (1988).[4] Therefore, from (5.4) in Shaman and Stine (1988), the j-th element of

[4]Note that the notation in this chapter is different from that in Shaman and Stine (1988). For example, ϕ_j $(j = 1, \cdots, p)$ corresponds to $-\alpha_j$ $(j = 1, \cdots, p)$ in Shaman and Stine (1988).

(A) is given by $(T-p)^{-1} \cdot (m+1) \sum_{i=0}^{j-1} (\phi_i - \phi_{p-i})$, where $\phi_0 = -1$, so that

$$(A) = -\frac{1}{T-p} F_p \cdot \{(m+1) B_{3p}\} \phi^* + o(T^{-1}), \tag{3.21}$$

where $F_p = [0_{p\times 1}, I_p]$, $\phi^* = [-1, \phi']'$ and B_{3p} is defined in Section 3 and Patterson (2000).

For (B), we can see from Lemma 4 that

$$(B) = R^{-1} E\left[(\tilde{R} - R) R^{-1} \left\{\tilde{r} - r - (\tilde{R} - R)\phi\right\}\right] + o(T^{-1}), \tag{3.22}$$

where $(\tilde{R})_{ij} = \tilde{r}_{ij}$ and $\tilde{r} = [\tilde{r}_{01}, \cdots, \tilde{r}_{0p}]'$. Hence, the first-order bias (B) is the same as the one in Shaman and Stine (1988), which is given by the sum of (5.1) and (5.3) in Shaman and Stine (1988). Therefore, we have

$$(B) = \frac{1}{T-p} F_p (B_{1p} + B_{2p}) \phi^* + o(T^{-1}), \tag{3.23}$$

where B_{1p} and B_{2p} are defined in Section 3 and Patterson (2000).

From (3.21) and (3.23), we obtain

$$\begin{aligned} E(\hat{\phi}) &= \phi + (A) - (B) + o(T^{-1}) \\ &= \phi - \frac{1}{T-p} F_p \left\{B_{1p} + B_{2p} + (m+1) B_{3p}\right\} \phi^* + o(T^{-1}) \\ &= \phi - \frac{1}{T-p} \left(K_p^{(m)} + B_p^{(m)} \phi\right) + o(T^{-1}). \blacksquare \end{aligned}$$

3.9 Appendix B: Proofs of Theorem 2" and Some Related Lemmas

Because the AR(p) model is a special case of the AR(∞) model, we only prove the results for AR(∞) errors. Lemma 1 and Theorems 2 and 2' can be proved similarly. Note that p_T becomes a fixed number for the finite order AR model and thus, for example, the order given by $o(p_T/T)$ in the following lemmas becomes $o(1/T)$ in the AR(p) case.

In this appendix, we use the vector norm $\|x\|_\infty = \max_{1\le i\le n} |x_i|$ for an $n \times 1$ vector $x = [x_1, \cdots, x_n]'$, and a matrix norm $\|A\|_\infty = \max_{1\le i\le n} \left(\sum_{j=1}^n |a_{ij}|\right)$ for an $n \times n$ matrix $A = (a_{ij})$. This matrix norm is sub-multiplicative, that is, $\|AB\|_\infty \le \|A\|_\infty \cdot \|B\|_\infty$ holds for $n \times n$ matrices A and B (cf. Hannan and Deistler, 1988, p.266). Moreover, $|x'Ay| \le n \cdot \|x\|_\infty \cdot \|A\|_\infty \cdot \|y\|_\infty$ holds for $n \times 1$ vectors x, y, and an $n \times n$ matrix A.

Lemma 1" *Under Assumptions 1", 2, and L, the following relations hold:*

$$(a) \quad E\left[\iota'\hat{\phi}\right] = \iota'\phi - \frac{1}{T-p_T}\iota'(K_{p_T} + B_{p_T}\phi) + o\left(\frac{p_T}{T}\right),$$
$$(b) \quad E\left[\iota'(\hat{\phi}-\phi)(\hat{\phi}-\phi)'\iota\right] = \frac{1}{T-p_T}\sigma_\varepsilon^2\iota'R^{-1}\iota + o\left(\frac{p_T}{T}\right),$$

where ϕ is defined by (3.12)*, ι is a $p_T \times 1$ vector of ones and*

$$\hat{\phi} = \begin{bmatrix} \hat{\phi}_{p_T,1} \\ \hat{\phi}_{p_T,2} \\ \vdots \\ \hat{\phi}_{p_T,p_T} \end{bmatrix} = \begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} & \cdots & \hat{r}_{1,p_T} \\ \hat{r}_{21} & \hat{r}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{r}_{p_T-1,p_T} \\ \hat{r}_{p_T,1} & \cdots & \hat{r}_{p_T,p_T-1} & \hat{r}_{p_T,p_T} \end{bmatrix}^{-1} \begin{bmatrix} \hat{r}_{10} \\ \hat{r}_{20} \\ \vdots \\ \hat{r}_{p_T,0} \end{bmatrix}.$$

Proof of Lemma 1"

Proof of (a). Using (3.17) and the relation $\phi = R^{-1}r$, we have,

$$\begin{aligned}
\iota'\hat{\phi} &= \iota'\hat{R}^{-1}\hat{r} \\
&= \iota'\phi + \iota'R^{-1}(\hat{r}-r) - \iota'R^{-1}(\hat{R}-R)\phi - \iota'R^{-1}(\hat{R}-R)R^{-1}(\hat{r}-r) \\
&\quad + \iota'R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)\phi + \iota'R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)R^{-1}(\hat{r}-r) \\
&\quad - \iota'R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)\hat{R}^{-1}\hat{r} \\
&= \iota'\phi + (a) - (b) - (c) + (d) + (e) - (f), \quad \text{say.}
\end{aligned}$$

First, let us consider (a). Since Lemma 3 holds uniformly in $0 \le i \le p_T$ and $0 \le j \le p_T$, we have

$$E(\hat{r}_{ij}) = r_{ij} - \frac{m+1}{T-p_T}\omega + \xi_{ij}, \tag{3.24}$$

where $\xi_{ij} = o(T^{-1})$ uniformly in $0 \le i \le p_T$ and $0 \le j \le p_T$. Therefore,

$$\begin{aligned}
E[(a)] &= \iota'R^{-1}E(\hat{r}-r) \\
&= -\iota'R^{-1}\cdot\frac{m+1}{T-p_T}\omega\iota + \iota'R^{-1}\xi \\
&= (a_1) + (a_2), \quad \text{say,}
\end{aligned}$$

where $\xi = [\xi_{10}, \cdots, \xi_{p_T,0}]'$. Since $\|R^{-1}\|_\infty = O(1)$ (cf. den Haan and Levin, 1998), we obtain

$$\begin{aligned}
|(a_2)| &\le p_T \cdot \|\iota\|_\infty \cdot \|R^{-1}\|_\infty \cdot \|\xi\|_\infty \\
&= p_T \cdot O(1) \cdot O(1) \cdot o(T^{-1}) = o(p_T/T).
\end{aligned}$$

For (b), we have

$$\begin{aligned} E[(b)] &= \iota' R^{-1} E(\hat{R} - R)\phi \\ &= \iota' R^{-1} \cdot \frac{m+1}{T - p_T} \omega \iota' \phi + \iota' R^{-1} \Xi \phi \\ &= (b_1) + (b_2), \quad \text{say,} \end{aligned}$$

where Ξ is a $p_T \times p_T$ matrix whose (i,j) element is ξ_{ij}. Since $\Xi\phi = \left[\sum_{k=1}^{p_T} \xi_{1k}\phi_{p_T,k}, \cdots, \sum_{k=1}^{p_T} \xi_{p_T,k}\phi_{p_T,k}\right]'$, we have

$$\begin{aligned} \|\Xi\phi\|_\infty &= \max_{1 \le j \le p_T} \left| \sum_{k=1}^{p_T} \xi_{jk}\phi_{p_T,k} \right| \\ &\le \left(\sum_{k=1}^{p_T} |\phi_{p_T,k}| \right) \cdot \max_{1 \le j,k \le p_T} |\xi_{jk}| \\ &= O(1) \cdot o(T^{-1}) = o(T^{-1}). \end{aligned}$$

Therefore,

$$\begin{aligned} |(b_2)| &\le p_T \cdot \|\iota\|_\infty \cdot \|R^{-1}\|_\infty \cdot \|\Xi\phi\|_\infty \\ &= p_T \cdot O(1) \cdot O(1) \cdot o(T^{-1}) = o(p_T/T). \end{aligned}$$

Combining these results, we have

$$E[(a) - (b)] = (a_1) - (b_1) + o(p_T/T), \tag{3.25}$$

where $(a_1) - (b_1)$ corresponds to the first order bias of (A) given in (3.20) in the proof of Theorem 1.

We next consider (c). Since the result of Lemma 4 holds uniformly in $0 \le i \le p_T$, $0 \le j \le p_T$, $0 \le i' \le p_T$, and $0 \le j' \le p_T$, we have

$$E\left[(\hat{r}_{ij} - r_{ij})(\hat{r}_{i'j'} - r_{i'j'})\right] = \frac{1}{T - p_T} b_{ij,i'j'} + \xi_{ij,i'j'}, \tag{3.26}$$

where $b_{ij,i'j'}$ is the first-order bias of $(\hat{r}_{ij} - r_{ij})(\hat{r}_{i'j'} - r_{i'j'})$, and $\xi_{ij,i'j'} = O(T^{-3/2})$ uniformly in $0 \le i \le p_T$, $0 \le j \le p_T$, $0 \le i' \le p_T$, and $0 \le j' \le p_T$. Now, we have

$$(\hat{R} - R)R^{-1}(\hat{r} - r) = \begin{bmatrix} \sum_{\ell=1}^{p_T} \sum_{i'=1}^{p_T} r^{i'j'} (\hat{r}_{1i'} - r_{1i'})(\hat{r}_{j'0} - r_{j'0}) \\ \vdots \\ \sum_{\ell=1}^{p_T} \sum_{i'=1}^{p_T} r^{i'j'} (\hat{r}_{p_T,i'} - r_{p_T,i'})(\hat{r}_{j'0} - r_{j'0}) \end{bmatrix},$$

so that

$$\begin{aligned} E[(\hat{R} - R)R^{-1}(\hat{r} - r)] &= \begin{bmatrix} \sum_{j'=1}^{p_T} \sum_{i'=1}^{p_T} r^{i'j'} \cdot \frac{1}{T - p_T} b_{1i',j'0} + \sum_{j'=1}^{p_T} \sum_{i'=1}^{p_T} r^{kj'} \xi_{1i',j'0} \\ \vdots \\ \sum_{j'=1}^{p_T} \sum_{i'=1}^{p_T} r^{i'j'} \cdot \frac{1}{T - p_T} b_{p_T i',j'0} + \sum_{j'=1}^{p_T} \sum_{i'=1}^{p_T} r^{i'j'} \xi_{p_T i',j'0} \end{bmatrix} \\ &= B_1 + \tilde{\xi}, \quad \text{say,} \end{aligned}$$

where r^{ij} is the (i,j) element of R^{-1}. By (3.26), we obtain

$$\begin{aligned}
\|\tilde{\xi}\|_\infty &= \max_{1\le j\le p_T}\left|\sum_{j'=1}^{p_T}\sum_{i'=1}^{p_T} r^{i'j'}\xi_{ji',j'0}\right| \\
&\le \left(\sum_{j'=1}^{p_T}\sum_{i'=1}^{p_T}|r^{i'j'}|\right)\cdot \max_{1\le j,i',j'\le p_T}|\xi_{ji',j'0}| \\
&= O(p_T)\cdot O(T^{-3/2}) = O(p_T/T^{3/2}),
\end{aligned}$$

where $\sum_{j'=1}^{p_T}\sum_{i'=1}^{p_T}|r^{i'j'}| = O(p_T)$ holds because $\|R^{-1}\|_\infty = O(1)$. Therefore,

$$\begin{aligned}
E[(c)] &= \iota' R^{-1}E[(\hat{R}-R)R^{-1}(\hat{r}-r)] \\
&= \iota' R^{-1}B_1 + \iota' R^{-1}\tilde{\xi} \\
&= (c_1)+(c_2), \quad \text{say}
\end{aligned}$$

Note that

$$\begin{aligned}
|(c_2)| &\le p_T\cdot\|\iota\|_\infty\cdot\|R^{-1}\|_\infty\cdot\|\tilde{\xi}\|_\infty \\
&= p_T\cdot O(1)\cdot O(1)\cdot O(p_T/T^{3/2}) = O(p_T^2/T^{3/2}).
\end{aligned}$$

For (d), because the (i,j) element of $(\hat{R}-R)R^{-1}(\hat{R}-R)$ is given by $\sum_{j'=1}^{p_T}\sum_{i'=1}^{p_T} r^{i'j'}(\hat{r}_{ii'}-r_{ii'})(\hat{r}_{j'j}-r_{j'j})$, we have

$$\begin{aligned}
&E[(\hat{R}-R)R^{-1}(\hat{R}-R)] \\
= &\begin{bmatrix}
\sum_{j'=1}^{p_T}\sum_{i'=1}^{p_T} r^{i'j'}\left(\frac{1}{T-p_T}b_{1i',j'1}+\xi_{1i',j'1}\right) & \cdots & \sum_{j'=1}^{p_T}\sum_{i'=1}^{p_T} r^{i'j'}\left(\frac{1}{T-p_T}b_{1i',j'p_T}+\xi_{1i',j'p_T}\right) \\
\vdots & \ddots & \vdots \\
\sum_{j'=1}^{p_T}\sum_{i'=1}^{p_T} r^{i'j'}\left(\frac{1}{T-p_T}b_{p_Ti',j'1}+\xi_{p_Ti',j'1}\right) & \cdots & \sum_{j'=1}^{p_T}\sum_{i'=1}^{p_T} r^{i'j'}\left(\frac{1}{T-p_T}b_{p_Ti',j'p_T}+\xi_{p_Ti',j'p_T}\right)
\end{bmatrix} \\
= &\ B_2+\tilde{\Xi}, \quad \text{say},
\end{aligned}$$

and each element of $\tilde{\Xi}$ is uniformly $O(p_T/T^{3/2})$. Therefore, we have

$$\begin{aligned}
E[(d)] &= \iota' R^{-1}E[(\hat{R}-R)R^{-1}(\hat{R}-R)]\phi \\
&= \iota' R^{-1}B_2\phi + \iota' R^{-1}\tilde{\Xi}\phi \\
&= (d_1)+(d_2), \quad \text{say}.
\end{aligned}$$

Note that

$$\begin{aligned}
|(d_2)| &\le p_T\cdot\|\iota\|_\infty\cdot\|R^{-1}\|_\infty\cdot\|\tilde{\Xi}\|_\infty\cdot\|\phi\|_\infty \\
&= p_T\cdot O(1)\cdot O(1)\cdot O(p_T^2/T^{3/2})\cdot O(1) = O(p_T^3/T^{3/2}).
\end{aligned}$$

Combining the above results, we have

$$E[-(c)+(d)] = -(c_1)+(d_1)+o(p_T/T), \tag{3.27}$$

where $-(c_1)+(d_2)$ corresponds to the first order bias of (B) given in (3.22) in the proof of Theorem 1.

For (e), because $\|\hat{R}-R\|_\infty = O_p(p_T/\sqrt{T})$ and $\|\hat{r}-r\|_\infty = O_p(T^{-1/2})$ (cf. den Haan and Levin, 1998), we have

$$\begin{aligned}
|(e)| &\leq p_T \cdot \|\iota\|_\infty \left\{\|R^{-1}\|_\infty \cdot \|\hat{R}-R\|_\infty\right\}^2 \|R^{-1}\|_\infty \cdot \|\hat{r}-r\|_\infty \\
&= p_T \cdot O(1) \cdot \left\{O(1)\cdot O_p(p_T/\sqrt{T})\right\}^2 \cdot O(1) \cdot O_p(T^{-1/2}) \\
&= O_p(p_T^3/T^{3/2}).
\end{aligned}$$

Finally, let us consider (f), which can be expressed as $(f) = \iota' R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)\left\{\phi+(\hat{\phi}-\phi)\right\}$.
Since $(\hat{R}-R)\phi = \left[\sum_{j'=1}^{p_T}(\hat{r}_{1j'}-r_{1j'})\phi_{p_T,j'}, \cdots, \sum_{j'=1}^{p_T}(\hat{r}_{p_T,j'}-r_{p_T,j'})\phi_{p_T,j'}\right]'$, we have

$$\begin{aligned}
\|(\hat{R}-R)\phi\|_\infty &= \max_{1\leq j\leq p_T}\left|\sum_{j'=1}^{p_T}(\hat{r}_{jj'}-r_{jj'})\phi_{p_T,j'}\right| \\
&\leq \max_{1\leq j,j'\leq p_T}|\hat{r}_{jj'}-r_{jj'}| \cdot \left(\sum_{j'=1}^{p_T}|\phi_{p_T,j'}|\right) \\
&= O_p(T^{-1/2})\cdot O(1) \\
&= O_p(T^{-1/2}).
\end{aligned}$$

Hence,

$$\begin{aligned}
|(f)| &\leq p_T \cdot \|\iota\|_\infty \left\{\|R^{-1}\|_\infty \cdot \|\hat{R}-R\|_\infty\right\}^2 \|R^{-1}\|_\infty \left\{\|(\hat{R}-R)\phi\|_\infty + \|\hat{R}-R\|_\infty \cdot \|\hat{\phi}-\phi\|_\infty\right\} \\
&= p_T \cdot O(1) \cdot \left\{O(1)\cdot O_p(p_T/\sqrt{T})\right\}^2 \cdot O(1) \cdot \left\{O_p(T^{-1/2}) + O_p(p_T/\sqrt{T})\cdot O_p(T^{-1/2})\right\} \\
&= O_p(p_T^3/T^{3/2}).
\end{aligned}$$

Therefore, we have $E(\iota'\hat{\phi}) = \iota'\phi + (a_1) - (b_1) - (c_1) + (d_1) + o(p_T/T)$ because $p_T^4/T \to 0$. Since the first-order bias of $\iota'\hat{\phi}$ given by $(a_1)-(b_1)-(c_1)+(d_1)$ is exactly equal to the one derived in Appendix A, we obtain the desired result. ■

Proof of (b). By defining $\eta_{p_T,t} = \sum_{j=1}^{p_T}(\phi_j - \phi_{p_T,j})u_{t-j} + \sum_{j=p_T+1}^{\infty}\phi_j u_{t-j}$, we can see that

$$u_t = \sum_{j=1}^{p_T}\phi_{p_T,j}u_{t-j} + \eta_{p_T,t} + \varepsilon_t. \tag{3.28}$$

Therefore, we have, for $\ell = 1, \cdots, m+1$,

$$\begin{aligned}
\bar{u}_\ell &= \frac{1}{T_\ell - T_{\ell-1}} \sum_{t=T_{\ell-1}+1}^{T_\ell} u_t \\
&= \sum_{j=1}^{p_T} \phi_{p_T,j} \left(\frac{1}{T_\ell - T_{\ell-1}} \sum_{t=T_{\ell-1}+1}^{T_\ell} u_{t-j} \right) + \frac{1}{T_\ell - T_{\ell-1}} \sum_{t=T_{\ell-1}+1}^{T_\ell} \eta_{p_T,t} + \frac{1}{T_\ell - T_{\ell-1}} \sum_{t=T_{\ell-1}+1}^{T_\ell} \varepsilon_t \\
&= \sum_{j=1}^{p_T} \phi_{p_T,j} \bar{u}_\ell + \bar{\eta}_\ell + \bar{\varepsilon}_\ell + \sum_{j=1}^{p_T} \phi_{p_T,j} \left\{ \frac{1}{T_\ell - T_{\ell-1}} \sum_{k=1}^{j} (u_{T_{\ell-1}+1-k} - u_{T_\ell+1-k}) \right\}, \qquad (3.29)
\end{aligned}$$

where $\bar{\eta}_\ell = (T_\ell - T_{\ell-1})^{-1} \sum_{t=T_{\ell-1}+1}^{T_\ell} \eta_{p_T,t}$ and $\bar{\varepsilon}_\ell = (T_\ell - T_{\ell-1})^{-1} \sum_{t=T_{\ell-1}+1}^{T_\ell} \varepsilon_t$. From (3.28) and (3.29), we have, for $t = T_{\ell-1} + p_T + 1, \cdots, T_\ell$,

$$\begin{aligned}
\hat{u}_t &= u_t - \bar{u}_\ell \\
&= \sum_{j=1}^{p_T} \phi_{p_T,j} \hat{u}_{t-j} + (\eta_{p_T,t} - \bar{\eta}_\ell) + (\varepsilon_t - \bar{\varepsilon}_\ell) + h_\ell, \qquad (3.30)
\end{aligned}$$

where $h_\ell = -\sum_{j=1}^{p_T} \phi_{p_T,j} \left\{ (T_\ell - T_{\ell-1})^{-1} \sum_{k=1}^{j} (u_{T_{\ell-1}+1-k} - u_{T_\ell+1-k}) \right\}$.

Similarly, for $t = T_\ell + 1, \cdots, T_\ell + p_T$ $(\ell = 1, \cdots, m)$, we have

$$\begin{aligned}
\hat{u}_t &= u_t - \bar{u}_\ell \cdot 1\{t \le T_\ell\} - \bar{u}_{\ell+1} \cdot 1\{t > T_\ell\} \\
&= \sum_{j=1}^{p_T} \phi_{p_T,j} u_{t-j} + \eta_{p_T,t} + \varepsilon_t - \bar{u}_\ell \cdot 1\{t \le T_\ell\} - \bar{u}_{\ell+1} \cdot 1\{t > T_\ell\} \\
&= \sum_{j=1}^{p_T} \phi_{p_T,j} \left[\hat{u}_{t-j} + \bar{u}_\ell \cdot 1\{t-j \le T_\ell\} + \bar{u}_{\ell+1} \cdot 1\{t-j > T_\ell\} \right] + \eta_{p_T,t} + \varepsilon_t \\
&\qquad - \bar{u}_\ell \cdot 1\{t \le T_\ell\} - \bar{u}_{\ell+1} \cdot 1\{t > T_\ell\} \\
&= \sum_{j=1}^{p_T} \phi_{p_T,j} \hat{u}_{t-j} + \eta_{p_T,t} + \varepsilon_t + \tilde{h}_{t,\ell}, \qquad (3.31)
\end{aligned}$$

where $\tilde{h}_{t,\ell} = \sum_{j=1}^{p_T} \phi_{p_T,j} \left[\bar{u}_\ell \cdot 1\{t-j \le T_\ell\} + \bar{u}_{\ell+1} \cdot 1\{t-j > T_\ell\} \right] - \bar{u}_\ell \cdot 1\{t \le T_\ell\} - \bar{u}_{\ell+1} \cdot 1\{t > T_\ell\}$.

From (3.30) and (3.31), we obtain

$$\hat{u}_t = \begin{cases} \underline{\hat{u}}_{t-1}' \phi + (\eta_{p_T,t} - \bar{\eta}_\ell) + (\varepsilon_t - \bar{\varepsilon}_\ell) + h_\ell & \text{for } t = T_{\ell-1} + p_T + 1, \cdots, T_\ell \ (\ell = 1, \cdots, m+1), \\ \underline{\hat{u}}_{t-1}' \phi + \eta_{p_T,t} + \varepsilon_t + \tilde{h}_{t,\ell} & \text{for } t = T_\ell + 1, \cdots, T_\ell + p_T \ (\ell = 1, \cdots, m), \end{cases} \qquad (3.32)$$

where $\underline{\hat{u}}_t = [\hat{u}_t, \cdots, \hat{u}_{t-p_T+1}]'$.

Since $\hat{\phi} = \left(\sum_{t=p_T+1}^{T} \underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}\right)^{-1}\left(\sum_{t=p_T+1}^{T} \underline{\hat{u}}_{t-1}\hat{u}_t\right)$, we obtain, using (3.32),

$$\begin{aligned}
&\sqrt{T-p_T}\,(\hat{\phi}-\phi)\\
=\;& \left(\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}\right)^{-1}\\
&\times\left[\frac{1}{\sqrt{T-p_T}}\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_\ell)\right)+\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}\underline{\hat{u}}_{t-1}\eta_{p_T,t}\right)\right\}\right.\\
&\quad+\frac{1}{\sqrt{T-p_T}}\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_\ell)\right)+\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}\underline{\hat{u}}_{t-1}\varepsilon_t\right)\right\}\\
&\quad\left.+\frac{1}{\sqrt{T-p_T}}\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}h_\ell\right)+\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}\underline{\hat{u}}_{t-1}\tilde{h}_{t,\ell}\right)\right\}\right]\\
=\;& \hat{R}^{-1}[(A)+(B)+(C)],\quad \text{say.}
\end{aligned}$$

First, let us consider (A), which can be expressed as

$$\begin{aligned}
(A) \;=\;& \frac{1}{\sqrt{T-p_T}}\left[\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\eta_{p_T,t}-\sum_{\ell=1}^{m+1}\left(\sum_{t=T_\ell+p_T+1}^{T_\ell}\underline{u}_{t-1}\bar{\eta}_\ell\right)-\sum_{\ell=1}^{m+1}\left(\bar{u}_{\ell\iota}\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\eta_{p_T,t}\right)\right.\\
&\left.+\sum_{\ell=1}^{m+1}((T_\ell-p_T)\bar{u}_{\ell\iota}\bar{\eta}_\ell)+\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}(\underline{\hat{u}}_{t-1}-\underline{u}_{t-1})\eta_{p_T,t}\right)\right]\\
=\;& (A_1)-(A_2)-(A_3)+(A_4)+(A_5),\quad \text{say,}
\end{aligned}$$

where $\underline{u}_t = [u_t, \cdots, u_{t-p_T+1}]'$.

By den Haan and Levin (1998), Assumption L(b) implies

$$\sum_{j=1}^{p_T}|\phi_j-\phi_{p_T,j}| = o\left(\frac{p_T}{T}\right). \tag{3.33}$$

Therefore, from Assumption L(b) and (3.33), we obtain

$$\begin{aligned}
&E\left|\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}u_{t-\ell}\eta_{p_T,t}\right|\\
=\;& E\left|\sum_{j=1}^{p_T}(\phi_j-\phi_{p_T,j})\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}u_{t-\ell}u_{t-j}\right)+\sum_{j=p_T+1}^{\infty}\phi_j\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}u_{t-\ell}u_{t-j}\right)\right|\\
\le\;& \left(\sum_{j=1}^{p_T}|\phi_j-\phi_{p_T,j}|+\sum_{j=p_T+1}^{\infty}|\phi_j|\right)\cdot\sup_{j\ge1}E\left|\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}u_{t-\ell}u_{t-j}\right|\\
=\;& o(p_T/T)\cdot O(\sqrt{T})\\
=\;& o(p_T/\sqrt{T})
\end{aligned}$$

uniformly in $1 \leq \ell \leq p_T$, so that $\|(A_1)\|_\infty = o_p(p_T/\sqrt{T})$.

Similarly, since $E\left|(T-p_T)^{-1}\sum_{t=p_T+1}^{T}\eta_{p_T,t}\right| = o(p_T/T)$, we have $\|(A_2)\|_\infty = o_p(p_T/T)$. In the same way, we obtain $\|(A_3)\|_\infty = o_p(p_T/T)$, $\|(A_4)\|_\infty = o_p(p_T/T)$, and $\|(A_5)\|_\infty = O_p(p_T/\sqrt{T})$. Therefore, $\|(A)\|_\infty = O_p(p_T/\sqrt{T})$.

For (B), since $(T-p_T)^{-1/2}\sum_{t=p_T+1}^{T} u_{t-\ell} = O_p(1)$ uniformly in $1 \leq \ell \leq p_T$, $\bar{\varepsilon}_\ell = O_p(T^{-1/2})$ for $\ell = 1, \cdots, m+1$, we have

$$\left\| (B) - \frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}_{t-1}\varepsilon_t \right\|_\infty = O_p(T^{-1/2}).$$

Now let us consider (C). Since

$$\begin{aligned} |h_\ell| &\leq \left(\sum_{j=1}^{p_T} |\phi_{p_T,j}|\right) \cdot \frac{1}{T_\ell - T_{\ell-1}} \sum_{k=1}^{p_T} \left(|u_{T_{\ell-1}+1-k}| + |u_{T_\ell+1-k}|\right) \\ &= O(1) \cdot O_p(p_T/T) \\ &= O_p(p_T/T) \end{aligned}$$

for $1 \leq \ell \leq m+1$, we have

$$\begin{aligned} \|(C)\|_\infty &\leq \sum_{\ell=1}^{m+1} \left(\left\| \frac{1}{\sqrt{T-p_T}} \sum_{t=T_{\ell-1}+p_T+1}^{T_\ell} \underline{\hat{u}}_{t-1} \right\|_\infty \cdot |h_\ell| \right) + \sum_{\ell=1}^{m} \left\| \frac{1}{\sqrt{T-p_T}} \sum_{t=T_\ell+1}^{T_\ell+p_T} \underline{\hat{u}}_{t-1}\tilde{h}_t \right\|_\infty \\ &= O_p(1) \cdot O_p(p_T/T) + O_p(p_T/\sqrt{T}) \\ &= O_p(p_T/\sqrt{T}). \end{aligned}$$

Therefore, we obtain

$$\sqrt{T-p_T}\,(\hat{\phi}-\phi) = \hat{R}^{-1}\left(\frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}_{t-1}\varepsilon_t + \zeta_T \right), \tag{3.34}$$

where $\|\zeta_T\|_\infty = O_p(p_T/\sqrt{T})$.

Then we evaluate the expectation of $\iota'(\hat{\phi}-\phi)(\hat{\phi}-\phi)'\iota$ up to $O(p_T/T)$. Using (3.34), this

can be expressed as

$$\begin{aligned}
& \iota'(\hat{\phi}-\phi)(\hat{\phi}-\phi)'\iota \\
= & \frac{1}{T-p_T}\iota'\left\{R^{-1}+(\hat{R}^{-1}-R^{-1})\right\}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t+\zeta_T\right) \\
& \times\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t+\zeta'_T\right)\left\{R^{-1}+(\hat{R}^{-1}-R^{-1})\right\}\iota \\
= & \frac{1}{T-p_T}\iota'R^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t\right)R^{-1}\iota \\
& +o_p(p_T/T), \qquad (3.35)
\end{aligned}$$

because $\|R^{-1}\|_\infty = O(1)$, $\|\hat{R}^{-1}-R^{-1}\|_\infty = O_p(p_T/\sqrt{T})$, $\|(T-p_T)^{-1/2}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\|_\infty = O_p(1)$, and $\|\zeta_T\|_\infty = O_p(p_T/\sqrt{T})$.

Since $E(\varepsilon_t|\mathcal{F}_{t-1})=0$ and $E(\varepsilon_t^2|\mathcal{F}_{t-1})=\sigma_\varepsilon^2$, we have

$$\begin{aligned}
E\left[\frac{1}{T-p_T}\left(\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\left(\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t\right)\right] &= E\left[\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\underline{u}'_{t-1}\varepsilon_t^2\right] \\
&= \sigma_\varepsilon^2 R. \qquad (3.36)
\end{aligned}$$

Therefore, from (3.35) and (3.36), we obtain

$$E\left[\iota'(\hat{\phi}-\phi)(\hat{\phi}-\phi)'\iota\right]=\frac{1}{T-p_T}\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T). \blacksquare$$

Lemma 2" *Under Assumptions 1", 2, and L, the following relations hold:*

(a) $E\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right]=(1-\iota'\phi)^2$
$$+\frac{1}{T-p_T}\left\{2(1-\iota'\phi)\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)+\sigma_\varepsilon^2\iota'R^{-1}\iota\right\}+o\left(\frac{p_T}{T}\right),$$

(b) $E\left[\hat{\sigma}_\varepsilon^2\right]=\sigma_\varepsilon^2-\frac{p_T+m+1}{T-p_T}\sigma_\varepsilon^2+o\left(\frac{p_T}{T}\right),$

(c) $Var\left[\hat{\sigma}_\varepsilon^2\right]=\frac{1}{T-p_T}\left\{E(\varepsilon_t^4)-\sigma_\varepsilon^4\right\}+o(T^{-1}),$

(d) $Cov\left[\hat{\sigma}_\varepsilon^2,\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right]=o\left(\frac{p_T}{T}\right).$

Proof of Lemma 2"

Proof of (a). Here we define $\psi=-\hat{\phi}+\phi-(T-p_T)^{-1}\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)$. Then, from Lemma 1", we obtain

$$\begin{aligned}
E(\iota'\psi) &= o(p_T/T), \qquad (3.37)\\
Var(\iota'\psi) &= \frac{1}{T-p_T}\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T). \qquad (3.38)
\end{aligned}$$

Since $1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}=1-\iota'\hat{\phi}=1-\iota'\phi+(T-p_T)^{-1}\iota'(K_{p_T}+B_{p_T}\phi)+\iota'\psi$, we have

$$\begin{aligned}
E\left[\left(1-\textstyle\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] &= E\left[\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)+\iota'\psi\right\}^2\right] \\
&= \left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)\right\}^2 \\
&\quad +2\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)\right\}E\left[\iota'\psi\right]+E\left[\left(\iota'\psi\right)^2\right] \\
&= (a)+2\cdot(b)+(c), \quad \text{say.} \qquad (3.39)
\end{aligned}$$

By (3.37) and (3.38), we have

$$\begin{aligned}
(a) &= (1-\iota'\phi)^2+\frac{2}{T-p_T}(1-\iota'\phi)\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)+o(p_T/T), \\
(b) &= o(p_T/T), \\
(c) &= \frac{1}{T-p_T}\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T).
\end{aligned}$$

Therefore, the expectation up to $O(p_T/T)$ is given by

$$\begin{aligned}
E\left[\left(1-\textstyle\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] &= (1-\iota'\phi)^2+\frac{2}{T-p_T}(1-\iota'\phi)\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right) \\
&\quad +\frac{1}{T-p_T}\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T). \blacksquare
\end{aligned}$$

Proof of (b). For $t=T_{\ell-1}+p_T+1,\cdots,T_\ell$ $(\ell=1,\cdots,m+1)$, $\hat{\varepsilon}_t$ can be expressed as

$$\begin{aligned}
\hat{\varepsilon}_t &= (u_t-\bar{u}_\ell)-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}(u_{t-j}-\bar{u}_\ell) \\
&= (u_t-\bar{u}_\ell)-\sum_{j=1}^{p_T}\phi_{p_T,j}(u_{t-j}-\bar{u}_\ell)-\sum_{j=1}^{p_T}(\hat{\phi}_{p_T,j}-\phi_{p_T,j})(u_{t-j}-\bar{u}_\ell) \\
&= (\varepsilon_t-\bar{\varepsilon}_\ell)-(\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}+(\eta_{p_T,t}-\bar{\eta}_\ell)+h_\ell, \qquad (3.40)
\end{aligned}$$

where the last equality holds because $\varepsilon_t = u_t-\sum_{j=1}^{p_T}\phi_{p_T,j}u_{t-j}-\eta_{p_T,t}$ and $\bar{\varepsilon}_\ell = \bar{u}_\ell-\sum_{j=1}^{p_T}\phi_{p_T,j}\bar{u}_\ell-\bar{\eta}_\ell+h_\ell$.

For $t=T_\ell+1,\cdots,T_\ell+p_T$ $(\ell=1,\cdots,m)$, we have

$$\begin{aligned}
\hat{\varepsilon}_t &= \hat{u}_t-\sum_{j=1}^{p_T}\phi_{p_T,j}\hat{u}_{t-j}-\sum_{j=1}^{p_T}(\hat{\phi}_{p_T,j}-\phi_{p_T,j})\hat{u}_{t-j} \\
&= \varepsilon_t+\eta_{p_T,t}+O_p(p_T/\sqrt{T}). \qquad (3.41)
\end{aligned}$$

Using (3.40) and (3.41) and noting that $|h_\ell| = O_p(p_T/T)$ for $1 \le \ell \le m+1$, we have

$$\begin{aligned}
\hat{\sigma}_\varepsilon^2 &= \frac{1}{T-p_T}\left[\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}(\varepsilon_t-\bar{\varepsilon}_\ell)^2\right)+\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}\varepsilon_t^2\right)\right\}\right. \\
&\quad -2\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}(\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_\ell)\right)\right\} \\
&\quad +\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\left((\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}\right)^2\right)\right\} \\
&\quad +\left\{2\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}(\eta_{p_T,t}-\bar{\eta}_\ell)\left((\varepsilon_t-\bar{\varepsilon}_\ell)-(\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}\right)\right)+2\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}\eta_{p_T,t}\varepsilon_t\right)\right. \\
&\quad \left.\left.+\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}(\eta_{p_T,t}-\bar{\eta}_\ell)^2\right)+\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}\eta_{p_T,t}^2\right)\right\}\right]+o_p\left(\frac{p_T}{T}\right) \\
&= (A)-2\cdot(B)+(C)+(D)+o_p\left(\frac{p_T}{T}\right), \quad \text{say.}
\end{aligned}$$

First, consider the term (A). Since

$$(A)=\sum_{\ell=1}^{m+1}\left\{\frac{T_\ell-T_{\ell-1}-p_T}{T-p_T}\left(\frac{1}{T_\ell-T_{\ell-1}-p_T}\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}(\varepsilon_t-\bar{\varepsilon}_\ell)^2\right)\right\},$$

we have

$$\begin{aligned}
E[(A)] &= \sigma_\varepsilon^2-\sum_{\ell=1}^{m+1}\left\{(\lambda_\ell-\lambda_{\ell-1})\cdot\frac{1}{(\lambda_\ell-\lambda_{\ell-1})(T-p_T)}\sigma_\varepsilon^2\right\}+o(T^{-1}) \qquad (3.42)\\
&= \sigma_\varepsilon^2-\frac{m+1}{T-p_T}\sigma_\varepsilon^2+o(T^{-1}).
\end{aligned}$$

Next, let us consider (B). Since

$$\frac{1}{\sqrt{T-p_T}}\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_\ell)\right)\right\}=\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t+\tilde{\zeta}_T,$$

where $\|\tilde{\zeta}_T\|_\infty = O_p(\sqrt{p_T/T})$, we have

$$\begin{aligned}
(B) &= \frac{1}{T-p_T}\cdot\sqrt{T-p_T}(\hat{\phi}-\phi)'\left[\frac{1}{\sqrt{T-p_T}}\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_\ell)\right)\right\}\right] \\
&= \frac{1}{T-p_T}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t+\zeta'_T\right)\left\{R^{-1}+(\hat{R}^{-1}-R^{-1})\right\} \\
&\quad\times\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t+\tilde{\zeta}_T\right) \\
&= \frac{1}{T-p_T}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t\right)R^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)+o_p\left(\frac{p_T}{T}\right),
\end{aligned}$$

because $\|\zeta_T\|_\infty = O_p(p_T/\sqrt{T})$.

From (3.36), we obtain

$$\begin{aligned}
&E\left[\frac{1}{T-p_T}\left(\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t\right)R^{-1}\left(\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\right] \\
&= tr\left[R^{-1}E\left(\frac{1}{T-p_T}\left(\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\left(\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t\right)\right)\right] \\
&= tr\left[R^{-1}\cdot\sigma_\varepsilon^2 R\right] \\
&= p_T\sigma_\varepsilon^2, \qquad (3.43)
\end{aligned}$$

so that $E[(B)] = (T-p_T)^{-1}\cdot p_T\sigma_\varepsilon^2 + o(p_T/T)$.

Next, let us consider (C):

$$\begin{aligned}
(C) &= (\hat{\phi}-\phi)'\left[\frac{1}{T-p_T}\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}\right)\right\}\right](\hat{\phi}-\phi) \\
&= (\hat{\phi}-\phi)'\hat{\hat{R}}(\hat{\phi}-\phi),
\end{aligned}$$

where $\hat{\hat{R}} = (T-p_T)^{-1}\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}\right)\right\}$.

Since $\|\hat{\hat{R}} - R\|_\infty = O_p(p_T/\sqrt{T})$, we have

$$\begin{aligned}
(C) &= \frac{1}{T-p_T}\left\{\sqrt{T-p_T}\,(\hat{\phi}-\phi)'\right\}\hat{\hat{R}}\left\{\sqrt{T-p_T}\,(\hat{\phi}-\phi)\right\} \\
&= \frac{1}{T-p_T}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}'\varepsilon_t+\zeta_T'\right)\left\{R^{-1}+(\hat{R}^{-1}-R^{-1})\right\}\left\{R+(\hat{\hat{R}}-R)\right\} \\
&\qquad\times\left\{R^{-1}+(\hat{R}^{-1}-R^{-1})\right\}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t+\zeta_T\right) \\
&= \frac{1}{T-p_T}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}'\varepsilon_t\right)R^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)+o_p\left(\frac{p_T}{T}\right),
\end{aligned}$$

so that we obtain $E[(C)] = (T-p_T)^{-1}\cdot p_T\sigma_\varepsilon^2 + o(p_T/T)$, using (3.43).

Finally, let us consider (D), which can be expressed as

$$\begin{aligned}
(D) &= \frac{1}{T-p_T}\left[2\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}(\eta_{p_T,t}-\bar{\eta}_\ell)(\varepsilon_t-\bar{\varepsilon}_\ell)\right)+\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}\eta_{p_T,t}\varepsilon_t\right)\right\}\right. \\
&\qquad -2(\hat{\phi}-\phi)'\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_\ell)\right)\right\} \\
&\qquad \left.+\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}(\eta_{p_T,t}-\bar{\eta}_\ell)^2\right)+\sum_{\ell=1}^{m}\left(\sum_{t=T_\ell+1}^{T_\ell+p_T}\eta_{p_T,t}^2\right)\right\}\right] \\
&= 2\cdot(D_1)-2\cdot(D_2)+(D_3), \quad \text{say.}
\end{aligned}$$

First, consider the term (D_1). By Assumption L(b) and (3.33), we have

$$\begin{aligned}
E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\eta_{p_T,t}\varepsilon_t\right| &\le \left(\sum_{j=1}^{p_T}|\phi_j-\phi_{p_T,j}|+\sum_{j=p_T+1}^{\infty}|\phi_j|\right)\cdot\sup_{j\ge1}E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}u_{t-j}\varepsilon_t\right| \\
&= o(p_T/T)\cdot O(1) \\
&= o(p_T/T),
\end{aligned}$$

and thus $(D_1) = o_p(p_T/T)$.

Then, let us consider (D_2). Here we define

$$P = \frac{1}{T-p_T}\left\{\sum_{\ell=1}^{m+1}\left(\sum_{t=T_{\ell-1}+p_T+1}^{T_\ell}\underline{\hat{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_\ell)\right)\right\}.$$

Then, $\|P\|_\infty = o_p(p_T/T)$ because $(T-p_T)^{-1}\sum_{t=p_T+1}^{T}u_{t-\ell}\eta_{p_T,t} = o_p(p_T/T)$ uniformly in $1\le\ell\le p_T$.

Since $(D_2) = (T-p_T)^{-1/2} \cdot \left\{\sqrt{T-p_T}(\hat{\phi}-\phi)\right\} P$, we have

$$\begin{aligned}
\|(D_2)\|_\infty &\leq \frac{1}{\sqrt{T-p_T}} \cdot p_T \cdot \left(\left\| \frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}_{t-1}\varepsilon_t \right\|_\infty + \|\zeta_T\|_\infty \right) \cdot \|\hat{R}^{-1}\|_\infty \cdot \|P\|_\infty \\
&= O(T^{-1/2}) \cdot p_T \cdot \left\{ O_p(1) + O_p(p_T/\sqrt{T}) \right\} \cdot O_p(1) \cdot o_p(p_T/T) \\
&= o_p(p_T^2/T^{3/2}).
\end{aligned}$$

Then, let us consider (D_3). First, we have

$$\begin{aligned}
E\left| \frac{1}{T-p_T} \sum_{t=p_T+1}^{T} \eta_{p_T,t}^2 \right| &= E\left| \frac{1}{T-p_T} \sum_{t=p_T+1}^{T} \left\{ \sum_{j=1}^{p_T} (\phi_j - \phi_{p_T,j}) u_{t-j} + \sum_{j=p_T+1}^{\infty} \phi_j u_{t-j} \right\}^2 \right| \\
&\leq E\left| \frac{1}{T-p_T} \sum_{t=p_T+1}^{T} \left\{ \sum_{j=1}^{p_T} (\phi_j - \phi_{p_T,j}) u_{t-j} \right\}^2 \right| \\
&\quad + 2E\left| \frac{1}{T-p_T} \sum_{t=p_T+1}^{T} \left\{ \sum_{j=1}^{p_T} (\phi_j - \phi_{p_T,j}) u_{t-j} \right\} \left\{ \sum_{j=p_T+1}^{\infty} \phi_j u_{t-j} \right\} \right| \\
&\quad + E\left| \frac{1}{T-p_T} \sum_{t=p_T+1}^{T} \left\{ \sum_{j=p_T+1}^{\infty} \phi_j u_{t-j} \right\}^2 \right| \\
&= (D_3 - 1) + 2 \cdot (D_3 - 2) + (D_3 - 3), \quad \text{say.}
\end{aligned}$$

Since

$$\begin{aligned}
(D_3 - 1) &\leq \left(\sum_{j=1}^{p_T} |\phi_j - \phi_{p_T,j}| \right)^2 \cdot \sup_{s,t} E|u_s u_t| \\
&= o(p_T^2/T^2) \cdot O(1) \\
&= o(p_T^2/T^2),
\end{aligned}$$

and similarly $(D_3 - 2) = o(p_T^2/T^2)$ and $(D_3 - 3) = o(p_T^2/T^2)$, we have $E\left|(T-p_T)^{-1}\sum_{t=p_T+1}^{T} \eta_{p_T,t}^2\right| = o(p_T^2/T^2)$, so that $(D_3) = o_p(p_T^2/T^2)$. Thus we have $(D) = o_p(p_T/T)$.

Using the above results, we obtain

$$E(\hat{\sigma}_\varepsilon^2) = \sigma_\varepsilon^2 - \frac{p_T + m + 1}{T - p_T}\sigma_\varepsilon^2 + o(p_T/T). \blacksquare$$

Proof of (c).

Since $Var\left[\left(1 - \sum_{j=1}^{p_T} \hat{\phi}_{p_T,j}\right)^2\right] = E\left[\left(1 - \sum_{j=1}^{p_T} \hat{\phi}_{p_T,j}\right)^4\right] - \left\{E\left[\left(1 - \sum_{j=1}^{p_T} \hat{\phi}_{p_T,j}\right)^2\right]\right\}^2$,
We only need to obtain $E\left[\left(1 - \sum_{j=1}^{p_T} \hat{\phi}_{p_T,j}\right)^4\right]$ to prove (c).

Here we have

$$\begin{aligned}\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^4 &= \left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi+\iota'\psi\right)\right\}^4 \\ &= \left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)\right\}^4 \\ &\quad +4\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)\right\}^3(\iota'\psi) \\ &\quad +6\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)\right\}^2(\iota'\psi)^2 \\ &\quad +4\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)\right\}(\iota'\psi)^3 \\ &\quad +(\iota'\psi)^4 \\ &= (A)+(B)+(C)+(D)+(E), \quad \text{say.}\end{aligned}$$

First,

$$(A)=(1-\iota'\phi)^4+\frac{4}{T-p_T}(1-\iota'\phi)^3\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)+o(p_T/T).$$

Then, since $E(\iota'\psi)=o(p_T/T)$, we have $E[(B)]=o(p_T/T)$.

Next, we evaluate the expectation of (C). Because $E[(\iota'\psi)^2]=Var(\iota'\psi)+\{E(\iota'\psi)\}^2=(T-p_T)^{-1}\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T)$, we have

$$\begin{aligned}E[(C)] &= 6\left\{(1-\iota'\phi)^2+O(p_T/T)\right\}\cdot\left\{\frac{1}{T-p_T}\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T)\right\} \\ &= \frac{6}{T-p_T}(1-\iota'\phi)^2\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T).\end{aligned}$$

Finally, let us consider (D) and (E). Here we have $\iota'\psi=O_p(\sqrt{p_T/T})$ from (3.38), so that $(\iota'\psi)^3=O_p(p_T^{3/2}/T^{3/2})$ and $(\iota'\psi)^4=O_p(p_T^2/T^2)$. Since $p_T^4/T\to 0$ as $T\to\infty$, we obtain $(\iota'\psi)^3=o_p(p_T/T)$ and $(\iota'\psi)^4=o_p(p_T/T)$, and thus $E[(D)]$ and $E[(E)]$ are both $o(p_T/T)$.

Therefore,

$$\begin{aligned}E\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^4\right] &= (1-\iota'\phi)^4+\frac{4}{T-p_T}(1-\iota'\phi)^3\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right) \\ &\quad +\frac{6}{T-p_T}(1-\iota'\phi)^2\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T).\end{aligned}$$

Using the above result and Lemma 2" (a), we obtain

$$\begin{aligned}Var\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] &= E\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^4\right]-\left\{E\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right]\right\}^2 \\ &= \frac{4}{T-p_T}(1-\iota'\phi)^2\sigma_\varepsilon^2\iota'R^{-1}\iota+o(p_T/T).\blacksquare\end{aligned}$$

Proof of (d). We only need to obtain $E\left[(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j})^2\hat{\sigma}_\varepsilon^2\right]$ to prove (d).

From (3.39), we have

$$\begin{aligned}\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\hat{\sigma}_\varepsilon^2 &= \left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)\right\}^2\hat{\sigma}_\varepsilon^2 \\ &\quad +2\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)\right\}(\iota'\psi)\hat{\sigma}_\varepsilon^2+(\iota'\psi)^2\hat{\sigma}_\varepsilon^2 \\ &= (a)+2\cdot(b)+(c), \quad \text{say.}\end{aligned}$$

For (a), we obtain

$$E[(a)]=(1-\iota'\phi)^2\sigma_\varepsilon^2-\frac{p_T+m+1}{T-p_T}(1-\iota'\phi)^2\sigma_\varepsilon^2+\frac{2}{T-p_T}\sigma_\varepsilon^2(1-\iota'\phi)\iota'\left(K_{p_T}^{(m)}+B_{p_T}^{(m)}\phi\right)+o(p_T/T),$$

using the result of Lemma 2" (b).

For (b), we need to calculate $E[(\iota'\psi)\hat{\sigma}_\varepsilon^2]$ up to $O(p_T/T)$. Since

$$\begin{aligned}\sqrt{T-p_T}\,\iota'\psi &= \iota'\hat{R}^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t+\zeta_T\right)+o_p(1), \\ \sqrt{T-p_T}\,(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2) &= \frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}(\varepsilon_t^2-\sigma_\varepsilon^2)+o_p(1),\end{aligned}$$

we have

$$\begin{aligned}&\iota'\psi(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2) \\ =\ &\frac{1}{T-p_T}\iota'\left\{R^{-1}+(\hat{R}^{-1}-R^{-1})\right\}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t+\zeta_T\right) \\ &\quad\times\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}(\varepsilon_t^2-\sigma_\varepsilon^2)\right)+o_p(p_T/T) \\ =\ &\frac{1}{T-p_T}\iota'R^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}(\varepsilon_t^2-\sigma_\varepsilon^2)\right)+o_p(p_T/T).\end{aligned}$$

Here we have

$$E\left[\frac{1}{T-p_T}\left(\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\left(\sum_{t=p_T+1}^{T}(\varepsilon_t^2-\sigma_\varepsilon^2)\right)\right]=0$$

because ε_t is a martingale difference sequence with a finite 4th moment and satisfies $E(\varepsilon_t^2|\mathcal{F}_{t-1})=\sigma_\varepsilon^2$ and $E(\varepsilon_t^3|\mathcal{F}_{t-1})=\kappa_3$. Therefore, we have $E[\iota'\psi(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2)]=o(p_T/T)$, and thus $E[\iota'\psi\hat{\sigma}_\varepsilon^2]=E[\iota'\psi(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2)]+E[\iota'\psi]\sigma_\varepsilon^2=o(p_T/T)$, and $E[(b)]=o(p_T/T)$.

For (c), since

$$\begin{aligned}(\iota'\psi)^2\hat{\sigma}_\varepsilon^2 &= (\iota'\psi)^2\left\{\sigma_\varepsilon^2 + O_p(T^{-1/2})\right\} \\ &= (\iota'\psi)^2\sigma_\varepsilon^2 + O_p(p_T/T^{3/2})\end{aligned}$$

and $E[(\iota'\psi)^2] = (T-p_T)^{-1}\sigma_\varepsilon^2\iota'R^{-1}\iota + o(p_T/T)$, we have $E[(c)] = (T-p_T)^{-1}\sigma_\varepsilon^2\iota'R^{-1}\iota + o(p_T/T)$.

Using the results above and Lemma 2" (a) and (b), we obtain the desired result. ■

Proof of Theorem 2"

Here we slightly modify the relation (3.8). When $X - E(X) = O_p(p_T/\sqrt{T})$, $Y - E(Y) = O_p(p_T/\sqrt{T})$, $E(X) \neq 0$, and $E(Y) \neq 0$, we have

$$E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)}\left[1 - \frac{Cov(X,Y)}{E(X)E(Y)} + \frac{Var(Y)}{\{E(Y)\}^2}\right] + o(p_T/T),$$

because $p_T^4/T \to 0$. Therefore, using the results of Lemma 2", we obtain the desired result. ■

Table 3.1: Values of $K_p^{(m)}$ and $B_p^{(m)}$ for $p = 1, \cdots, 5$

p	$K_p^{(m)}$	$B_p^{(m)}$
1	$m+1$	$m+3$
2	$\begin{bmatrix} m+1 \\ m+2 \end{bmatrix}$	$\begin{bmatrix} 1 & m+1 \\ 0 & m+4 \end{bmatrix}$
3	$\begin{bmatrix} m+1 \\ m+2 \\ m+1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & m+2 \\ -m-1 & m+4 & m+1 \\ 0 & 0 & m+5 \end{bmatrix}$
4	$\begin{bmatrix} m+1 \\ m+2 \\ m+1 \\ m+2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & m+1 \\ -m-1 & 2 & m+1 & m+2 \\ -m-2 & 0 & m+5 & m+1 \\ 0 & 0 & 0 & m+6 \end{bmatrix}$
5	$\begin{bmatrix} m+1 \\ m+2 \\ m+1 \\ m+2 \\ m+1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 & m+2 \\ -m-1 & 2 & 0 & m+2 & m+1 \\ -m-2 & -m-1 & m+5 & m+1 & m+2 \\ -m-1 & 0 & 0 & m+6 & m+1 \\ 0 & 0 & 0 & 0 & m+7 \end{bmatrix}$

Table 3.2: Bias and MSE of the long-run variance estimators under DGP1 (1 break) with AR(1) errors: $u_t = \phi u_{t-1} + \varepsilon_t$

			Bias					MSE				
			$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$
$T = 100$	$\delta = 1$	$\hat{\omega}_{kernel}$	−0.069	−0.193	−0.278	−0.380	−0.562	0.040	0.082	0.137	0.216	0.378
		$\hat{\omega}_{AR}$	−0.050	−0.230	−0.163	−0.211	−0.343	0.033	0.140	0.135	0.171	0.296
		$\hat{\omega}_{AR,BC}$	−0.030	−0.198	−0.096	−0.111	−0.173	0.034	0.148	0.150	0.197	0.403
	$\delta = 2$	$\hat{\omega}_{kernel}$	−0.050	−0.168	−0.244	−0.336	−0.520	0.038	0.077	0.127	0.200	0.360
		$\hat{\omega}_{AR}$	−0.037	−0.201	−0.124	−0.155	−0.270	0.032	0.137	0.132	0.171	0.315
		$\hat{\omega}_{AR,BC}$	−0.016	−0.167	−0.051	−0.043	−0.066	0.035	0.149	0.158	0.220	0.533
	known	$\hat{\omega}_{kernel}$	−0.038	−0.155	−0.230	−0.326	−0.519	0.038	0.075	0.124	0.198	0.360
		$\hat{\omega}_{AR}$	−0.026	−0.186	−0.104	−0.141	−0.268	0.033	0.137	0.131	0.174	0.318
		$\hat{\omega}_{AR,BC}$	−0.005	−0.151	−0.029	−0.026	−0.062	0.036	0.151	0.160	0.230	0.540
$T = 200$	$\delta = 1$	$\hat{\omega}_{kernel}$	−0.034	−0.130	−0.183	−0.251	−0.380	0.020	0.047	0.077	0.123	0.226
		$\hat{\omega}_{AR}$	−0.024	−0.118	−0.072	−0.104	−0.180	0.015	0.077	0.060	0.090	0.171
		$\hat{\omega}_{AR,BC}$	−0.014	−0.097	−0.035	−0.046	−0.068	0.016	0.080	0.063	0.098	0.212
	$\delta = 2$	$\hat{\omega}_{kernel}$	−0.025	−0.117	−0.166	−0.227	−0.352	0.020	0.045	0.074	0.117	0.220
		$\hat{\omega}_{AR}$	−0.018	−0.102	−0.054	−0.078	−0.138	0.015	0.076	0.060	0.090	0.178
		$\hat{\omega}_{AR,BC}$	−0.008	−0.080	−0.016	−0.017	−0.016	0.015	0.080	0.066	0.103	0.240
	known	$\hat{\omega}_{kernel}$	−0.019	−0.110	−0.158	−0.221	−0.351	0.020	0.044	0.072	0.116	0.220
		$\hat{\omega}_{AR}$	−0.013	−0.093	−0.045	−0.069	−0.136	0.015	0.075	0.061	0.091	0.179
		$\hat{\omega}_{AR,BC}$	−0.002	−0.070	−0.006	−0.008	−0.014	0.016	0.080	0.067	0.105	0.242

Table 3.3: Bias and MSE of the long-run variance estimators under DGP2 (2 breaks) with AR(1) errors: $u_t = \phi u_{t-1} + \varepsilon_t$

			Bias					MSE				
			$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$
$T = 100$	$\delta = 1$	$\hat{\omega}_{kernel}$	−0.120	−0.259	−0.367	−0.493	−0.684	0.046	0.101	0.177	0.286	0.495
		$\hat{\omega}_{AR}$	−0.089	−0.303	−0.274	−0.345	−0.512	0.039	0.152	0.162	0.202	0.350
		$\hat{\omega}_{AR,BC}$	−0.060	−0.262	−0.182	−0.217	−0.336	0.039	0.157	0.165	0.192	0.348
	$\delta = 2$	$\hat{\omega}_{kernel}$	−0.085	−0.215	−0.310	−0.422	−0.634	0.040	0.087	0.148	0.239	0.447
		$\hat{\omega}_{AR}$	−0.063	−0.256	−0.203	−0.253	−0.412	0.035	0.142	0.141	0.177	0.318
		$\hat{\omega}_{AR,BC}$	−0.033	−0.209	−0.098	−0.098	−0.168	0.038	0.157	0.166	0.210	0.451
	known	$\hat{\omega}_{kernel}$	−0.061	−0.189	−0.283	−0.405	−0.633	0.039	0.081	0.139	0.232	0.447
		$\hat{\omega}_{AR}$	−0.042	−0.225	−0.165	−0.226	−0.409	0.033	0.138	0.135	0.176	0.321
		$\hat{\omega}_{AR,BC}$	−0.011	−0.172	−0.052	−0.060	−0.162	0.038	0.159	0.174	0.230	0.474
$T = 200$	$\delta = 1$	$\hat{\omega}_{kernel}$	−0.061	−0.171	−0.242	−0.330	−0.488	0.021	0.055	0.093	0.154	0.287
		$\hat{\omega}_{AR}$	−0.042	−0.171	−0.131	−0.184	−0.300	0.016	0.085	0.064	0.096	0.184
		$\hat{\omega}_{AR,BC}$	−0.027	−0.139	−0.073	−0.099	−0.154	0.016	0.087	0.062	0.093	0.192
	$\delta = 2$	$\hat{\omega}_{kernel}$	−0.043	−0.145	−0.207	−0.284	−0.442	0.020	0.049	0.083	0.135	0.263
		$\hat{\omega}_{AR}$	−0.030	−0.138	−0.094	−0.131	−0.225	0.015	0.079	0.060	0.091	0.176
		$\hat{\omega}_{AR,BC}$	−0.014	−0.103	−0.033	−0.037	−0.052	0.016	0.084	0.064	0.102	0.235
	known	$\hat{\omega}_{kernel}$	−0.031	−0.131	−0.191	−0.273	−0.441	0.020	0.047	0.079	0.132	0.263
		$\hat{\omega}_{AR}$	−0.020	−0.119	−0.075	−0.115	−0.222	0.015	0.077	0.060	0.090	0.177
		$\hat{\omega}_{AR,BC}$	−0.004	−0.083	−0.012	−0.018	−0.048	0.016	0.083	0.068	0.107	0.238

Table 3.4: Bias and MSE of the long-run variance estimators under DGP1 (1 break) with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t$ with $\phi_2 = -0.3$

			Bias					MSE				
			$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = 0.7$	$\phi_1 = 0.9$	$\phi_1 = 1.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = 0.7$	$\phi_1 = 0.9$	$\phi_1 = 1.1$
$T = 100$	$\delta = 1$	$\hat{\omega}_{kernel}$	0.218	−0.010	−0.145	−0.276	−0.488	0.093	0.064	0.104	0.175	0.322
		$\hat{\omega}_{AR}$	−0.061	−0.059	−0.076	−0.105	−0.201	0.121	0.154	0.175	0.215	0.300
		$\hat{\omega}_{AR,BC}$	−0.009	0.009	0.010	0.015	−0.004	0.137	0.183	0.220	0.301	0.526
	$\delta = 2$	$\hat{\omega}_{kernel}$	0.239	0.022	−0.106	−0.233	−0.447	0.105	0.070	0.103	0.169	0.318
		$\hat{\omega}_{AR}$	−0.024	−0.017	−0.026	−0.064	−0.170	0.133	0.167	0.195	0.230	0.321
		$\hat{\omega}_{AR,BC}$	0.031	0.055	0.067	0.063	0.042	0.155	0.206	0.258	0.341	0.620
	known	$\hat{\omega}_{kernel}$	0.251	0.034	−0.094	−0.222	−0.444	0.112	0.074	0.106	0.172	0.319
		$\hat{\omega}_{AR}$	−0.017	−0.011	−0.027	−0.065	−0.164	0.135	0.168	0.192	0.227	0.329
		$\hat{\omega}_{AR,BC}$	0.039	0.062	0.065	0.062	0.051	0.159	0.208	0.254	0.339	0.646
$T = 200$	$\delta = 1$	$\hat{\omega}_{kernel}$	0.177	0.018	−0.071	−0.158	−0.304	0.059	0.044	0.065	0.104	0.199
		$\hat{\omega}_{AR}$	−0.044	−0.051	−0.062	−0.079	−0.121	0.050	0.057	0.069	0.090	0.157
		$\hat{\omega}_{AR,BC}$	−0.014	−0.016	−0.019	−0.020	−0.017	0.053	0.061	0.075	0.102	0.204
	$\delta = 2$	$\hat{\omega}_{kernel}$	0.190	0.036	−0.052	−0.135	−0.278	0.064	0.047	0.066	0.104	0.200
		$\hat{\omega}_{AR}$	−0.028	−0.033	−0.042	−0.056	−0.107	0.051	0.059	0.071	0.092	0.151
		$\hat{\omega}_{AR,BC}$	0.002	0.003	0.002	0.005	−0.001	0.055	0.064	0.080	0.108	0.198
	known	$\hat{\omega}_{kernel}$	0.195	0.041	−0.046	−0.128	−0.275	0.067	0.049	0.067	0.105	0.202
		$\hat{\omega}_{AR}$	−0.025	−0.030	−0.039	−0.057	−0.105	0.051	0.058	0.069	0.089	0.151
		$\hat{\omega}_{AR,BC}$	0.006	0.006	0.005	0.004	0.001	0.055	0.064	0.078	0.104	0.199

Table 3.5: Bias and MSE of the long-run variance estimators under DGP2 (2 break) with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t$ with $\phi_2 = -0.3$

			Bias					MSE				
			$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = 0.7$	$\phi_1 = 0.9$	$\phi_1 = 1.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = 0.7$	$\phi_1 = 0.9$	$\phi_1 = 1.1$
$T = 100$	$\delta = 1$	$\hat{\omega}_{kernel}$	0.167	−0.088	−0.247	−0.404	−0.630	0.067	0.057	0.120	0.223	0.434
		$\hat{\omega}_{AR}$	−0.142	−0.153	−0.179	−0.232	−0.368	0.113	0.145	0.171	0.208	0.292
		$\hat{\omega}_{AR,BC}$	−0.068	−0.056	−0.058	−0.073	−0.140	0.122	0.168	0.212	0.280	0.428
	$\delta = 2$	$\hat{\omega}_{kernel}$	0.202	−0.033	−0.181	−0.333	−0.587	0.084	0.059	0.106	0.192	0.403
		$\hat{\omega}_{AR}$	−0.077	−0.079	−0.098	−0.153	−0.307	0.121	0.154	0.176	0.209	0.288
		$\hat{\omega}_{AR,BC}$	0.004	0.030	0.038	0.026	−0.043	0.149	0.205	0.253	0.338	0.544
	known	$\hat{\omega}_{kernel}$	0.227	−0.011	−0.159	−0.317	−0.584	0.097	0.064	0.107	0.193	0.403
		$\hat{\omega}_{AR}$	−0.061	−0.068	−0.095	−0.152	−0.297	0.124	0.153	0.176	0.207	0.296
		$\hat{\omega}_{AR,BC}$	0.023	0.043	0.042	0.030	−0.025	0.156	0.207	0.257	0.341	0.594
$T = 200$	$\delta = 1$	$\hat{\omega}_{kernel}$	0.145	−0.033	−0.140	−0.247	−0.427	0.045	0.038	0.068	0.121	0.247
		$\hat{\omega}_{AR}$	−0.085	−0.097	−0.115	−0.145	−0.214	0.050	0.058	0.070	0.092	0.157
		$\hat{\omega}_{AR,BC}$	−0.040	−0.044	−0.050	−0.058	−0.067	0.051	0.059	0.073	0.098	0.192
	$\delta = 2$	$\hat{\omega}_{kernel}$	0.168	0.001	−0.100	−0.203	−0.384	0.054	0.041	0.065	0.112	0.235
		$\hat{\omega}_{AR}$	−0.055	−0.063	−0.077	−0.100	−0.180	0.050	0.058	0.070	0.092	0.149
		$\hat{\omega}_{AR,BC}$	−0.007	−0.007	−0.008	−0.007	−0.026	0.055	0.064	0.080	0.112	0.198
	known	$\hat{\omega}_{kernel}$	0.179	0.012	−0.088	−0.191	−0.381	0.059	0.043	0.066	0.112	0.236
		$\hat{\omega}_{AR}$	−0.047	−0.055	−0.071	−0.098	−0.177	0.050	0.057	0.066	0.087	0.149
		$\hat{\omega}_{AR,BC}$	0.001	0.001	−0.002	−0.005	−0.022	0.055	0.064	0.077	0.105	0.202

Table 3.6: Bias and MSE of the long-run variance estimators under DGP1 (1 break) with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t$ with $\phi_2 = 0.3$

			Bias					MSE				
			$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$
$T = 100$	$\delta = 1$	$\hat{\omega}_{kernel}$	−0.312	−0.437	−0.519	−0.510	−0.578	0.122	0.205	0.304	0.319	0.417
		$\hat{\omega}_{AR}$	−0.219	−0.258	−0.351	−0.337	−0.432	0.174	0.183	0.288	0.285	0.382
		$\hat{\omega}_{AR,BC}$	−0.151	−0.192	−0.273	−0.210	−0.232	0.194	0.208	0.336	0.351	0.573
	$\delta = 2$	$\hat{\omega}_{kernel}$	−0.292	−0.421	−0.495	−0.475	−0.545	0.111	0.192	0.285	0.297	0.406
		$\hat{\omega}_{AR}$	−0.179	−0.224	−0.301	−0.274	−0.375	0.171	0.179	0.284	0.285	0.404
		$\hat{\omega}_{AR,BC}$	−0.106	−0.154	−0.213	−0.127	−0.133	0.202	0.215	0.354	0.401	0.771
	known	$\hat{\omega}_{kernel}$	−0.286	−0.416	−0.489	−0.469	−0.545	0.108	0.188	0.281	0.294	0.406
		$\hat{\omega}_{AR}$	−0.164	−0.206	−0.281	−0.262	−0.374	0.171	0.182	0.286	0.289	0.404
		$\hat{\omega}_{AR,BC}$	−0.089	−0.131	−0.188	−0.110	−0.133	0.206	0.225	0.365	0.419	0.773
$T = 200$	$\delta = 1$	$\hat{\omega}_{kernel}$	−0.244	−0.387	−0.460	−0.406	−0.446	0.076	0.157	0.241	0.210	0.279
		$\hat{\omega}_{AR}$	−0.083	−0.104	−0.136	−0.148	−0.223	0.072	0.085	0.124	0.137	0.224
		$\hat{\omega}_{AR,BC}$	−0.038	−0.054	−0.073	−0.058	−0.068	0.077	0.093	0.138	0.160	0.310
	$\delta = 2$	$\hat{\omega}_{kernel}$	−0.234	−0.380	−0.447	−0.387	−0.425	0.072	0.152	0.230	0.199	0.271
		$\hat{\omega}_{AR}$	−0.066	−0.085	−0.112	−0.116	−0.186	0.072	0.086	0.124	0.140	0.233
		$\hat{\omega}_{AR,BC}$	−0.020	−0.033	−0.046	−0.020	−0.018	0.079	0.096	0.143	0.174	0.352
	known	$\hat{\omega}_{kernel}$	−0.230	−0.377	−0.443	−0.384	−0.425	0.070	0.150	0.228	0.197	0.271
		$\hat{\omega}_{AR}$	−0.058	−0.075	−0.101	−0.109	−0.185	0.073	0.087	0.124	0.142	0.233
		$\hat{\omega}_{AR,BC}$	−0.011	−0.022	−0.033	−0.012	−0.017	0.081	0.099	0.145	0.179	0.354

Table 3.7: Bias and MSE of the long-run variance estimators under DGP2 (2 break) with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t$ with $\phi_2 = 0.3$

			Bias					MSE				
			$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$
$T = 100$	$\delta = 1$	$\hat{\omega}_{kernel}$	−0.362	−0.476	−0.576	−0.603	−0.702	0.153	0.239	0.355	0.399	0.525
		$\hat{\omega}_{AR}$	−0.327	−0.362	−0.478	−0.501	−0.608	0.201	0.210	0.326	0.350	0.450
		$\hat{\omega}_{AR,BC}$	−0.243	−0.284	−0.398	−0.371	−0.429	0.209	0.218	0.354	0.357	0.461
	$\delta = 2$	$\hat{\omega}_{kernel}$	−0.324	−0.447	−0.538	−0.547	−0.664	0.129	0.213	0.319	0.350	0.490
		$\hat{\omega}_{AR}$	−0.257	−0.301	−0.402	−0.399	−0.539	0.180	0.189	0.302	0.309	0.431
		$\hat{\omega}_{AR,BC}$	−0.157	−0.209	−0.297	−0.219	−0.294	0.209	0.224	0.382	0.405	0.653
	known	$\hat{\omega}_{kernel}$	−0.312	−0.436	−0.526	−0.537	−0.663	0.122	0.205	0.310	0.343	0.490
		$\hat{\omega}_{AR}$	−0.227	−0.268	−0.367	−0.378	−0.538	0.177	0.186	0.297	0.307	0.431
		$\hat{\omega}_{AR,BC}$	−0.119	−0.164	−0.247	−0.187	−0.291	0.218	0.239	0.402	0.432	0.658
$T = 200$	$\delta = 1$	$\hat{\omega}_{kernel}$	−0.275	−0.408	−0.501	−0.469	−0.540	0.091	0.174	0.274	0.254	0.341
		$\hat{\omega}_{AR}$	−0.145	−0.173	−0.227	−0.252	−0.360	0.079	0.093	0.142	0.150	0.239
		$\hat{\omega}_{AR,BC}$	−0.079	−0.100	−0.140	−0.131	−0.171	0.080	0.094	0.148	0.154	0.271
	$\delta = 2$	$\hat{\omega}_{kernel}$	−0.254	−0.393	−0.475	−0.432	−0.506	0.081	0.162	0.252	0.228	0.320
		$\hat{\omega}_{AR}$	−0.108	−0.134	−0.176	−0.188	−0.297	0.074	0.088	0.133	0.140	0.235
		$\hat{\omega}_{AR,BC}$	−0.039	−0.056	−0.081	−0.049	−0.075	0.080	0.098	0.152	0.172	0.341
	known	$\hat{\omega}_{kernel}$	−0.246	−0.388	−0.467	−0.426	−0.506	0.077	0.158	0.246	0.224	0.319
		$\hat{\omega}_{AR}$	−0.092	−0.114	−0.154	−0.175	−0.296	0.073	0.087	0.129	0.140	0.235
		$\hat{\omega}_{AR,BC}$	−0.020	−0.032	−0.054	−0.033	−0.074	0.083	0.101	0.154	0.179	0.343

Table 3.8: Bias and MSE of the long-run variance estimators under DGP1 (1 break) with MA(1) errors: $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

			Bias					MSE				
			$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$
$T = 100$	$\delta = 1$	$\hat{\omega}_{kernel}$	9.398	0.492	−0.069	−0.173	−0.207	94.996	0.359	0.039	0.085	0.114
		$\hat{\omega}_{AR}$	3.958	0.332	−0.052	−0.080	−0.215	22.291	0.345	0.032	0.147	0.227
		$\hat{\omega}_{AR,BC}$	4.157	0.370	−0.032	−0.014	−0.110	24.245	0.378	0.033	0.172	0.295
	$\delta = 2$	$\hat{\omega}_{kernel}$	9.399	0.519	−0.050	−0.143	−0.172	95.003	0.388	0.038	0.081	0.109
		$\hat{\omega}_{AR}$	3.973	0.368	−0.039	−0.039	−0.164	22.451	0.376	0.030	0.151	0.238
		$\hat{\omega}_{AR,BC}$	4.172	0.407	−0.018	0.031	−0.049	24.418	0.413	0.031	0.184	0.341
	known	$\hat{\omega}_{kernel}$	9.454	0.533	−0.038	−0.128	−0.157	96.123	0.405	0.038	0.081	0.109
		$\hat{\omega}_{AR}$	3.783	0.379	−0.028	−0.018	−0.148	20.869	0.394	0.031	0.155	0.250
		$\hat{\omega}_{AR,BC}$	3.977	0.418	−0.008	0.054	−0.024	22.711	0.432	0.033	0.193	0.372
$T = 200$	$\delta = 1$	$\hat{\omega}_{kernel}$	7.183	0.397	−0.035	−0.106	−0.123	53.887	0.213	0.020	0.050	0.067
		$\hat{\omega}_{AR}$	2.625	0.286	−0.024	−0.014	−0.101	8.829	0.188	0.014	0.086	0.142
		$\hat{\omega}_{AR,BC}$	2.716	0.308	−0.014	0.024	−0.028	9.375	0.201	0.014	0.094	0.172
	$\delta = 2$	$\hat{\omega}_{kernel}$	7.194	0.411	−0.026	−0.090	−0.105	54.044	0.225	0.020	0.049	0.066
		$\hat{\omega}_{AR}$	2.674	0.305	−0.018	0.006	−0.076	9.169	0.200	0.014	0.088	0.146
		$\hat{\omega}_{AR,BC}$	2.765	0.326	−0.008	0.045	−0.001	9.731	0.214	0.014	0.099	0.182
	known	$\hat{\omega}_{kernel}$	7.215	0.417	−0.020	−0.083	−0.098	54.361	0.230	0.020	0.049	0.066
		$\hat{\omega}_{AR}$	2.600	0.308	−0.012	0.014	−0.067	8.700	0.203	0.015	0.092	0.149
		$\hat{\omega}_{AR,BC}$	2.690	0.329	−0.002	0.053	0.011	9.240	0.217	0.015	0.103	0.189

Table 3.9: Bias and MSE of the long-run variance estimators under DGP2 (2 breaks) with MA(1) errors: $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

			Bias					MSE				
			$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$
$T = 100$	$\delta = 1$	$\hat{\omega}_{kernel}$	9.376	0.425	−0.120	−0.253	−0.306	94.611	0.291	0.046	0.105	0.144
		$\hat{\omega}_{AR}$	4.268	0.239	−0.090	−0.197	−0.354	25.486	0.272	0.038	0.154	0.238
		$\hat{\omega}_{AR,BC}$	4.613	0.301	−0.061	−0.106	−0.231	29.163	0.316	0.037	0.169	0.261
	$\delta = 2$	$\hat{\omega}_{kernel}$	9.302	0.476	−0.085	−0.201	−0.245	93.110	0.341	0.040	0.089	0.121
		$\hat{\omega}_{AR}$	4.113	0.314	−0.064	−0.123	−0.269	23.770	0.328	0.034	0.145	0.224
		$\hat{\omega}_{AR,BC}$	4.449	0.379	−0.034	−0.020	−0.122	27.229	0.384	0.034	0.181	0.300
	known	$\hat{\omega}_{kernel}$	9.432	0.503	−0.061	−0.172	−0.217	95.694	0.372	0.038	0.085	0.118
		$\hat{\omega}_{AR}$	3.826	0.331	−0.043	−0.084	−0.244	21.216	0.350	0.032	0.149	0.238
		$\hat{\omega}_{AR,BC}$	4.148	0.397	−0.013	0.027	−0.076	24.355	0.410	0.034	0.199	0.369
$T = 200$	$\delta = 1$	$\hat{\omega}_{kernel}$	7.136	0.360	−0.061	−0.155	−0.185	53.195	0.183	0.022	0.057	0.077
		$\hat{\omega}_{AR}$	2.645	0.238	−0.042	−0.076	−0.189	8.957	0.164	0.016	0.084	0.139
		$\hat{\omega}_{AR,BC}$	2.797	0.274	−0.027	−0.018	−0.088	9.886	0.183	0.016	0.091	0.160
	$\delta = 2$	$\hat{\omega}_{kernel}$	7.155	0.389	−0.043	−0.124	−0.148	53.464	0.206	0.020	0.052	0.070
		$\hat{\omega}_{AR}$	2.724	0.279	−0.030	−0.034	−0.142	9.447	0.185	0.015	0.084	0.137
		$\hat{\omega}_{AR,BC}$	2.879	0.316	−0.015	0.027	−0.033	10.418	0.208	0.015	0.098	0.176
	known	$\hat{\omega}_{kernel}$	7.202	0.401	−0.031	−0.109	−0.134	54.169	0.216	0.020	0.051	0.069
		$\hat{\omega}_{AR}$	2.591	0.284	−0.019	−0.020	−0.125	8.645	0.190	0.015	0.088	0.143
		$\hat{\omega}_{AR,BC}$	2.742	0.321	−0.004	0.043	−0.008	9.550	0.214	0.015	0.105	0.194

Chapter 4

Improving the Finite Sample Performance of Tests for a Shift in Mean

It is widely known that structural break tests based on the long-run variance estimator, which is estimated under the alternative, suffer from serious size distortion when the errors are serially correlated. In this chapter, we propose bias-corrected tests for a shift in mean by correcting the bias of the long-run variance estimator up to $O(T^{-1})$. Simulation results show that the proposed tests have good size and high power.[1]

4.1 Introduction

Testing for structural breaks has been a longstanding problem and various tests have been proposed in the econometric and statistical literature. One of the frequently used tests for parameter constancy against the general alternative is the CUSUM test based on recursive residuals proposed by Brown, Durbin, and Evans (1975), and this test was further developed based on OLS residuals by Ploberger and Krämer (1992). By specifying a random walk as the alternative, optimal tests for parameter constancy were investigated by Nyblom and Mäkeläinen (1983), Nyblom (1986, 1989), and Nabeya and Tanaka (1988), among others, while the point optimal test for general regression models was studied by Elliott and Müller

[1]The published version is Yamazaki and Kurozumi (2015b), "Improving the Finite Sample Performance of Tests for a Shift in Mean", *Journal of Statistical Planning and Inference* 167, 144–173. (DOI:10.1016/j.jspi.2015.05.002)

(2006). On the other hand, it is often the case that a one-time structural change with an unknown change point is considered as the alternative and the sup-type test by Andrews (1993) and the mean- and exponential-type tests developed by Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996) are widely used in practical analyses. For a general discussion on structural changes, see, for example, Csörgő and Horváth (1997), Perron (2006), and Aue and Horváth (2013).

In practice, when we test for structural breaks in time-series models, we need to take serial correlation into account, and thus we have to estimate the long-run variance of the errors. If we estimate the long-run variance under the null hypothesis of no structural breaks, then it is known that the above tests suffer from the so-called non-monotonic power problem, that is, the power initially rises under the alternative, but as the magnitude of the break increases, the power eventually falls and tends to zero. This problem was investigated by Vogelsang (1999), Crainiceanu and Vogelsang (2007), Deng and Perron (2008), and Perron and Yamamoto (2014). The reason for this problem is that the long-run variance estimator takes significantly large values as the magnitude of the break increases.

On the other hand, if we estimate the long-run variance under the alternative, then the tests suffer from size distortion; they tend to over-reject the null hypothesis. This is because the long-run variance is under-estimated, so that the test statistics tend to take large values under the null hypothesis of no break.

In order to cope with the problem associated with the estimation of the long-run variance, several methods have been proposed. Kejriwal (2009) proposed to estimate the long-run variance using the residuals under both the null and alternative hypotheses. By using this hybrid estimator, we can reduce size distortion, but the power becomes extremely low when the error is strongly serially correlated. Juhl and Xiao (2009) proposed to estimate the long-run variance using the residuals of the nonparametric regression to mitigate the non-monotonic power problem. However, the finite sample performance of this test crucially depends on the choice of the bandwidth in the nonparametric regression. While these papers tried to improve the accuracy of the long-run variance estimator, there are several methods with which we do not have to consistently estimate the long-run variance. Sayginsoy and Vogelsang (2011) and Yang and Vogelsang (2011) proposed fixed-b sup-Wald and fixed-b sup-LM tests, respectively, which are robust to I(0)/I(1) errors. The fixed-b framework is based on Kiefer and Vogelsang (2005), which used an inconsistent long-run variance estimator where the bandwidth is proportional to the sample size. The fixed-b sup-Wald and sup-LM tests have relatively good sizes under the null hypothesis, but there is a loss of power due to

the inconsistent estimation of the long-run variance. On the other hand, Shao and Zhang (2010) proposed a self-normalized test based on the CUSUM test. The basic idea of self-normalization is similar to the fixed-b approach. Although the finite sample performance of these tests are improved, compared to the frequently used tests, such as the original CUSUM and sup-type tests, the existing methods do not seem to be satisfactory in terms of both size and power.

In this chapter, we develop an accurate long-run variance estimator and propose to use it to improve the finite sample property of the structural change tests. This estimator can be obtained by correcting the bias up to $O(T^{-1})$, where T is the sample size. The key feature of our method is that bias correction is achieved by taking a structural break into account. The advantage of our method is that tests with our long-run variance estimator can control the empirical size well, while maintaining high power. The simulation results show that the proposed tests have a higher power than other tests, such as the fixed-b test. Moreover, the power difference between our bias-corrected tests and the original (bias-uncorrected) tests is very minor, and it becomes negligible as the sample size increases. This result is in contrast to some other tests, which suffer from asymptotic power loss.

The remainder of this chapter is organized as follows. In Section 4.2, we introduce the model and the test statistic. The derivation of the bias term is discussed in Section 4.3, and the bias correction method is explained in Section 4.4. The case with general error processes is discussed in Section 4.5. Simulation results are given in Section 4.6, and Section 4.7 concludes the chapter. All mathematical proofs are delegated to the appendix.

4.2 Model and Test Statistic

Let us consider the following mean-shift model:

$$y_t = \mu + \delta \cdot DU_t(T_b^0) + u_t, \quad t = 1, \cdots, T, \tag{4.1}$$

where $DU_t(T_b^0) = 1\{t > T_b^0\}$, and $1\{\cdot\}$ is the indicator function. We assume that u_t is a zero-mean stationary process and that the break date T_b^0 is unknown.

The testing problem is

$$H_0: \quad \delta = 0 \qquad \text{vs.} \qquad H_1: \quad \delta \neq 0. \tag{4.2}$$

Under H_0, there is no shift in mean, whereas under H_1, there is a one-time break.

In order to test for a shift in mean, we need to estimate the long-run variance of u_t defined by $\omega = \sum_{\ell=-\infty}^{\infty} E(u_t u_{t-\ell})$ for the scale adjustment, which can be consistently estimated by

the kernel method. As it is known that tests with ω estimated under the null hypothesis suffer from the non-monotonic power problem, as pointed out by Vogelsang (1999), we exclude the case where the long-run variance is estimated under the null hypothesis, and focus on the case where it is estimated under the alternative of a one-time break. That is, we consider the following kernel estimator of ω as a benchmark:

$$\hat{\omega}(T_b) = \hat{\gamma}_0 + 2\sum_{j=1}^{T-1} k\left(\frac{j}{m}\right)\hat{\gamma}_j, \tag{4.3}$$

where $k(\cdot)$ is the kernel function, m is the bandwidth, $\hat{\gamma}_j$ is the estimator of the jth autocovariance of u_t defined by $\hat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}$, the residuals $\hat{u}_t$ are obtained under the alternative with the supposed break date T_b, and

$$\hat{u}_t = \begin{cases} y_t - \bar{y}_1 & \text{for } t = 1, \cdots, T_b, \\ y_t - \bar{y}_2 & \text{for } t = T_b + 1, \cdots, T, \end{cases} \tag{4.4}$$

where $\bar{y}_1 = T_b^{-1}\sum_{t=1}^{T_b} y_t$ and $\bar{y}_2 = (T - T_b)^{-1}\sum_{t=T_b+1}^{T} y_t$. Note that T_b is specified by a researcher and it is not necessarily consistent with T_b^0. We suppress the dependency of $\hat{\gamma}_j$ and $\hat{u}_t$ on T_b for notational simplicity.

When the parametric structure is framed for u_t, we may use, instead of the kernel estimator, the autoregressive spectral density estimator of ω based on the AR(p) model given by

$$\hat{\omega}_{AR}(T_b) = \frac{\hat{\sigma}_\varepsilon^2}{\left(1 - \sum_{j=1}^{p}\hat{\phi}_j\right)^2}, \tag{4.5}$$

where $\hat{u}_t = \sum_{j=1}^{p}\hat{\phi}_j\hat{u}_{t-j} + \hat{\varepsilon}_t$ with $\hat{\phi}_j$ $(j = 1, \cdots, p)$ being the OLS estimator, and $\hat{\sigma}_\varepsilon^2 = (T-p)^{-1}\sum_{t=p+1}^{T}\hat{\varepsilon}_t^2$.

In this chapter, we mainly consider the following two structural change tests, which have been commonly used in many practical analyses, with $\hat{\omega}^*(T_b)$ denoting either $\hat{\omega}(T_b)$ in (4.3) or $\hat{\omega}_{AR}(T_b)$ in (4.5), as the estimator of ω.

Sup-Wald test

Following Andrews (1993), the sup-Wald statistic for testing problem (4.2) is given by

$$\text{sup-}W = \max_{T_b \in [\varepsilon T, (1-\varepsilon)T]} W(T_b), \quad \text{where} \quad W(T_b) = \frac{SSR_0 - SSR(T_b)}{\hat{\omega}^*(T_b)}, \tag{4.6}$$

where SSR_0 is the sum of squared residuals under H_0, $SSR(T_b)$ is the sum of squared residuals under the alternative of a one-time break with the break date T_b, and ε is the trimming parameter.

CUSUM test

The CUSUM test statistic proposed by Ploberger and Krämer (1992) is originally defined as

$$CUSUM = \max_{T_b \in [1, T-1]} \left| \frac{T^{-1/2} \sum_{t=1}^{T_b} \tilde{u}_t}{\sqrt{\tilde{\omega}}} \right|,$$

where $\tilde{u}_t$ is the residual under H_0, and the long-run variance estimator $\tilde{\omega}$ is estimated under the null hypothesis. As explained in Crainiceanu and Vogelsang (2007) and Deng and Perron (2008), this test suffers from the non-monotonic power problem because the long-run variance is estimated under the null hypothesis of no break. In order to avoid this problem, we again consider estimating the long-run variance under the alternative of a one-time break. Then, the test statistic should be modified as

$$CUSUM_{H_1} = \max_{T_b \in [1, T-1]} \left| \frac{T^{-1/2} \sum_{t=1}^{T_b} \tilde{u}_t}{\sqrt{\hat{\omega}^*(T_b)}} \right|. \tag{4.7}$$

4.3 Derivation of the Bias

In this section, we derive the bias of the reciprocal of the long-run variance estimator up to $O(T^{-1})$, under the assumption that the correct specification for u_t is the AR(p) model. The case with general error processes will be discussed later. Throughout this chapter, we define the bias as the expectation up to $O(T^{-1})$, ignoring the $o_p(T^{-1})$ terms.[2] Note that since our purpose is to control the size of the tests by precisely estimating the long-run variance, the bias is derived under the null hypothesis of no break, whereas $\hat{u}_t$ is obtained assuming a one-time break at T_b, which is given by

$$\hat{u}_t = \begin{cases} u_t - \bar{u}_1 & \text{for } t = 1, \cdots, T_b, \\ u_t - \bar{u}_2 & \text{for } t = T_b + 1, \cdots, T. \end{cases} \tag{4.8}$$

Here, we note that T_b is not the actual structural break date, but the prespecified possible break date which is necessary for calculation of the long-run variance estimator.

To derive the bias term, we make the following assumptions when $p \geq 1$:

Assumption 1 *$\{u_t\}$ follows a zero-mean stationary AR(p) process: $u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + \varepsilon_t$, where $1 - \sum_{j=1}^{p} \phi_j z^j \neq 0$ for $|z| \leq 1$, and $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$ and $E(\varepsilon_t^3 | \mathcal{F}_{t-1}) = \kappa_3$.*

[2] If we need to evaluate the expectation without ignoring the $o_p(T^{-1})$ terms, we have to make additional assumptions about the existence of higher-order moments.

Assumption 2 *$T_b/T \to \lambda \in (0,1)$ as $T \to \infty$.*

When $p = 0$, we use the following Assumption 1', instead of Assumption 1.

Assumption 1' *$u_t = \varepsilon_t$ for all t, where $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\varepsilon^2$.*

Assumptions 1 and 1' exclude the case where $\{u_t\}$ is a unit root process. Assumption 2 is standard for structural break models.

We derive the bias of the reciprocal of $\hat{\omega}_{AR}$, which is given by

$$\frac{1}{\hat{\omega}_{AR}} = \begin{cases} \dfrac{\left(1 - \sum_{j=1}^p \hat{\phi}_j\right)^2}{\hat{\sigma}_\varepsilon^2} & \text{for } p \geq 1, \\ \dfrac{1}{\hat{\sigma}_\varepsilon^2} & \text{for } p = 0. \end{cases} \tag{4.9}$$

Here, we consider the bias of the reciprocal of $\hat{\omega}_{AR}$ because the long-run variance estimator is placed in the denominator of the test statistics.

In general, when random variables X and Y satisfy $X - E(X) = O_p(T^{-1/2})$, $Y - E(Y) = O_p(T^{-1/2})$, $E(X) \neq 0$, and $E(Y) \neq 0$, the following relation holds:

$$E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)}\left[1 - \frac{Cov(X,Y)}{E(X)E(Y)} + \frac{Var(Y)}{\{E(Y)\}^2}\right] + o(T^{-1}), \tag{4.10}$$

which can be obtained by the Taylor expansion of $f(x,y) = x/y$ around $(x,y) = (E(X), E(Y))$, and by taking expectations, ignoring the $o_p(T^{-1})$ terms. See Mood, Graybill, and Boes (1974, p.181).

Therefore, in order to derive the bias of (4.9) up to $O(T^{-1})$, we need to obtain $E[(1 - \sum_{j=1}^p \hat{\phi}_j)^2]$, $E[\hat{\sigma}_\varepsilon^2]$, $Var[\hat{\sigma}_\varepsilon^2]$, and $Cov[\hat{\sigma}_\varepsilon^2, (1 - \sum_{j=1}^p \hat{\phi}_j)^2]$ for $p \geq 1$. When $p = 0$, we only need $E[\hat{\sigma}_\varepsilon^2]$ and $Var[\hat{\sigma}_\varepsilon^2]$.

The following lemma gives the results for $p \geq 1$:

Lemma 1 *Under Assumptions 1 and 2, the following relations hold:*

(a) $E\left[\left(1 - \sum_{j=1}^p \hat{\phi}_j\right)^2\right] = (1 - \iota'\phi)^2 + \frac{1}{T-p}\left\{2(1 - \iota'\phi)\iota'\left(K_p^{(1)} + B_p^{(1)}\phi\right) + \sigma_\varepsilon^2 \iota' R^{-1}\iota\right\} + o(T^{-1}),$

(b) $E\left[\hat{\sigma}_\varepsilon^2\right] = \sigma_\varepsilon^2 - \frac{p+2}{T-p}\sigma_\varepsilon^2 + o(T^{-1}),$

(c) $Var\left[\hat{\sigma}_\varepsilon^2\right] = \frac{1}{T-p}\left\{E(\varepsilon_t^4) - \sigma_\varepsilon^4\right\} + o(T^{-1}),$

(d) $Cov\left[\hat{\sigma}_\varepsilon^2, \left(1 - \sum_{j=1}^p \hat{\phi}_j\right)^2\right] = o(T^{-1}),$

where R is a $p \times p$ matrix whose (i,j) element is given by $\gamma_{|i-j|} = E(u_t u_{t-|i-j|})$, ι is a $p \times 1$ vector of ones, and $K_p^{(1)}$ and $B_p^{(1)}$ are defined in Chapter 3.

By (4.10) and Lemma 1, we obtain the first-order bias of the reciprocal of the long-run variance estimator for $p \geq 1$:

Theorem 1 *Under Assumptions 1 and 2, the expectation of $1/\hat{\omega}_{AR}$ up to $O(T^{-1})$ is given by*

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{(1-\iota'\phi)^2}{\sigma_\varepsilon^2} + \frac{1}{T-p}\left[\frac{1}{\sigma_\varepsilon^2}\left\{2(1-\iota'\phi)\iota'\left(K_p^{(1)} + B_p^{(1)}\phi\right) + \sigma_\varepsilon^2 \iota' R^{-1}\iota + (p+2)(1-\iota'\phi)^2\right\} + \frac{(1-\iota'\phi)^2}{\sigma_\varepsilon^2}\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o(T^{-1}).$$

Remark 1 *When $p = 1$, the expectation of $1/\hat{\omega}_{AR}$ is given by*

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{(1-\phi_1)^2}{\sigma_\varepsilon^2} + \frac{1}{T-1}\cdot\frac{1}{\sigma_\varepsilon^2}\left[(1-\phi_1)(8+6\phi_1) + (1-\phi_1)^2\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o(T^{-1}),$$

if the long-run variance is estimated using the residuals under the alternative.

On the other hand, if we use the residuals under the null hypothesis of no structural break to estimate the long-run variance, the expectation can be shown to be given by

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{(1-\phi_1)^2}{\sigma_\varepsilon^2} + \frac{1}{T-1}\cdot\frac{1}{\sigma_\varepsilon^2}\left[(1-\phi_1)(5+5\phi_1) + (1-\phi_1)^2\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o(T^{-1}).$$

Therefore, we can see that the first-order bias of $1/\hat{\omega}_{AR}$ with the residuals under the alternative hypothesis is larger than the one with the residuals under the null hypothesis.

Similarly, when $p = 0$, we obtain the following lemma and theorem:

Lemma 1' *Under Assumptions 1' and 2, the following relations hold:*

$$(a) \quad E\left[\hat{\sigma}_\varepsilon^2\right] = \sigma_\varepsilon^2 - \frac{2}{T}\sigma_\varepsilon^2 + o(T^{-1}),$$
$$(b) \quad Var\left[\hat{\sigma}_\varepsilon^2\right] = \frac{1}{T}\left\{E(\varepsilon_t^4) - \sigma_\varepsilon^4\right\} + o(T^{-1}).$$

Theorem 1' *Under Assumptions 1' and 2, the expectation of $1/\hat{\omega}_{AR}$ up to $O(T^{-1})$ is given by*

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{1}{\sigma_\varepsilon^2} + \frac{1}{T}\left[\frac{2}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\varepsilon^2}\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o(T^{-1}).$$

Remark 2 *The first-order bias of $1/\hat{\omega}_{AR}$ does not depend on the maintained break fraction λ.*

4.4 Bias-Corrected Test

In this section, we propose the correction of the bias of (4.9) using Theorems 1 and 1', and explain how to use our bias-corrected estimator in order to test for a shift in mean.

4.4.1 Bias correction of the reciprocal of the long-run variance estimator

In this subsection, we obtain the bias-corrected estimator of the reciprocal of the long-run variance.

Since the first-order bias of (4.9) is given by Theorems 1 and 1', the bias-corrected estimator of $1/\omega$ is

$$\left(\frac{1}{\hat{\omega}_{AR}}\right)_{BC} = \frac{1}{\hat{\omega}_{AR}} - \hat{b}, \tag{4.11}$$

where

$$\hat{b} = \begin{cases} \dfrac{1}{T-p}\left[\dfrac{1}{\hat{\sigma}_\varepsilon^2}\left\{2(1-\iota'\hat{\phi})\iota'\left(K_p^{(1)}+B_p^{(1)}\hat{\phi}\right)+\hat{\sigma}_\varepsilon^2\iota'\hat{R}^{-1}\iota \right.\right. \\ \qquad \left.\left. +(p+2)(1-\iota'\hat{\phi})^2\right\}+\dfrac{(1-\iota'\hat{\phi})^2}{\hat{\sigma}_\varepsilon^2}\left\{\dfrac{\widehat{E(\varepsilon_t^4)}}{\hat{\sigma}_\varepsilon^4}-1\right\}\right] & \text{for } p\geq 1, \\ \dfrac{1}{T}\left[\dfrac{2}{\hat{\sigma}_\varepsilon^2}+\dfrac{1}{\hat{\sigma}_\varepsilon^2}\left\{\dfrac{\widehat{E(\varepsilon_t^4)}}{\hat{\sigma}_\varepsilon^4}-1\right\}\right] & \text{for } p=0, \end{cases}$$

and $\hat{\phi}$, $\hat{\sigma}_\varepsilon^2=(T-p)^{-1}\sum_{t=p+1}^T\hat{\varepsilon}_t^2$, $\widehat{E(\varepsilon_t^4)}=(T-p)^{-1}\sum_{t=p+1}^T\hat{\varepsilon}_t^4$, and $\hat{\gamma}_{ij}$ for the (i,j) element of $\hat{R}$ are the least squares estimators of ϕ, σ_ε^2, $E(\varepsilon_t^4)$, and γ_{ij}, respectively.[3]

For example, when $p=1$, the correcting term is given by

$$\hat{b}=\frac{1}{T-1}\cdot\frac{1}{\hat{\sigma}_\varepsilon^2}\left[(1-\hat{\phi})(8+6\hat{\phi})+(1-\hat{\phi})^2\left\{\frac{\widehat{E(\varepsilon_t^4)}}{\hat{\sigma}_\varepsilon^4}-1\right\}\right].$$

4.4.2 Tests based on the bias-corrected long-run variance estimator

The bias-corrected test statistic can be obtained by using the bias-corrected estimator (4.11). For example, the bias-corrected sup-Wald test statistic is given by

$$\text{sup-}W_{BC} = \max_{T_b\in[\varepsilon T,(1-\varepsilon)T]} W_{BC}(T_b), \tag{4.12}$$

where

$$W_{BC}(T_b) = \left(\frac{1}{\hat{\omega}_{AR}}\right)_{BC}\cdot\left(SSR_0 - SSR(T_b)\right).$$

[3]Other consistent estimators can also be plugged in.

Similarly, the bias-corrected CUSUM test statistic[4] is given by

$$CUSUM_{H_1,BC} = \max_{T_b \in [\varepsilon T, (1-\varepsilon)T]} \left| \sqrt{\left(\frac{1}{\hat{\omega}_{AR}} \right)_{BC}} \cdot T^{-1/2} \sum_{t=1}^{T_b} \tilde{u}_t \right|. \tag{4.13}$$

Note that $\hat{\omega}_{AR}$ in (4.12) and (4.13) depends on T_b, so that this bias correction procedure needs to be repeatedly applied to each $\hat{\omega}_{AR}$ for all possible T_b with the repeatedly estimated lag order. However, we suppress the dependency on T_b for notational simplicity.

Since the correcting terms are $O_p(T^{-1})$, the asymptotic distribution of the test statistic under the null hypothesis is exactly the same as that of the original test, and thus we do not have to modify critical values in order to apply the bias-corrected test. Moreover, even under the alternative, it can be shown that the first-order bias is asymptotically negligible, so that there is no asymptotic power loss.

4.5 Extension to the Model with General Error Processes

In this section, we consider the case where the error term u_t is generated by a stationary AR(∞) process. In this case, we make the following assumption:

Assumption 1" *$u_t = \sum_{j=1}^{\infty} \phi_j u_{t-j} + \varepsilon_t$, where $1 - \sum_{j=1}^{\infty} \phi_j z^j \neq 0$ for $|z| \leq 1$, $\sum_{j=1}^{\infty} |\phi_j| < \infty$, and $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$ and $E(\varepsilon_t^3 | \mathcal{F}_{t-1}) = \kappa_3$.*

Although only the absolute summability of $\{\phi_j\}$ is assumed in Assumption 1", we may require the higher order summability of $\{\phi_j\}$, as explained below.

Since the error term is an infinite order AR process, we need to truncate the lag order at some point p_T and consider estimating the AR(p_T) model. The following assumption is concerned with the lag truncation point p_T.

Assumption L

$$\begin{aligned} &(a) \quad p_T \to \infty \quad \textit{and} \quad p_T^4 / T \to 0 \quad \textit{as} \quad T \to \infty. \\ &(b) \quad \textstyle\sum_{j=p_T+1}^{\infty} |\phi_j| = o(p_T / T) \qquad \textit{as} \quad T \to \infty. \end{aligned} \tag{4.14}$$

Assumption L(a) gives the upper bound of the divergence rate of p_T. This rate guarantees the consistency of the autoregressive spectral density estimator as proved by Berk

[4] We use a trimming for the CUSUM test so that Assumption 2 is satisfied.

(1974) and den Haan and Levin (1998), although condition (4.14) is stronger than theirs. Assumption L(b) not only imposes the lower bound of p_T but is also related with the higher order summability of $\{\phi_j\}$. For example, when $\sum_{j=0}^{\infty} j^{3+\alpha}|\phi_j| < \infty$ holds and p_T is greater than $O(T^{1/(4+\alpha)})$ for some $\alpha > 0$, Assumption L(b) is satisfied. Note that this assumption is satisfied if u_t is generated by a finite-order ARMA process and $p_T = O(T^{\delta})$ for some $\delta > 0$, because $|\phi_j|$ declines geometrically to zero.

The next theorem gives the bias of the reciprocal of the autoregressive spectral density estimator up to $O(p_T/T)$:

Theorem 1" *Under Assumptions 1", 2, and L, the expectation of* $1/\hat{\omega}_{AR}$ *up to* $O(p_T/T)$ *is given by*

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{(1-\iota'\phi)^2}{\sigma_\varepsilon^2} + \frac{1}{T-p_T}\left[\frac{1}{\sigma_\varepsilon^2}\left\{2(1-\iota'\phi)\iota'\left(K_{p_T}^{(1)} + B_{p_T}^{(1)}\phi\right) + \sigma_\varepsilon^2\iota' R^{-1}\iota + (p_T+2)(1-\iota'\phi)^2\right\}\right.$$
$$\left. + \frac{(1-\iota'\phi)^2}{\sigma_\varepsilon^2}\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o\left(\frac{p_T}{T}\right),$$

where

$$\phi = \begin{bmatrix} \phi_{p_T,1} \\ \phi_{p_T,2} \\ \vdots \\ \phi_{p_T,p_T} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p_T-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{p_T-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{p_T} \end{bmatrix}. \tag{4.15}$$

This first-order bias is exactly the same as the one in Theorem 1. Therefore, we can implement the bias correction as explained in Section 4.4.

4.6 Simulation Results

4.6.1 Biases and mean squared errors of the estimators of $1/\omega$

In this subsection, we investigate the finite sample performance of the estimators of the reciprocal of the long-run variance. The data generating process is as follows:

$$y_t = \mu + u_t, \quad t = 1, \cdots, T. \tag{4.16}$$

We consider the following cases for the error processes of u_t:

$$\begin{cases} \text{AR(1)} : u_t = \phi u_{t-1} + \varepsilon_t, & \varepsilon_t \sim i.i.d.\ N\left(0,\ (1-\phi)^2\right), \\ \text{AR(2)} : u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t, & \varepsilon_t \sim i.i.d.\ N\left(0,\ (1-\phi_1-\phi_2)^2\right), \\ \text{MA(1)} : u_t = \varepsilon_t + \theta\varepsilon_{t-1}, & \varepsilon_t \sim i.i.d.\ N\left(0,\ \frac{1}{(1+\theta)^2}\right), \end{cases}$$

where the variance of ε_t is selected so that $1/\omega = 1$.

We compare the biases and mean squared errors (MSEs) of the following estimators:

(i): $1/\hat{\omega}_{kernel}$: the reciprocal of the kernel estimator given by (4.3).

(ii): $1/\hat{\omega}_{AR}$: the reciprocal of the autoregressive spectral density estimator given by (4.5).

(iii): $(1/\hat{\omega}_{AR})_{BC}$: the bias-corrected estimator given by (4.11).

Throughout the simulation in this subsection, we set $T_b = 0.5T$. For the kernel estimator (4.3), we use the quadratic spectral kernel with the bandwidth parameter selected by Andrews' (1991) rule. When we implement the AR(p) regression to obtain the autoregressive spectral density estimator, we select the lag length p by the Bayesian Information Criterion (BIC) with the maximum lag length 5. In this simulation, the number of replications is 10,000.

Tables 4.1–4.3 give the simulation results. As we can see from Table 4.1, when the error term follows a stationary AR(1) process, the bias-corrected estimator has much less bias than the other ones. Moreover, the bias-corrected estimator has less MSE, compared to other estimators. Table 4.2 shows the results with AR(2) errors. In this case, we can see similar results. When u_t follows an MA(1) process, we can see from Table 4.3 that the bias-corrected estimator has less bias in most cases, and the MSE of the bias-corrected estimator is comparable to that of other estimators. Overall, the bias-corrected estimator performs well in finite samples.

4.6.2 Finite sample performance of the tests

In this subsection, we investigate the finite sample performance of the tests through a Monte Carlo experiment. The data generating process is as follows:

$$y_t = \mu + \delta \cdot DU_t(T_b^0) + u_t, \quad \mu = 0, \quad \delta = \frac{c}{\sqrt{T}}, \quad T_b^0 = 0.5T.$$

We consider the following four processes of u_t:

$$\begin{cases} \text{AR}(1): u_t = \phi u_{t-1} + \varepsilon_t, & \varepsilon_t \sim i.i.d.\ N\left(0,\ 1-\phi^2\right), \\ \text{AR}(2): u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t, & \varepsilon_t \sim i.i.d.\ N\left(0,\ \frac{(1+\phi_2)\{(1-\phi_2)^2-\phi_1^2\}}{1-\phi_2}\right), \\ \text{MA}(1): u_t = \varepsilon_t + \theta\varepsilon_{t-1}, & \varepsilon_t \sim i.i.d.\ N\left(0,\ \frac{1}{1+\theta^2}\right), \\ \text{ARMA}(3,3): u_t = \phi^3 u_{t-3} + \varepsilon_t + \theta\varepsilon_{t-3}, & \varepsilon_t \sim i.i.d.\ N\left(0,\ \frac{1-\phi^6}{1+2\phi^3\theta+\theta^2}\right), \end{cases}$$

where the variance of ε_t is selected so that $Var(u_t) = 1$.

In this subsection, we compare the sizes and powers of the following tests:

(sup-Wald test)

(i): sup-W: the sup-Wald test (4.6) with the long-run variance estimator given by (4.3).

(ii): sup-W_{AR}: the sup-Wald test (4.6) with the long-run variance estimator given by (4.5).

(iii): sup-W_{BC}: the bias-corrected sup-Wald test (4.12).

(iv): sup-W_{kej}: the sup-Wald test (4.6) with the hybrid long-run variance estimator by Kejriwal (2009).

(v): fixed-b sup-W: the fixed-b sup-Wald test based on Sayginsoy and Vogelsang (2011), where we use the J statistic as a scaling factor. We use the Daniell kernel with the feasible integrated power optimal data-dependent bandwidth as described in Sayginsoy and Vogelsang (2011), and a 10% trimming for this test.

(CUSUM test)

(i): $CUSUM_{H_1}$: the CUSUM test with a 15% trimming, which is given by

$$CUSUM_{H_1} = \max_{T_b \in [0.15T, 0.85T]} \left| \frac{T^{-1/2} \sum_{t=1}^{T_b} \tilde{u}_t}{\sqrt{\hat{\omega}^*(T_b)}} \right|. \tag{4.17}$$

We use the long-run variance estimator given by (4.3).

(ii): $CUSUM_{H_1,AR}$: the CUSUM test (4.17) with the long-run variance estimator given by (4.5).

(iii): $CUSUM_{H_1,BC}$: the bias-corrected CUSUM test (4.13).

(iv): SN: the self-normalizing method by Shao and Zhang (2010).

For the kernel estimator (4.3), we use the quadratic spectral kernel with the bandwidth parameter selected by Andrews' (1991) rule to estimate the long-run variance, except for the fixed-b sup-Wald test. When we implement the AR(p) regression to obtain the autoregressive spectral density estimator, we select the lag length p by the BIC, where the maximum lag length is 5. For the sup-Wald and CUSUM tests, we use a 15% trimming, except for the fixed-b sup-Wald test. The number of replications is 2,000, and the nominal size is 0.05.

Empirical sizes of the tests

Tables 4.4–4.9 show the empirical sizes of the tests. When the error follows an AR(1) process, we can see from Table 4.4 that the original sup-Wald test tends to over-reject the null hypothesis as ϕ gets larger. By using the autoregressive spectral density estimator, we can mitigate the over-rejection problem, except the case where $\phi = 0.2$, but the test still has size distortion. We need to note that, when $\phi = 0.2$, the sup-W_{AR} test has a larger size distortion than the original sup-Wald test because the lag length selected by the BIC is sometimes too short in finite samples. The bias-corrected sup-Wald test performs much better than the bias-uncorrected tests, in particular when u_t is strongly serially correlated. The empirical sizes of the sup-Wald test based on Kejriwal (2009) and the fixed-b sup-Wald test are relatively close to the nominal one, although the fixed-b test is rather conservative. We observe similar results for the CUSUM test. The bias-corrected CUSUM test ($CUSUM_{H_1,BC}$) has much less size distortion than the bias-uncorrected CUSUM tests ($CUSUM_{H_1}$ and $CUSUM_{H_1,AR}$), unless $\phi = 0.2$. Moreover, the $CUSUM_{H_1,BC}$ test performs better than the self-normalization based test when ϕ is large. As the sample size increases, the sizes of all tests get closer to the nominal one.

Tables 4.5 and 4.6 show the empirical sizes with AR(2) errors. We can see that the relative performance holds when $\phi_2 = -0.3$, compared to the case with AR(1), whereas when $\phi_2 = 0.3$ and $T = 100$, all the tests tend to over-reject the null hypothesis, including the bias-corrected tests. In this case, only the fixed-b sup-Wald test has relatively good size. However, as the sample size increases, the performance of the bias-corrected tests greatly improves, and it is superior to that of the other tests. When the error follows an MA(1) process, we can see from Table 4.7 that the bias-corrected tests have good finite sample properties.

Tables 4.8 and 4.9 give the results with ARMA(3,3) errors. Note that an important feature of this error process is that the autocovariance satisfies $\gamma_j = E(u_t u_{t-j}) \neq 0$ if and only if $j = 0, \pm 3, \pm 6, \cdots$. Therefore, it is difficult to fit with the autoregressive process.

In this case, we can see from Table 4.8 that most tests tend to under-reject the null hypothesis when $\theta = -0.3$ and $\phi \leq 0.6$. When $\theta = -0.3$ and $\phi = 0.8$, all tests are over-sized. In this case, the fixed-b sup-Wald test and the self-normalization based test have relatively good size.

When $\theta = 0.3$, we can see from Table 4.9 that the original sup-Wald test and Kejriwal's (2009) test are severely over-sized, especially when ϕ is large. The sup-Wald test with the autoregressive spectral density estimator performs better than the original sup-Wald test

when ϕ is large. The bias-corrected sup-Wald test performs well when ϕ is large, although it tends to be over-sized when ϕ is small. The fixed-b sup-Wald test has relatively good size when ϕ is small, but it is over-sized when ϕ is large. We can see similar results for the CUSUM test.

Size-adjusted power of the tests

We compare the size-adjusted power of the tests.[5][6] Figure 4.1 shows the size-adjusted powers with AR(1) errors and $T = 100$. We can see from Figure 4.1 that, when $\phi = 0.6$, the bias-corrected sup-Wald test is more powerful than the sup-W_{kej} and fixed-b sup-Wald tests, while for the CUSUM test, the bias-corrected test performs much better than the self-normalization based test. We can see that the power difference between the bias-corrected and bias-uncorrected tests is relatively small. Similar results are obtained when $\phi = 0.8$. Although the power loss due to bias correction is slightly larger than that of the case with $\phi = 0.6$, the bias-corrected test has higher power than the other tests.

As in Figure 4.2, when $T = 200$, the power difference between the bias-corrected and bias-uncorrected tests is much smaller than the case when $T = 100$. In this case, the bias-corrected test still outperforms the other tests.

Figures 4.3 and 4.4 show the size-adjusted power of the sup-Wald tests with ARMA(3,3) errors.[7] When $\theta = -0.3$, all tests have similar size-adjusted power, although the sup-W_{kej} test has slightly non-monotonic power when $T = 100$. When $\theta = 0.3$, the size-adjusted power of the bias-corrected test is lower than that of the original sup-Wald test and the Kejriwal's (2009) test. As the sample size increases, we can see that the finite sample performance of the bias-corrected test improves.

Overall, our bias-corrected tests have good finite sample property, in terms of both size and power. Even when it is difficult to fit the AR approximation, such as the ARMA(3,3) case with $\theta = 0.3$, the bias-corrected test has less size distortion than the original tests in

[5]Because the critical value of the fixed-b sup-Wald test is data-dependent, we adjust the size of the other tests to the empirical size of the fixed-b sup-Wald test with nominal one 0.05 in the case with AR(1) errors.

[6]Since the size-adjusted powers of the sup-W_{AR} and $CUSUM_{H_1,AR}$ tests are almost the same as those of the sup-W and $CUSUM_{H_1}$ tests, respectively, in the AR(1) case, we omit the results of sup-W_{AR} and $CUSUM_{H_1,AR}$ tests.

[7]We omit the results of the fixed-b sup-Wald test because its critical value is data-dependent, and this test has serious size distortion in some cases. Our preliminary simulation results show that the size-adjusted power of the fixed-b sup-Wald test is similar to that of the sup-W_{AR} test, if the size is adjusted to the empirical size of the fixed-b sup-Wald test.

most cases, although it has low size-adjusted power, especially when the sample size is small.

Comparison of the finite sample performance of bias-corrected tests

Here, we focus on only the bias-corrected versions of the tests commonly used in the literature and compare their sizes and powers. We consider the sup-Wald test (sup-W_{BC}) by Andrews (1993), the mean-Wald test (mean-W_{BC}) and the exponential-Wald test (exp-W_{BC}) by Andrews, Lee, and Ploberger (1996), the locally best invariant test against the random walk alternative by Nabeya and Tanaka (1988) (which we denote as $LM_{H_1,BC}$), the asymptotically point optimal test against the random walk alternative by Elliott and Müller (2006) (which we denote as $qLL_{H_1,BC}$), and the $CUSUM_{H_1,BC}$ test given by (4.17). Since the original LM, qLL, and CUSUM tests use the long-run variance estimator under the null hypothesis and they have non-monotonic power, we consider estimating the long-run variance under the alternative of a one-time break. For the LM and qLL tests, we use the residuals under the alternative with break date $\hat{T}_b = \arg\min_{T_b \in [0.15T, 0.85T]} SSR(T_b)$. For the CUSUM test, we use the bias-corrected test statistic (4.13).

The empirical sizes with AR(1) errors are given in Table 4.10 (we omit the other cases to save space). We observe that the bias-corrected mean-Wald test has relatively good size, while the other tests are slightly over-sized.

The size-adjusted powers of the tests are given in Figure 4.5. We observe that the bias-corrected CUSUM test performs best, while the mean-Wald, LM, and qLL tests suffer from power loss, in particular when the errors are strongly serially correlated, or when the sample size is small.

Overall, we can see that the bias-corrected CUSUM test with the long-run variance estimated under the alternative has the best finite sample properties, against the alternative of a one-time break. However, it is not clear whether this bias-corrected CUSUM test outperforms other tests against various kinds of the alternative, such as multiple breaks or time-varying parameter models.

4.7 Conclusion

We have proposed a bias correction to the long-run variance estimator, which is estimated under the alternative hypothesis of a one-time break. We have derived the first-order bias of the reciprocal of the long-run variance estimator, taking a structural break into account. By Monte Carlo simulations, we have found that our bias-corrected tests have better finite

sample properties than the existing tests.

So far, we have considered tests for a mean shift, but it is also in our interest to consider bias correction to test for structural change in general regression models. We wish to investigate such topics in future studies.

4.8 Appendix: Proofs of Theorem 1" and Related Lemma

Because the AR(p) model is a special case of the AR(∞) model, we only prove the results for AR(∞) errors. Lemmas 1 and 1', and Theorems 1 and 1' can be proved similarly. Note that p_T becomes a fixed number for the finite order AR model and thus, for example, the order given by $o(p_T/T)$ in the following lemmas becomes $o(1/T)$ in the AR(p) case.

Lemma 1" *Under Assumptions 1", 2, and L, the following relations hold:*

(a) $$E\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] = (1-\iota'\phi)^2 + \frac{1}{T-p_T}\left\{2(1-\iota'\phi)\iota'\left(K_{p_T}^{(1)}+B_{p_T}^{(1)}\phi\right)+\sigma_\varepsilon^2\iota' R^{-1}\iota\right\}+o\left(\frac{p_T}{T}\right),$$

(b) $$E\left[\hat{\sigma}_\varepsilon^2\right] = \sigma_\varepsilon^2 - \frac{p_T+2}{T-p_T}\sigma_\varepsilon^2 + o\left(\frac{p_T}{T}\right),$$

(c) $$Var\left[\hat{\sigma}_\varepsilon^2\right] = \frac{1}{T-p_T}\left\{E(\varepsilon_t^4)-\sigma_\varepsilon^4\right\}+o(T^{-1}),$$

(d) $$Cov\left[\hat{\sigma}_\varepsilon^2, \left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] = o\left(\frac{p_T}{T}\right).$$

Proof of Lemma 1"

Since (a), (b), and (d) are proved in Lemma 2" in Chapter 3, we only need to prove (c).

Proof of (c).

Since $\hat{\sigma}_\varepsilon^2 = (T-p_T)^{-1}\sum_{t=p_T+1}^{T}\varepsilon_t^2 + O_p(p_T/T)$, we obtain

$$\begin{aligned}\sqrt{T-p_T}\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right) &= \frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}(\varepsilon_t^2-\sigma_\varepsilon^2)+O_p(p_T/\sqrt{T}) \\ &\xrightarrow{d} N\left(0, E(\varepsilon_t^4)-\sigma_\varepsilon^4\right),\end{aligned}$$

so that $Var(\hat{\sigma}_\varepsilon^2) = (T-p_T)^{-1}\left\{E(\varepsilon_t^4)-\sigma_\varepsilon^4\right\}+o(T^{-1})$. ■

Proof of Theorem 1"

When $X - E(X) = O_p(p_T/\sqrt{T})$, $Y - E(Y) = O_p(T^{-1/2})$, $E(X) \neq 0$, and $E(Y) \neq 0$, we have

$$E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)}\left[1 - \frac{Cov(X,Y)}{E(X)E(Y)} + \frac{Var(Y)}{\{E(Y)\}^2}\right] + o(p_T/T),$$

because $p_T^4/T \to 0$. Therefore, using the results of Lemma 1", we obtain the desired result. ■

Table 4.1: Bias and MSE of the estimators of $1/\omega$ with AR(1) errors: $u_t = \phi u_{t-1} + \varepsilon_t$

		Bias					MSE				
		$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$
$T = 100$	$1/\hat{\omega}_{kernel}$	0.078	0.256	0.439	0.755	1.852	0.056	0.166	0.419	1.144	6.445
	$1/\hat{\omega}_{AR}$	0.062	0.372	0.301	0.400	0.949	0.055	0.322	0.456	0.615	2.605
	$(1/\hat{\omega}_{AR})_{BC}$	0.017	0.281	0.091	0.033	0.055	0.043	0.277	0.362	0.351	1.095
$T = 200$	$1/\hat{\omega}_{kernel}$	0.038	0.165	0.260	0.415	0.875	0.024	0.078	0.168	0.382	1.515
	$1/\hat{\omega}_{AR}$	0.028	0.192	0.110	0.175	0.393	0.017	0.154	0.095	0.170	0.553
	$(1/\hat{\omega}_{AR})_{BC}$	0.006	0.136	0.007	0.006	0.009	0.016	0.144	0.075	0.120	0.301

Table 4.2: Bias and MSE of the estimators of $1/\omega$ with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t$

		Bias					MSE				
		$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = 0.7$	$\phi_1 = 0.9$	$\phi_1 = 1.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = 0.7$	$\phi_1 = 0.9$	$\phi_1 = 1.1$
		$\phi_2 = -0.3$					$\phi_2 = -0.3$				
$T = 100$	$1/\hat{\omega}_{kernel}$	-0.177	0.028	0.225	0.531	1.537	0.052	0.066	0.222	0.764	5.416
	$1/\hat{\omega}_{AR}$	0.138	0.169	0.212	0.314	0.662	0.175	0.230	0.303	0.492	1.515
	$(1/\hat{\omega}_{AR})_{BC}$	0.003	-0.008	-0.014	-0.018	-0.027	0.126	0.164	0.204	0.297	0.700
$T = 200$	$1/\hat{\omega}_{kernel}$	-0.146	0.000	0.116	0.277	0.707	0.035	0.042	0.100	0.269	1.223
	$1/\hat{\omega}_{AR}$	0.075	0.089	0.111	0.156	0.301	0.059	0.072	0.095	0.149	0.374
	$(1/\hat{\omega}_{AR})_{BC}$	0.002	0.003	0.002	0.001	-0.003	0.049	0.058	0.074	0.108	0.222
		$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$
		$\phi_2 = 0.3$					$\phi_2 = 0.3$				
$T = 100$	$1/\hat{\omega}_{kernel}$	0.470	0.784	1.201	1.356	2.309	0.345	0.766	1.978	3.260	10.586
	$1/\hat{\omega}_{AR}$	0.455	0.506	0.898	1.002	1.701	0.667	0.718	1.700	3.170	7.910
	$(1/\hat{\omega}_{AR})_{BC}$	0.255	0.302	0.657	0.476	0.403	0.556	0.580	1.501	2.490	4.027
$T = 200$	$1/\hat{\omega}_{kernel}$	0.335	0.637	0.954	0.827	1.193	0.167	0.461	1.236	1.123	2.637
	$1/\hat{\omega}_{AR}$	0.151	0.187	0.306	0.312	0.597	0.154	0.183	0.463	0.432	1.156
	$(1/\hat{\omega}_{AR})_{BC}$	0.035	0.054	0.138	0.037	0.019	0.130	0.152	0.416	0.299	0.590

Table 4.3: Bias and MSE of the estimators of $1/\omega$ with MA(1) errors: $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

		Bias					MSE				
		$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$
$T = 100$	$1/\hat{\omega}_{kernel}$	-0.899	-0.316	0.077	0.235	0.317	0.808	0.126	0.057	0.171	0.294
	$1/\hat{\omega}_{AR}$	-0.724	-0.166	0.067	0.214	0.518	0.549	0.168	0.071	0.369	0.894
	$(1/\hat{\omega}_{AR})_{BC}$	-0.754	-0.233	0.021	0.036	0.155	0.587	0.161	0.055	0.287	0.536
$T = 200$	$1/\hat{\omega}_{kernel}$	-0.874	-0.276	0.038	0.138	0.175	0.765	0.092	0.024	0.080	0.124
	$1/\hat{\omega}_{AR}$	-0.681	-0.182	0.028	0.079	0.232	0.476	0.093	0.019	0.124	0.297
	$(1/\hat{\omega}_{AR})_{BC}$	-0.701	-0.219	0.006	-0.017	0.034	0.503	0.099	0.017	0.103	0.219

Table 4.4: Empirical size of the tests with AR(1) errors: $u_t = \phi u_{t-1} + \varepsilon_t$

	$T = 100$					$T = 200$				
	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$
sup-W	0.075	0.107	0.145	0.207	0.335	0.065	0.088	0.111	0.143	0.227
sup-W_{AR}	0.072	0.140	0.128	0.132	0.214	0.064	0.105	0.079	0.085	0.125
sup-W_{BC}	0.061	0.126	0.101	0.078	0.102	0.058	0.096	0.066	0.062	0.069
sup-W_{kej}	0.064	0.077	0.069	0.060	0.045	0.060	0.069	0.064	0.058	0.043
fixed-b sup-W	0.014	0.019	0.026	0.027	0.036	0.032	0.029	0.028	0.031	0.037
CUSUM$_{H_1}$	0.067	0.092	0.132	0.188	0.312	0.055	0.077	0.093	0.123	0.195
CUSUM$_{H_1,AR}$	0.064	0.127	0.115	0.123	0.192	0.054	0.087	0.064	0.074	0.112
CUSUM$_{H_1,BC}$	0.057	0.114	0.086	0.073	0.094	0.048	0.083	0.053	0.053	0.063
SN	0.057	0.065	0.072	0.090	0.141	0.050	0.055	0.060	0.067	0.088

Table 4.5: Empirical size of the tests with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t,\ \phi_2 = -0.3$

	$T = 100$					$T = 200$				
	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = 0.7$	$\phi_1 = 0.9$	$\phi_1 = 1.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = 0.7$	$\phi_1 = 0.9$	$\phi_1 = 1.1$
sup-W	0.032	0.077	0.118	0.200	0.346	0.034	0.071	0.094	0.136	0.224
sup-W_{AR}	0.114	0.118	0.127	0.147	0.206	0.086	0.087	0.086	0.096	0.124
sup-W_{BC}	0.089	0.085	0.080	0.085	0.098	0.067	0.065	0.064	0.066	0.068
sup-W_{kej}	0.011	0.027	0.029	0.025	0.012	0.018	0.036	0.037	0.032	0.015
fixed-b sup-W	0.001	0.003	0.009	0.012	0.017	0.015	0.010	0.008	0.009	0.016
CUSUM$_{H_1}$	0.028	0.067	0.111	0.171	0.319	0.031	0.059	0.079	0.121	0.195
CUSUM$_{H_1,AR}$	0.104	0.104	0.109	0.129	0.181	0.073	0.073	0.076	0.081	0.107
CUSUM$_{H_1,BC}$	0.076	0.075	0.070	0.073	0.095	0.059	0.060	0.059	0.060	0.060
SN	0.041	0.052	0.059	0.070	0.103	0.043	0.045	0.052	0.059	0.072

Table 4.6: Empirical size of the tests with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t,\ \phi_2 = 0.3$

	$T = 100$					$T = 200$				
	$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$
sup-W	0.161	0.221	0.308	0.328	0.403	0.139	0.205	0.277	0.245	0.292
sup-W_{AR}	0.182	0.176	0.267	0.276	0.335	0.106	0.108	0.139	0.124	0.174
sup-W_{BC}	0.155	0.148	0.246	0.212	0.204	0.085	0.085	0.111	0.082	0.097
sup-W_{kej}	0.117	0.177	0.272	0.202	0.163	0.106	0.180	0.251	0.162	0.125
fixed-b sup-W	0.064	0.070	0.078	0.078	0.064	0.058	0.057	0.059	0.065	0.074
CUSUM$_{H_1}$	0.148	0.195	0.282	0.293	0.374	0.119	0.180	0.248	0.221	0.261
CUSUM$_{H_1,AR}$	0.161	0.161	0.248	0.252	0.309	0.088	0.092	0.119	0.113	0.163
CUSUM$_{H_1,BC}$	0.135	0.134	0.227	0.192	0.192	0.071	0.069	0.102	0.075	0.089
SN	0.059	0.073	0.088	0.114	0.172	0.052	0.060	0.064	0.079	0.103

Table 4.7: Empirical size of the tests with MA(1) errors: $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

	$T = 100$					$T = 200$				
	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$
sup-W	0.000	0.025	0.075	0.106	0.129	0.000	0.024	0.065	0.086	0.099
sup-W_{AR}	0.059	0.076	0.072	0.126	0.212	0.035	0.055	0.064	0.091	0.139
sup-W_{BC}	0.043	0.064	0.061	0.093	0.140	0.024	0.046	0.058	0.068	0.098
sup-W_{kej}	0.000	0.014	0.064	0.056	0.045	0.000	0.016	0.060	0.055	0.045
fixed-b sup-W	0.000	0.002	0.014	0.012	0.014	0.001	0.027	0.032	0.017	0.015
CUSUM$_{H_1}$	0.000	0.023	0.067	0.093	0.119	0.000	0.020	0.055	0.075	0.084
CUSUM$_{H_1,AR}$	0.044	0.067	0.064	0.108	0.190	0.021	0.047	0.054	0.079	0.120
CUSUM$_{H_1,BC}$	0.028	0.057	0.057	0.083	0.127	0.016	0.041	0.048	0.060	0.081
SN	0.000	0.026	0.057	0.065	0.067	0.000	0.033	0.050	0.055	0.057

Table 4.8: Empirical size of the tests with ARMA(3,3) errors: $u_t = \phi^3 u_{t-3} + \varepsilon_t + \theta\varepsilon_{t-3},\ \theta = -0.3$

	$T = 100$					$T = 200$				
	$\phi = 0$ (ϕ^3=0)	$\phi = 0.2$ (ϕ^3=0.008)	$\phi = 0.4$ (ϕ^3=0.064)	$\phi = 0.6$ (ϕ^3=0.216)	$\phi = 0.8$ (ϕ^3=0.512)	$\phi = 0$ (ϕ^3=0)	$\phi = 0.2$ (ϕ^3=0.008)	$\phi = 0.4$ (ϕ^3=0.064)	$\phi = 0.6$ (ϕ^3=0.216)	$\phi = 0.8$ (ϕ^3=0.512)
sup-W	0.005	0.005	0.008	0.038	0.256	0.002	0.002	0.006	0.027	0.278
sup-W_{AR}	0.054	0.053	0.040	0.043	0.231	0.044	0.044	0.035	0.025	0.220
sup-W_{BC}	0.037	0.037	0.030	0.035	0.210	0.034	0.035	0.027	0.022	0.204
sup-W_{kej}	0.003	0.004	0.005	0.029	0.227	0.001	0.002	0.003	0.022	0.261
fixed-b sup-W	0.000	0.000	0.001	0.005	0.085	0.003	0.003	0.004	0.012	0.100
CUSUM$_{H_1}$	0.004	0.004	0.008	0.031	0.230	0.001	0.002	0.004	0.018	0.243
CUSUM$_{H_1,AR}$	0.043	0.043	0.034	0.035	0.211	0.036	0.036	0.031	0.022	0.195
CUSUM$_{H_1,BC}$	0.034	0.032	0.026	0.026	0.190	0.027	0.027	0.022	0.020	0.180
SN	0.014	0.014	0.019	0.034	0.091	0.031	0.032	0.034	0.042	0.070

Table 4.9: Empirical size of the tests with ARMA(3,3) errors: $u_t = \phi^3 u_{t-3} + \varepsilon_t + \theta\varepsilon_{t-3},\ \theta = 0.3$

	$T = 100$					$T = 200$				
	$\phi = 0$ (ϕ^3=0)	$\phi = 0.2$ (ϕ^3=0.008)	$\phi = 0.4$ (ϕ^3=0.064)	$\phi = 0.6$ (ϕ^3=0.216)	$\phi = 0.8$ (ϕ^3=0.512)	$\phi = 0$ (ϕ^3=0)	$\phi = 0.2$ (ϕ^3=0.008)	$\phi = 0.4$ (ϕ^3=0.064)	$\phi = 0.6$ (ϕ^3=0.216)	$\phi = 0.8$ (ϕ^3=0.512)
sup-W	0.205	0.213	0.243	0.354	0.605	0.200	0.204	0.242	0.363	0.646
sup-W_{AR}	0.183	0.183	0.187	0.172	0.181	0.134	0.132	0.110	0.069	0.082
sup-W_{BC}	0.157	0.158	0.158	0.106	0.063	0.115	0.113	0.088	0.036	0.028
sup-W_{kej}	0.176	0.180	0.213	0.316	0.559	0.182	0.188	0.223	0.343	0.622
fixed-b sup-W	0.072	0.073	0.090	0.155	0.318	0.051	0.053	0.062	0.092	0.225
CUSUM$_{H_1}$	0.186	0.187	0.214	0.312	0.568	0.173	0.177	0.203	0.314	0.583
CUSUM$_{H_1,AR}$	0.165	0.167	0.172	0.157	0.168	0.115	0.113	0.092	0.063	0.070
CUSUM$_{H_1,BC}$	0.141	0.142	0.145	0.095	0.057	0.101	0.099	0.074	0.030	0.025
SN	0.065	0.065	0.070	0.085	0.141	0.055	0.055	0.057	0.064	0.084

Table 4.10: Empirical size of the bias-corrected tests with AR(1) errors: $u_t = \phi u_{t-1} + \varepsilon_t$

	$T = 100$					$T = 200$				
	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$
sup-W_{BC}	0.061	0.126	0.101	0.078	0.102	0.058	0.096	0.066	0.062	0.069
mean-W_{BC}	0.066	0.101	0.074	0.065	0.068	0.050	0.070	0.051	0.053	0.055
exp-W_{BC}	0.068	0.125	0.098	0.085	0.098	0.051	0.081	0.059	0.059	0.071
$LM_{H_1,BC}$	0.071	0.124	0.107	0.100	0.123	0.053	0.082	0.065	0.067	0.087
$qLL_{H_1,BC}$	0.067	0.167	0.099	0.070	0.088	0.054	0.102	0.058	0.054	0.064
CUSUM$_{H_1,BC}$	0.057	0.114	0.086	0.073	0.094	0.048	0.083	0.053	0.053	0.063

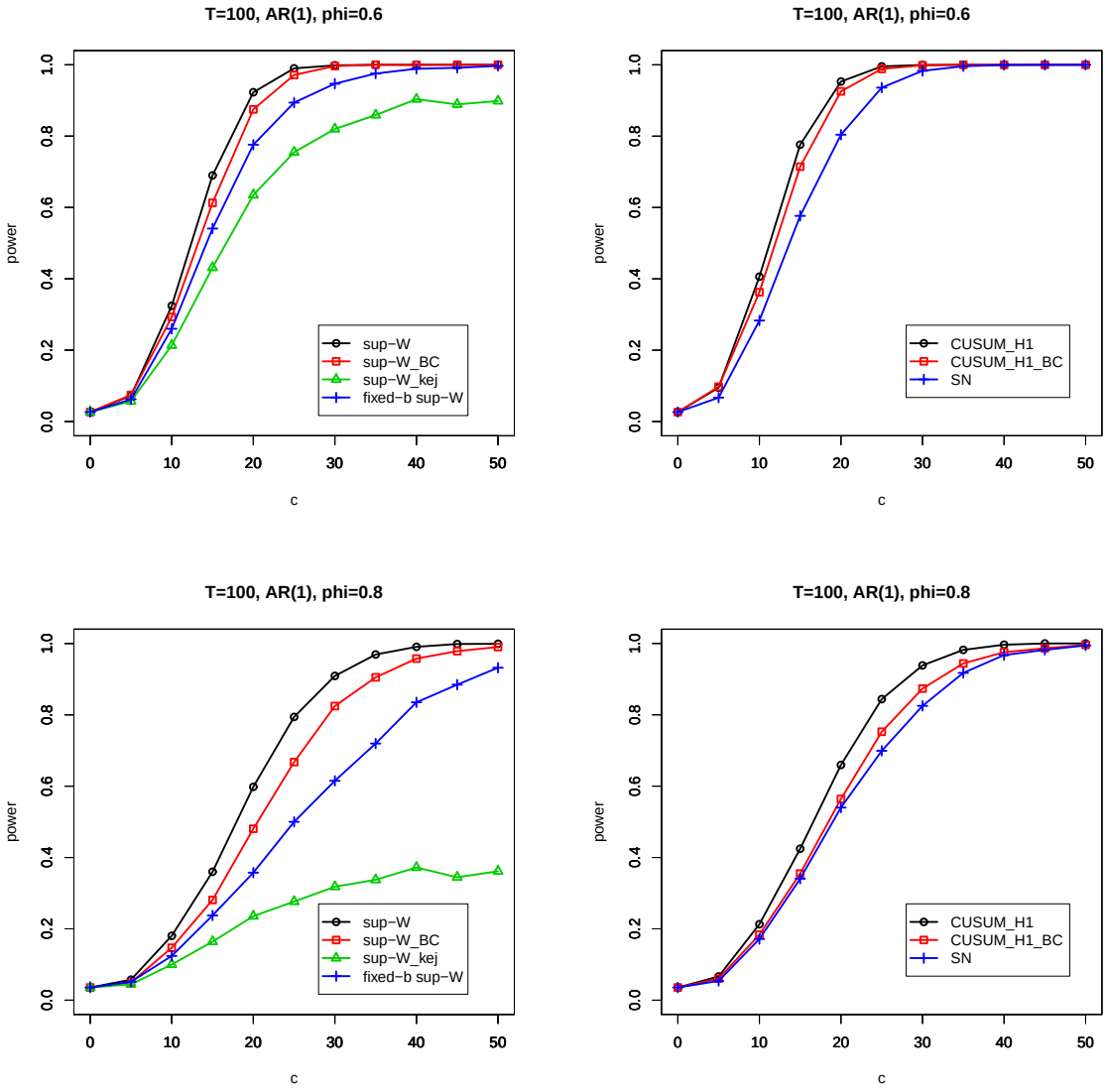

Figure 4.1: Size-adjusted power of the tests with AR(1) errors and $T = 100$

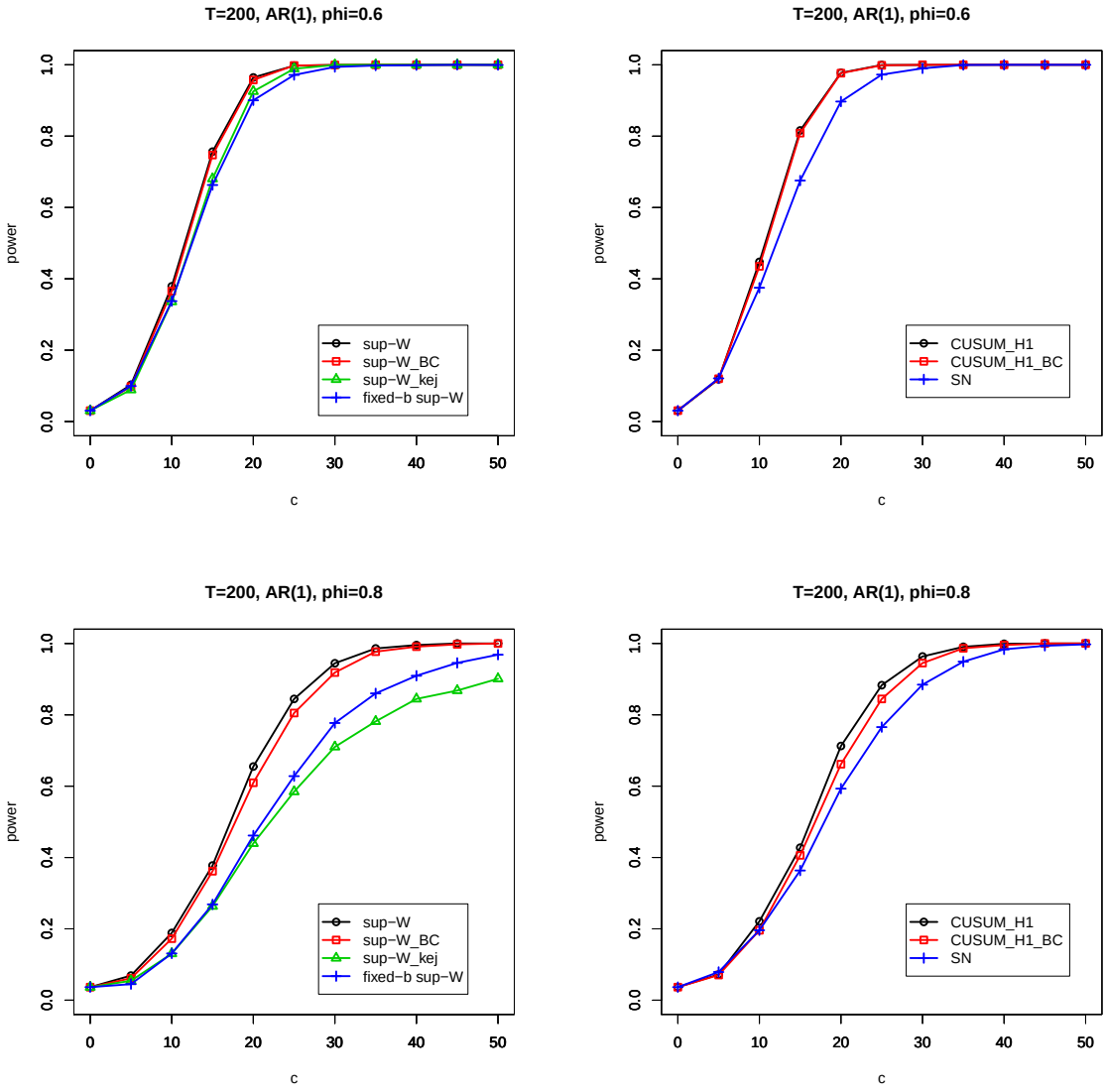

Figure 4.2: Size-adjusted power of the tests with AR(1) errors and $T = 200$

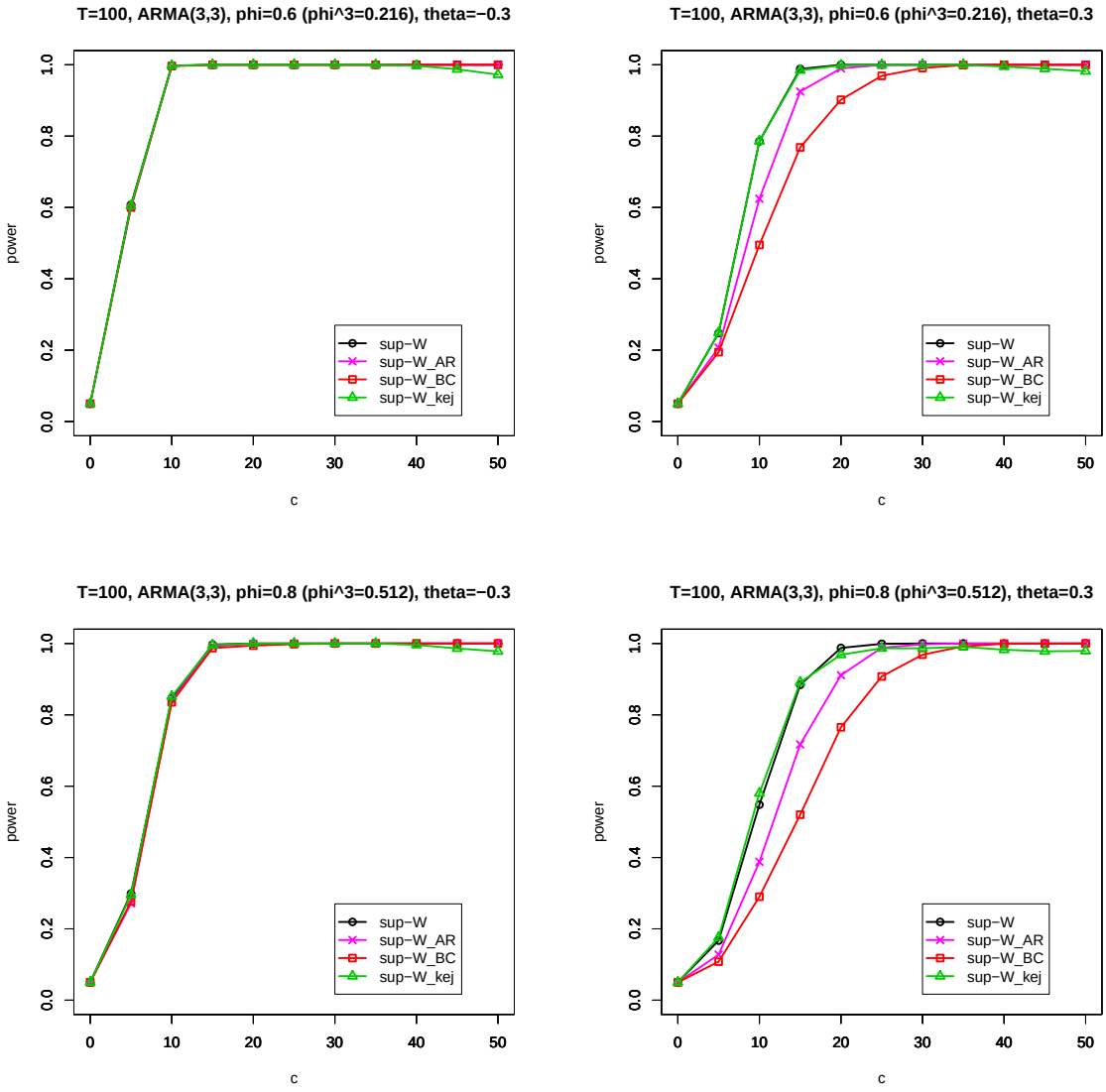

Figure 4.3: Size-adjusted power of the sup-Wald tests with ARMA(3,3) errors and $T = 100$

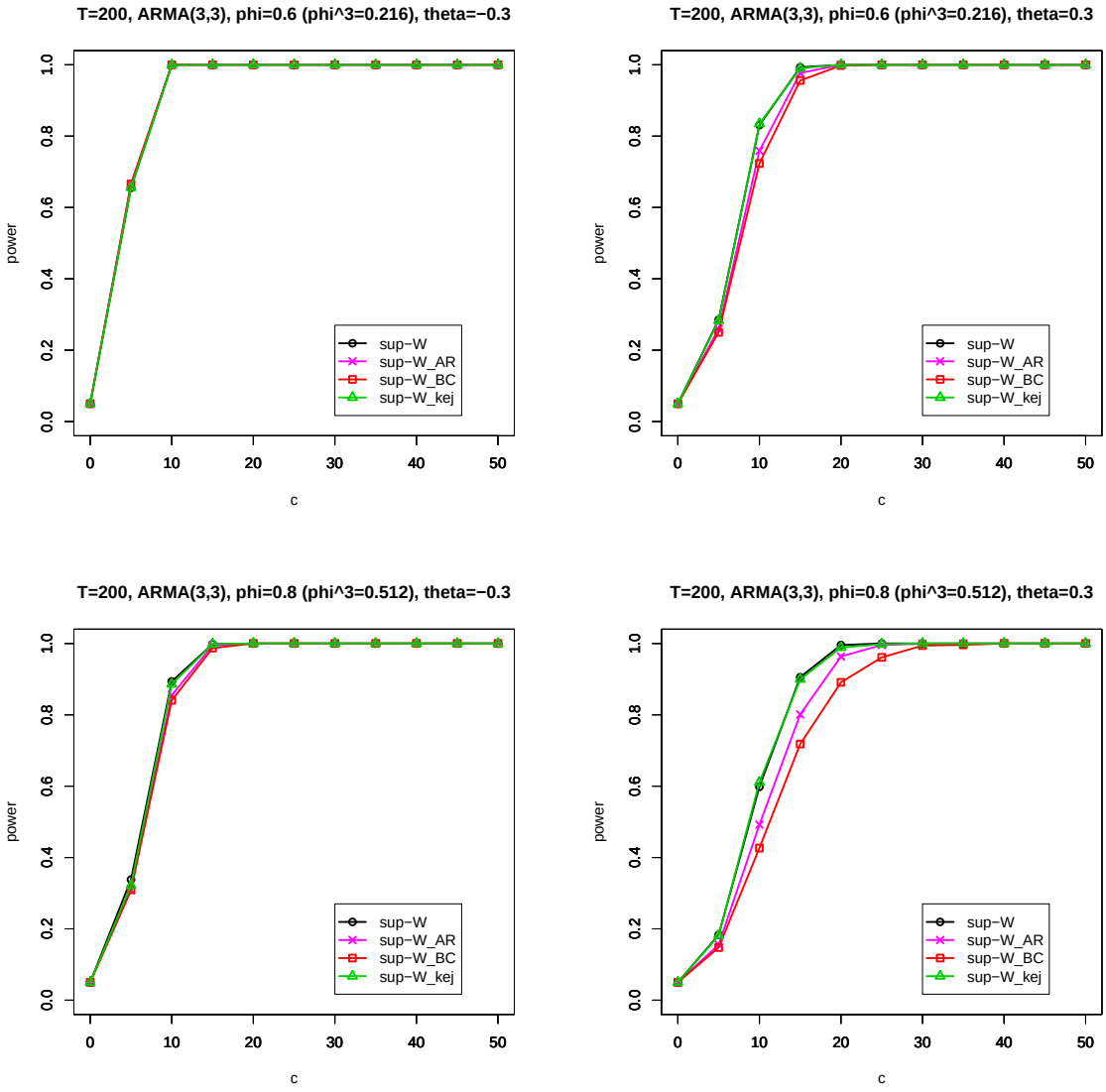

Figure 4.4: Size-adjusted power of the sup-Wald tests with ARMA(3,3) errors and $T = 200$

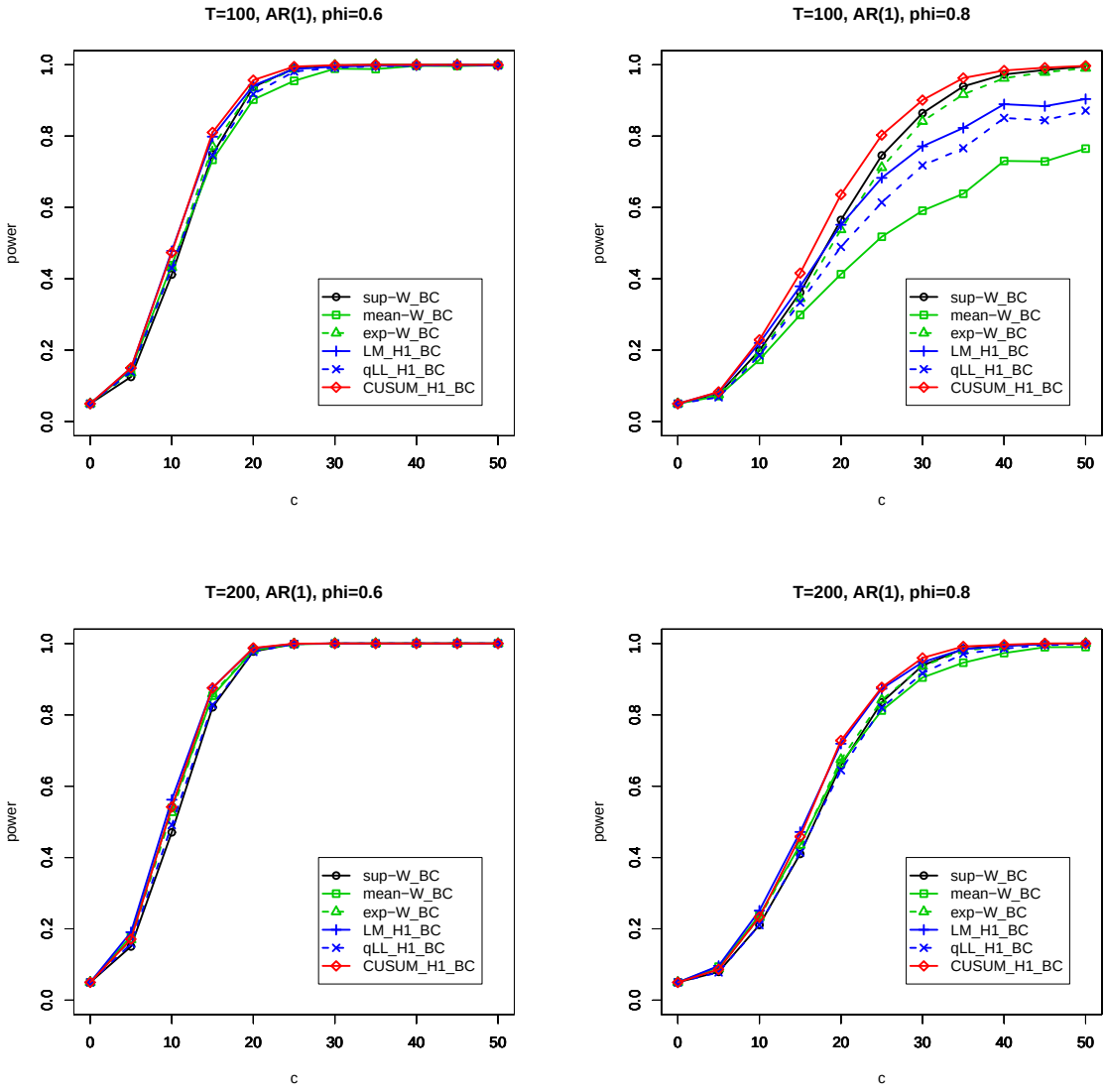

Figure 4.5: Size-adjusted power of the bias-corrected tests with AR(1) errors

Bibliography

[1] Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

[2] Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, 821-856.

[3] Andrews, D. W. K. and J-Y. Kim (2006), "Tests for Cointegration Breakdown over a Short Time Period," *Journal of Business and Economic Statistics* 24, 379-94.

[4] Andrews, D. W. K., I. Lee and W. Ploberger (1996), "Optimal Changepoint Tests for Normal Linear Regression," *Journal of Econometrics* 70, 9-38.

[5] Andrews, D. W. K. and W. Ploberger (1994), "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica* 62, 1383-1414.

[6] Aue, A. and L. Horváth (2013), "Structural Breaks in Time Series," *Journal of Time Series Analysis* 34, 1-16.

[7] Bai, J. (1997), "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics* 79, 551-563.

[8] Bai, J. (1999), "Likelihood Ratio Tests for Multiple Structural Changes," *Journal of Econometrics* 91, 299-323.

[9] 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.

[10] Bai, J. (2010), "Common Breaks in Means and Variances for Panel Data," *Journal of Econometrics* 157, 78-92.

[11] Bai, J., R. L. Lumsdaine and J. H. Stock (1998), "Testing For and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies* 65, 395-432.

[12] Bai, J. and P. Perron (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica* 66, 47-78.

[13] Berk, K. N. (1974), "Consistent Autoregressive Spectral Estimates," *Annals of Statistics* 2, 489-502.

[14] Brown, R. L., J. Durbin and J. M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society Series B* 37, 149-192.

[15] Canova, F. and Hansen, B. E. (1995), "Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability," *Journal of Business and Economic Statistics* 13, 237-252.

[16] Chan, J., L. Horváth and M. Hušková (2013), "Darling-Erdös Limit Results for Change-Point Detection in Panel Data," *Journal of Statistical Planning and Inference* 143, 955-970.

[17] Cox, D. R. and D. V. Hinkley (1974), *Theoretical Statistics*, Chapman and Hall, London.

[18] Crainiceanu, C. M. and T. J. Vogelsang (2007), "Nonmonotonic Power for Tests for a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation* 77, 457-476.

[19] Csörgő, M. and L. Horváth (1997), *Limit Theorems in Change-Point Analysis.* Wiley, New York.

[20] Deng, A. and P. Perron (2008), "A Non-Local Perspective on the Power Properties of the CUSUM and CUSUM of Squares Tests for Structural Change," *Journal of Econometrics* 142, 212-240.

[21] den Haan, W. J. and A. T. Levin (1997), "A Practitioner's Guide to Robust Covariance Matrix Estimation," in G. S. Maddala and C. R. Rao, ed., *Handbook of Econometrics, Vol.15: Robust Inference*. Elsevier, Amsterdam.

[22] den Haan, W. J. and A. T. Levin (1998), "Vector Autoregressive Covariance Matrix Estimation," working paper.

[23] Elliott, G. and U. K. Müller (2006), "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies* 73, 907-940.

[24] Hannan, E. J. and M. Deistler (1988), "The Statistical Theory of Linear Systems," John Wiley & Sons, New York.

[25] Hansen, B. E. (1992a), "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics* 10, 321-335.

[26] Hansen, B. E. (1992b), "Testing for Parameter Instability in Linear Models," *Journal of Policy Modeling* 14, 517-533.

[27] Horváth, L. and M. Hušková (2012), "Change-Point Detection in Panel Data," *Journal of Time Series Analysis* 33, 631-648.

[28] Juhl, T. and O. Lugovskyy (2014), "A Test for Slope Heterogeneity in Fixed Effects Models," *Econometric Reviews* 33, 906-935.

[29] Juhl, T. and Z. Xiao (2003), "Power Functions and Envelopes for Unit Root Tests," *Econometric Theory* 19, 240-253.

[30] Juhl, T. and Z. Xiao (2009), "Tests for Changing Mean with Monotonic Power," *Journal of Econometrics* 148, 14-24.

[31] Kejriwal M. (2009), "Tests for a Mean Shift with Good Size and Monotonic Power," *Economics Letters* 102, 78-82.

[32] Kejriwal M. and P. Perron (2010), "Testing for Multiple Structural Changes in Cointegrated Regression Models," *Journal of Business and Economic Statistics* 28, 503-522.

[33] Kiefer, N. M. and T. J. Vogelsang (2005), "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory* 21, 1130-1164.

[34] Kim, D. (2010), "Improved and Extended End-of-Sample Instability Tests Using a Feasible Quasi-generalized Least Squares Procedure," *Econometric Theory* 26, 994-1031.

[35] King, M. L. and G. H. Hillier (1985), "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model," *Journal of the Royal Statistical Society Series B* 47, 98-102.

[36] Kim, D. (2011), "Estimating a Common Deterministic Trend Break in Large Panels with Cross Sectional Dependence," *Journal of Econometrics* 164, 310-330.

[37] Kim, D. and P. Perron (2009), "Assessing the Relative Power of Structural Break Tests Using a Framework Based on the Approximate Bahadur Slope," *Journal of Econometrics* 149, 26-51.

[38] Kurozumi, E. (2003), "Some Properties of the Point Optimal Invariant Test for the Constancy of Parameters," *Journal of the Japan Statistical Society* 33, 169-180.

[39] Lin, C.-F. J. and T. Teräsvirta (1999), "Testing Parameter Constancy in Linear Models against Stochastic Stationary Parameters," *Journal of Econometrics* 90, 193-213.

[40] Mood, A. M., F. A. Graybill and D. C. Boes (1974), *Introduction to the Theory of Statistics*, Third Edition, McGraw-Hill, Singapore.

[41] Nabeya, S. and Tanaka, K. (1988), "Asymptotic Theory of a Test for the Constancy of Regression Coefficients against the Random Walk Alternative," *Annals of Statistics* 16, 218-235.

[42] Nyblom, J. (1986), "Testing for Deterministic Linear Trend in Time Series," *Journal of the American Statistical Association* 81, 545-549.

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

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

[45] Patterson, H. D. and R. Thompson (1971), "Recovery of Inter-Block Information when Block Sizes are Unequal," *Biometrika* 58, 545-554.

[46] Patterson, K. D. (2000), "Bias Reduction in Autoregressive Models," *Economics Letters* 68, 135-141.

[47] Perron, P. (2006), "Dealing with Structural Breaks," in T. C. Mills and K. Patterson, ed., *Palgrave Handbook of Econometrics, Vol.1: Econometric Theory*. Palgrave Macmillan, New York.

[48] Perron, P. and Y. Yamamoto (2014), "On the Usefulness or Lack Thereof of Optimality Criteria for Structural Change Tests," *Econometric Reviews*, forthcoming.

[49] Pesaran, M. H. and T. Yamagata (2008), "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics* 142, 50-93.

[50] Ploberger, W. and W. Krämer (1992), "The CUSUM Test with OLS Residuals," *Econometrica* 60, 271-285.

[51] Sayginsoy, Ö. and T. J. Vogelsang (2011), "Testing for a Shift in Trend at an Unknown Date: A Fixed-b Analysis of Heteroskedasticity Autocorrelation Robust OLS-Based Tests," *Econometric Theory* 27, 992-1025.

[52] Qu, Z. and P. Perron (2007), "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica* 75, 459-502.

[53] Shaman, P. and R. A. Stine (1988), "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association* 83, 842-848.

[54] Shao, X. and X. Zhang (2010), "Testing for Change Points in Time Series," *Journal of American Statistical Association* 105, 1228-1240.

[55] Shively, T. S. (1988), "An Analysis of Tests for Regression Coefficient Stability," *Journal of Econometrics* 39, 367-386.

[56] Sowell, B. (1996), "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework," *Econometrica* 64, 1085-1107.

[57] Stine, R. A. and P. Shaman (1989), "A Fixed Point Characterization for Bias of Autoregressive Estimators," *Annals of Statistics* 17, 1275-1284.

[58] Swamy, P. A. V. B. (1970), "Efficient Inference in a Random Coefficient Regression Model," *Econometrica* 38, 311-323.

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

[60] Velasco, C. and P. M. Robinson (2001), "Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean," *Econometric Theory* 17, 497-539.

[61] Vogelsang, T. J. (1999), "Sources of Nonmonotonic Power when Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics* 88, 283-299.

[62] Yamazaki, D. and E. Kurozumi (2015a), "Testing for Parameter Constancy in the Time Series Direction in Panel Data Models," *Journal of Statistical Computation and Simulation* 85, 2874-2902.

[63] Yamazaki, D. and E. Kurozumi (2015b), "Improving the Finite Sample Performance of Tests for a Shift in Mean," *Journal of Statistical Planning and Inference* 167, 144-173.

[64] Yang, J. and T. J. Vogelsang (2011), "Fixed-b Analysis of LM-Type Tests for a Shift in Mean," *Econometrics Journal* 14, 438-456.