

ROBUST INFERENCE ON INFINITE AND GROWING DIMENSIONAL TIME SERIES REGRESSION

ABHIMANYU GUPTA[✉] AND MYUNG HWAN SEO^{§†}

ABSTRACT. We develop a class of tests for time series models such as multiple regression with growing dimension, infinite-order autoregression and nonparametric sieve regression. Examples include the Chow test, Andrews and Ploberger (1994)-type exponential tests, and general linear restriction tests, all of growing rank p . Employing such increasing p asymptotics, we introduce a new scale correction to conventional test statistics. This accounts for a high-order long-run variance that emerges as p grows with sample size. We propose a bias correction via a null-imposed bootstrap to alleviate finite sample bias without sacrificing power unduly. A simulation study shows the importance of robustifying testing procedures against the high-order long-run variance even when p is moderate. The tests are illustrated with an application to the oil regressions in Hamilton (2003).

Keywords: Growing number of restrictions, High-order Long-run Variance (HLV), Nonparametric regression, Infinite-order autoregression.

Date: April 14, 2022 We thank the co-editor and three referees for many insightful comments that substantially improved the paper, as well as Donald Andrews, Xiaohong Chen, Jean-Jacques Forneron, Šárka Hudecová, Yuichi Kitamura, Joon Park, Pierre Perron, Zhongjun Qu and Yixiao Sun, audiences at Yale, Boston University, GOFCP 2019 (Trento), MEG 2019 (Columbus), LAMES 2019 (Puebla), Cambridge, ESWC 2020 and Binghamton. We are grateful to Joel Tropp for guiding us to useful results and Ekaterina Oparina for excellent research assistance.

[✉] Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: a.gupta@essex.ac.uk. Research supported by British Academy/Leverhulme Trust grant SRG\170956.

[§] Corresponding author. [†] Department of Economics, Seoul National University, Gwan-Ak Ro 1, Seoul, Korea. Email: myunghseo@snu.ac.kr. Research supported by the Ministry of Education of the Republic of Korea, National Research Foundation of Korea (NRF-2018S1A5A2A01033487) and LG Yonam Foundation, and partly written while visiting the Cowles Foundation at Yale University, whose hospitality is greatly acknowledged.

1. INTRODUCTION

This paper develops asymptotically valid tests for inference on infinite-order and growing dimensional time series regression models, revealing the presence of an hitherto undetected *nonlinear serial dependence* or *high-order long-run variance* (HLV) factor. This factor depends on the model error and regressors in a nonlinear fashion, and can appear in limit distributions when the data exhibit dependence and the number of restrictions grows. Chow tests, tests for linear restrictions and stability tests with unknown change point are all covered. Our theory, simulations and empirical results show the deleterious effect of ignoring the HLV term, and we propose a testing procedure that is robust to its presence. This is shown to possess desirable finite sample properties. While the HLV factor is revealed by our increasing dimension asymptotics, it can contaminate inference even in multiple regressions with a moderate number of covariates. Such specifications are ubiquitous in practice. Thus, the findings and recommendations of this paper are important for practitioners wishing to make correct inferences when data are dependent.

Models of infinite or growing dimension have been widely studied in the recent econometric literature, reflecting modern applications with rich sets of variables. For example, the asset pricing literature has suggested hundreds of potential risk factors to explain returns, see Feng et al. (2020). With a larger number of observations accumulating over time, it is natural to include more of these variables as covariates even without resorting to penalized estimation methods. In fact, an attitude that permits the number of covariates to grow as a function of sample size is tacitly adopted in the literature. In a survey, Koenker (1988) observed that the number of regressors in empirical work increases as the sample size n increases, roughly like $n^{1/4}$, suggesting that practitioners implicitly treat model complexity as a function of sample size. Finally, nonparametric methods such as series estimation have found wide applications in the economics and finance literature, see e.g. Jordà (2005), Chen (2007), Chen and Christensen (2015). These methods involve the approximation of an infinite-order model with a sequence of growing dimensional models. Taken together, this proliferation

of models highlights the importance of developing appropriate techniques for their study.

Our approach is to develop tests for null hypotheses that involve a growing number of restrictions p in time series regression, with p increasing slower than sample size. As a leading example, we consider the Chow test, due to Chow (1960), to test for a structural break at a prescribed time. This has the advantage of being a simple exclusion restriction test with wide applicability. After examining the key issues in this simple context, we present extensions to more general structural break testing with an unknown break date and testing of general linear restrictions. This extends specification tests with slowly growing p , see e.g. Hong and White (1995) and Gupta (2018), to time series regression. However, our testing problem is distinct from the so-called many restrictions setting in e.g. Calhoun (2011), Anatolyev (2012, 2019), Kline et al. (2020), amongst others, where the number of restrictions grows proportionally to sample size.

We derive the asymptotic distribution of the Chow test Wald statistic centered by p and normalized by $\sqrt{2p}$. This yields asymptotic normality with an unknown asymptotic variance \mathcal{V} , which we term the HLV, provided that p meets certain growth conditions. The HLV factor \mathcal{V} captures high-order autocovariances of the regressors and disturbances, echoing the long-run variance that appears in fixed dimensional time series regression, and vanishes under simplifying assumptions that remove these high-order autocovariances. The new HLV factor \mathcal{V} does not appear in fixed p asymptotic regimes, nor does it appear in the independent data setting of Hong and White (1995), who use the same transformation and obtain asymptotic standard normality.

We robustify the Chow test against the HLV in the spirit of heteroskedasticity autocorrelation robust (HAR) inference, see e.g. Kiefer and Vogelsang (2002), Sun (2014), and Lazarus et al. (2018), just to name a few. The resulting asymptotic distribution is mixed normal and pivotal. However, unlike conventional HAR inference, where the standard Wiener process characterizes mixed normality, our limit distribution is represented by two dependent centered Gaussian processes $W(r)$ and $\bar{W}(r)$ such that

$EW(r)^2 = r^2$ and $E\bar{W}(r)^2 = (1 - r)^2$. Although pivotal, the asymptotic distribution depends on the location of the hypothesized break date and thus we provide R code to compute the p-values.

The finite sample bias in the Wald statistic, or in quadratic statistics more generally, is a serious issue when p is large. See e.g. Kline et al. (2020) for more discussion and a bias correction proposal that works well even when p is proportional to the sample size but under independent sampling. Our simulations document that the problem is even worse in time series regression. Thus, we propose a bootstrap bias correction which imposes the null hypothesis in the resampling so as not to sacrifice power unduly. Even a small number of bootstrap iterations appear sufficient to reduce the bias, making computation easily manageable. Based on these findings, we recommend a bias-corrected and HLV-robust test to practitioners.

Our findings apply to more general settings. We present two extensions: structural break testing with an unknown break date and the testing of general linear restrictions of rank p . The HLV appears in both. In break testing, we index the statistic with an unidentified parameter under the null hypothesis and establish the weak convergence of this stochastic process as both n and p grow to infinity simultaneously. We then employ an exponential averaging of the process that is suitably modified from Andrews and Ploberger (1994) to account for the growing dimension p . Thus, we also contribute to the hitherto rather small nonparametric structural stability testing literature (see e.g. Wu and Chu (1993), Chen and Hong (2012) and Mohr and Neumeyer (2020)).

In simulations for a range of settings across regression with many covariates, long AR fits and sieve regression, we demonstrate that our statistic exhibits excellent size control without sacrificing power excessively. Failure to correct for the HLV can seriously affect inference, in general leading to over-rejection and often severely so. Such a pattern is shown to persist for the three types of tests that we provide: Chow, exclusion restrictions and structural break testing. In an empirical example based on Hamilton (2003, 2009), we show that using our bias-corrected and HLV-robust

tests can yield inferences that lead to new conclusions when considering the relation between oil prices and economic activity.

The paper is organized as follows. Section 2 introduces the model and the Chow test, along with some basic assumptions and examples. In Section 3 we provide an asymptotic theory and verify our high-level conditions for several examples. Section 4 presents the correction for the HLV discussed above by means of the HAR inference literature. Extensions to general structural break testing with an unknown break date and to the testing of general linear restrictions are given in Section 5. Section 6 contains a Monte Carlo study of finite sample performance, and Section 7 demonstrates our test with real data. All the proofs of theorems and lemmas are collected in two further appendices, the second of which is available online. Throughout the paper, cross-referenced items prefixed with ‘S’ can be found in this online supplementary appendix. An R-package to replicate the simulations and empirical example is available on the authors’ websites.¹

2. STRUCTURAL BREAK IN INFINITE-ORDER REGRESSION

We consider the issue of testing for a structural break at a known point in the conditional mean function of y_t given the information available up to $t-1$, ie $E(y_t|\mathcal{F}_{t-1})$, where \mathcal{F}_{t-1} denotes the filtration up to time $t-1$. In nonparametric regression, \mathcal{F}_{t-1} typically consists of a finite number of observable covariates z_t . In the context of the infinite order autoregressive AR(∞) model, \mathcal{F}_{t-1} is the collection of all the lagged dependent variables, $\{y_{t-j}\}_{j \geq 1}$. Alternatively, it can be viewed as a genuine high-dimensional regression model which may contain an infinite number of covariates and their lags. We allow for array structure but we do not introduce further notation to denote it unless necessary.

Given a sample size n , we estimate the unknown regression function via a growing-dimensional (or truncated) linear regression

$$(2.1) \quad y_t = x'_{nt} \beta_n + e_{nt},$$

¹sites.google.com/site/myunghseo/research, sites.google.com/site/abhimanyugupta85/research.

where x_{nt} and β_n are p -dimensional vectors and $p \rightarrow \infty$ as $n \rightarrow \infty$ to estimate $E(y_t|\mathcal{F}_{t-1})$ consistently. To be more precise, let

$$\varepsilon_t = y_t - E(y_t|\mathcal{F}_{t-1}),$$

β_n be the best linear predictor of y_t given x_{nt} , and $r_{nt} = E(y_t|\mathcal{F}_{t-1}) - x'_{nt}\beta_n$. Then, $e_{nt} = r_{nt} + \varepsilon_t$. Throughout the paper, let C (c) denote a generic finite constant, arbitrarily large (small) but independent of n , and ‘a.s.’ stand for ‘almost surely’. Introduce the following assumptions:

Assumption 1. *The martingale difference sequence $\{\varepsilon_t\}$ satisfies $\sigma_t^2 \leq C$, where $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$, and $E(\varepsilon_t^4|\mathcal{F}_{t-1}) \leq C$, a.s.*

The theory presented in the paper may not hold if in fact we only have $E(x_{nt}\varepsilon_t) = 0$ as the long run variance of $x_{nt}\varepsilon_t$ will then appear in the type of quadratic statistics that we consider.

Assumption 2. *For $a = 1, 2$,*

$$(2.2) \quad \sup_t E(r_{nt}^{2a}) = o(n^{-1}).$$

We discuss this assumption on the negligibility of the approximation error in more detail in Section 6, where specific examples are introduced. The subscript n will now usually be dropped, although we will emphasize this occasionally to remind the reader of the n -dependence of certain quantities.

Introduce a potential structural break for these models at a given time, say $t = [n\gamma]$, $\gamma \in \Gamma \subset (0, 1)$, Γ compact and $[\cdot]$ denoting the integer part of the argument. That is, $\beta = \beta_1$ if $t/n \leq \gamma$ and $\beta = \beta_2$ if $t/n > \gamma$. We write the model as

$$(2.3) \quad y_t = x'_t\beta_1 1\{t/n \leq \gamma\} + x'_t\beta_2 1\{t/n > \gamma\} + e_t = x'_t\delta_1 + x'_t\delta_2 1\{t/n > \gamma\} + e_t,$$

where $\delta_1 = \beta_1$, $\delta_2 = \beta_2 - \beta_1$, and $1\{\cdot\}$ denotes the indicator function. Consider the Wald test for the exclusion restriction $\delta_2 = 0$, namely the Chow test for the presence

of a structural break at a known date. We will examine the case where the break date is unknown *a priori* in Section 5.

Let $\hat{\delta}(\gamma)$ and $\hat{e}_t(\gamma)$ denote the OLS estimate and the OLS residuals, respectively, and $x_t(\gamma) := (x'_t, x'_t 1\{t/n > \gamma\})'$. Also, let $\hat{M}(\gamma) = n^{-1} \sum_{t=1}^n x_t(\gamma) x_t(\gamma)'$, and $\hat{\Omega}(\gamma)$ denote an estimator of $E \varepsilon_t^2 x_t(\gamma) x_t(\gamma)'$. For instance, $\hat{\Omega}(\gamma)$ can be set as $n^{-1} \sum_{t=1}^n x_t(\gamma) x_t(\gamma)' \hat{e}_t(\gamma)^2$ (the Eicker-White formula) or, assuming conditional homoskedasticity, it can be $\hat{\sigma}(\gamma)^2 \hat{M}(\gamma)$, where $\hat{\sigma}^2(\gamma) = n^{-1} \sum_{t=1}^n \hat{e}_t(\gamma)^2$. The choice depends on the case being considered. Then, the Wald statistic for the familiar Chow test is defined as

$$(2.4) \quad W_n(\gamma) := n \hat{\delta}_2(\gamma)' \left(R \hat{M}(\gamma)^{-1} \hat{\Omega}(\gamma) \hat{M}(\gamma)^{-1} R' \right)^{-1} \hat{\delta}_2(\gamma),$$

where $R = (0_{p \times p} : I_p)$ is a selection matrix.

When the dimension p of x_t grows with the sample size n , the Wald statistic diverges as it is approximately chi-squared distributed with degree of freedom p . Thus, a conventional approach, as used e.g. by de Jong and Bierens (1994) and Hong and White (1995) in the cross-sectional (independent data) framework is to introduce a new centering and scaling to define

$$(2.5) \quad \mathcal{Q}_n(\gamma) := (W_n(\gamma) - p) / \sqrt{2p},$$

since the mean and variance of a chi-square distribution with p degrees of freedom are p and $2p$, respectively. Furthermore, it has been established that the standard normal approximation of \mathcal{Q}_n is valid in their settings. Subsequent sections investigate how this conventional approach fails in the context of growing or infinite dimensional time series models, mirroring the failure of time series inference procedures without heteroskedasticity and autocorrelation correction or robustification.

3. ASYMPTOTIC DISTRIBUTION OF \mathcal{Q}_n

This section provides the asymptotic distributions of the test statistic under the null and also shows that the statistic has non-trivial power against local alternatives at an appropriate nonparametric rate.

There has been some recent interest in the so-called many regressor setting where p is allowed to be proportional to n , see e.g. Cattaneo et al. (2018) and Kline et al. (2020). We do not permit such a large p as our hypothesis of interest concerns a p -dimensional restriction and the design matrix of time series data faces more difficulties in satisfying the rank condition. In this regard, Chen and Lockhart (2001) provide an interesting example from an ANOVA design where the weak convergence of the empirical distribution of residuals from the linear regression with growing dimension fails when the dimension p is of order $n^{1/3}$. They compare various growth conditions for p in the literature and conclude that $p^3 \log^2 p = o(n)$ is nearly necessary for a general stochastic design. Heuristically, a hypothesis represented through the empirical distribution function imposes an infinite number of restrictions, like our structural break testing also does, and valid testing of such a hypothesis demands a tighter control on the growth rate of p .

3.1. Asymptotic Null Distribution. Define $\|A\| = \{\bar{\lambda}(A'A)\}^{\frac{1}{2}}$ for a generic matrix A , where $\underline{\lambda}$ (respectively $\bar{\lambda}$) denotes the smallest (largest) eigenvalue of a symmetric nonnegative definite matrix. Any limit stated as ' $n \rightarrow \infty$ ' is taken as both n and p grow to infinity simultaneously unless specified otherwise. We also introduce the $p \times p$ non-stochastic matrix sequences M and Ω and define

$$M(\gamma) = \begin{bmatrix} M & (1-\gamma)M \\ (1-\gamma)M & (1-\gamma)M \end{bmatrix}, \quad \Omega(\gamma) = \begin{bmatrix} \Omega & (1-\gamma)\Omega \\ (1-\gamma)\Omega & (1-\gamma)\Omega \end{bmatrix}.$$

Assumption 3. (i) $\sup_{i,t} Ex_{ti}^4 < \infty$.

(ii)

$$\begin{aligned} \sup_{\gamma \in \Gamma \cup \{1\}} \left\| n^{-1} \sum_{t=1}^{[n\gamma]} x_t x'_t - \gamma M \right\| + \sup_{\gamma \in \Gamma \cup \{1\}} \left\| n^{-1} \sum_{t=1}^{[n\gamma]} x_t x'_t \sigma_t^2 - \gamma \Omega \right\| &= O_p(\varkappa_p), \\ \sup_{\gamma \in \Gamma} \left\| \hat{\Omega}(\gamma) - \Omega(\gamma) \right\| &= O_p(v_p), \\ \underline{\lambda}(M) > \lambda_n, \underline{\lambda}(\Omega) &> \lambda_n, \end{aligned}$$

for some positive sequences of numbers \varkappa_p , v_p and λ_n satisfying

$$(3.1) \quad \lambda_n^{-4} \sqrt{p} (\lambda_n^{-1} \varkappa_p + v_p) \rightarrow 0 \text{ and } \lambda_n^4 p \rightarrow \infty.$$

(iii) $\overline{\lim}_{n \rightarrow \infty} \bar{\lambda}(M) < \infty$, $\overline{\lim}_{n \rightarrow \infty} \bar{\lambda}(\Omega) < \infty$.

Several factors determine the bound \varkappa_p for nonparametric series regression. It is proportional to $\sqrt{p/n}$ or p/\sqrt{n} up to logarithmic factors with iid data, depending on the choice of basis functions. For dependent data, the mixing decay rate also contributes to \varkappa_p . The exact rate v_p depends on a particular example. We formally introduce our examples of multiple linear regression, $\text{AR}(\infty)$ and nonparametric sieve regression in Section 6. Primitive conditions and expressions for \varkappa_p and v_p are given in Propositions SP.1 and SP.2 in Section S.D of the online supplementary appendix, using the results of Peligrad (1982), Newey (1997), Gonçalves and Kilian (2007) and Chen and Christensen (2015).

Recall that the eigenvalues of the Kronecker product of two symmetric matrices are the products of their eigenvalues, and γ is bounded away from zero and one. Thus, $M(\gamma)$ and $\Omega(\gamma)$ inherit the eigenvalue restrictions on M and Ω in Assumption 3 (ii) and (iii) uniformly in γ , up to positive constants.

To develop the distributional limit of $\mathcal{Q}_n(\gamma)$ where both n and p diverge simultaneously, we introduce more conditions. Now, for convenience we denote $\xi_t = \Omega^{-1/2} x_t \varepsilon_t$, $\Upsilon_t = E(\xi_t \xi'_t | \mathcal{F}_{t-1})$, and $\Xi_s = \sum_{t_1=1}^{s-1} \sum_{t_2=1}^{s-1} \xi_{t_1} \xi'_{t_2}$. The next assumption introduces the HLV factor \mathcal{V} formally.

Assumption 4. Suppose that $\overline{\lim}_{n \rightarrow \infty} \bar{\lambda}(\Upsilon_t) < C$ a.s., $\max_{1 \leq t \leq n} E((\xi'_t \xi_t)^2 | \mathcal{F}_{t-1}) = O(p^2)$ a.s., $\max_{1 \leq s, t \leq n, t \neq s} E(\xi'_s \xi_t)^4 = O(p^2)$, $\sum_{t=1}^n \sum_{s=1}^{t-1} \text{cov}(\text{tr}(\Upsilon_t \Xi_t), \text{tr}(\Upsilon_s \Xi_s)) = o(n^4 p^2)$, and there exists \mathcal{V} such that for m that is proportional to n ,

$$(3.2) \quad \lim_{m, p \rightarrow \infty} \left(\frac{1}{mp} \text{tr} \sum_{t_1=1}^m \sum_{t_2=1}^{[mr]} E \left[\Omega^{-1} x_m x'_m \varepsilon_m^2 \Omega^{-1} x_{t_1} x'_{t_2} \varepsilon_{t_1} \varepsilon_{t_2} \right] \right) = r\mathcal{V}, \text{ uniformly in } r \in (0, 1].$$

Note that under the special case where $\{x_t \varepsilon_t\}$ is an iid sequence, we have $\mathcal{V} = \lim_{m, p \rightarrow \infty} (m^{-1} \sum_{t=1}^{m-1} p^{-1} \text{tr} E(\Omega^{-1} x_m x'_m \varepsilon_m^2) E(\Omega^{-1} x_t x'_t \varepsilon_t^2)) = 1$, thus \mathcal{V} is an extra factor that appears in the limit due to nonlinear dependence in the data. In particular, it captures a *high-order serial correlation* of $x_t \varepsilon_t$, while $x_t \varepsilon_t$ itself does not have serial correlation since it is an martingale difference sequence.

The restriction on the summability rate of $\text{cov}(\text{tr}(\Upsilon_t \Xi_t), \text{tr}(\Upsilon_s \Xi_s))$ is related to the dimension p . To gain some insight, consider the case where the conditional moment Υ_t is homogeneous, so that $\Upsilon_t = I_p$ for all t . Then, some tedious algebra yields that $\text{cov}(\text{tr}(\Upsilon_t \Xi_t), \text{tr}(\Upsilon_s \Xi_s)) = O(n^2 p)$ uniformly over all s, t with $s < t$. This implies that the double sum of the covariances is $O(n^4 p)$ and thus meets the required condition as $p \rightarrow \infty$. Our assumption says that more generally this double sum over covariances must be $o(n^4 p^2)$ as $n, p \rightarrow \infty$.

For mean zero random variables $a_{1i}, a_{2j}, a_{3k}, a_{4l}$, let $\text{cum}_{ijkl}(a_{1i}, a_{2j}, a_{3k}, a_{4l})$ denote the fourth cumulant.

Assumption 5. $\{x_{ti} \varepsilon_t\}_{t \in \mathbb{Z}}$ is fourth order stationary for all $i = 1, \dots, p$. Furthermore, $\sup_{i,j=1,\dots,p} \sum_{t=-\infty}^{\infty} |c_{ij}(t)| < \infty$, where $c_{ij}(t) = E(x_{r,i} \varepsilon_r x_{r+t,j} \varepsilon_{r+t})$ for integer r , and $\sup_{i,j,k,l=1,\dots,p} \sum_{t_1, t_2, t_3=-\infty}^{\infty} |\text{cum}_{ijkl}(x_{0,i} \varepsilon_0, x_{t_1,j} \varepsilon_{t_1}, x_{t_2,k} \varepsilon_{t_2}, x_{t_3,l} \varepsilon_{t_3})| = O(n^2)$.

This assumption controls the temporal dependence in $\{x_t \varepsilon_t\}$ and is discussed in Andrews (1991b), for example, wherein sufficient conditions for it to hold are also provided. Then, the following theorem establishes the marginal convergence for a given γ , which is extended to weak convergence of the process in Theorem ST.2 in the online appendix.

Theorem 3.1. *Let Assumptions 1- 5 and \mathcal{H}_0 hold. Then $\mathcal{Q}_n(\gamma) \xrightarrow{d} \mathcal{N}(0, \mathcal{V})$, for a given $\gamma \in \Gamma$.*

Theorem 3.1 highlights a distinctive feature of testing growing number of restrictions in time series regressions. Unlike the independent cross-sectional case, we have to robustify the test against the HLV term \mathcal{V} . The provenance of this term can be illustrated by some formulae, details of which are contained in the full proofs. These proofs first establish (Theorem SL.6) that

$$(3.3) \quad \frac{W_n(\gamma) - p}{\sqrt{2p}} = \frac{\mathcal{R}_n(\gamma) - p}{\sqrt{2p}} + o_p(1),$$

where

$$(3.4) \quad \begin{aligned} \mathcal{R}_n(\gamma) &= [\gamma(1 - \gamma)n]^{-1} \left(\sum_{t=1}^{[n\gamma]} \varepsilon_t x_t - \gamma \sum_{t=1}^n \varepsilon_t x_t \right)' \Omega^{-1} \left(\sum_{t=1}^{[n\gamma]} \varepsilon_t x_t - \gamma \sum_{t=1}^n \varepsilon_t x_t \right) \\ &= \left(n^{-1/2} \sum_{t=1}^n \psi_t(\gamma) x_t \varepsilon_t \right)' \Omega^{-1} \left(n^{-1/2} \sum_{t=1}^n \psi_t(\gamma) x_t \varepsilon_t \right), \end{aligned}$$

and $\psi_t(\gamma) = (1(t/n \leq \gamma) - \gamma) / \sqrt{\gamma(1 - \gamma)}$. Also note that $n^{-1} \sum_{t=1}^n \psi_t(\gamma)^2 \rightarrow 1$. Thus, just as for the familiar Wald statistic, we have a quadratic form structure for $\mathcal{R}_n(\gamma)$. When p is fixed and there is no approximation error, we note that (3.3) has also been established by Andrews (1993), Cho and Vogelsang (2017) and Sun and Wang (2021).

This then yields the approximation

$$(3.5) \quad \frac{\mathcal{R}_n(\gamma) - p}{\sqrt{2p}} = \mathcal{S}_n(\gamma) + o_p(1),$$

where

$$\mathcal{S}_n(\gamma) = \frac{2}{\sqrt{2p}} \frac{1}{n} \sum_{t=2}^n \left(\Omega^{-\frac{1}{2}} \psi_t(\gamma) x_t \varepsilon_t \right)' \sum_{s < t} \left(\Omega^{-\frac{1}{2}} \psi_s(\gamma) x_s \varepsilon_s \right) = \frac{\sqrt{2}}{\sqrt{n}} \sum_{t=2}^n v_t(\gamma),$$

say, by Lemma SL.8. Then $\mathcal{V} = \lim_{n,p \rightarrow \infty} 2n^{-1} \sum_{s,t=2}^n \text{cov}(v_s(\gamma), v_t(\gamma))$, i.e. the limiting variance of $\mathcal{S}_n(\gamma)$. Note that the $v_t(\gamma)$ are defined as products of terms of the type $x_t \varepsilon_t$ and the cumulative sum of their lags, implying that the variances of the

$v_t(\gamma)$ themselves contain high-order covariance terms. This explains why we call \mathcal{V} a HLV despite the mds property of the $v_t(\gamma)$, which implies that $\{v_t(\gamma)\}$ is uncorrelated.

The next section establishes that the test based on \mathcal{Q}_n has nontrivial local power under suitable sequences of local alternatives, following which we study more detailed characteristics of \mathcal{V} and develop a HLV-robust test.

3.2. Local Alternatives. We consider a sequence of local alternatives that converge to the null at $p^{1/4}/\sqrt{n}$ -rate to study the local power properties of the test. This is slower than the usual $1/\sqrt{n}$ parametric rate and has been employed by a number of other authors, e.g. de Jong and Bierens (1994), Hong and White (1995), Gupta (2018). It is a cost of the nonparametric nature of the problem. Our sequence of local alternatives is:

$$(3.6) \quad \mathcal{H}_\ell : \delta_{2\ell} = 2^{1/4} \tau p^{1/4} / \sqrt{n},$$

where τ is a unit length $p \times 1$ vector.

Theorem 3.2. *Suppose that Assumptions 1- 5 and \mathcal{H}_ℓ hold and let $\tau_\infty = \lim_{n \rightarrow \infty} \tau' M \Omega^{-1} M \tau$. Then, $\mathcal{Q}_n(\gamma) \xrightarrow{d} \mathcal{N}(\tau_\infty \gamma(1 - \gamma), \mathcal{V})$.*

Note that $|\tau' M \Omega^{-1} M \tau| \leq \|\tau\| \|M\|^2 \|\Omega^{-1}\| = \bar{\lambda}(M)^2 / \underline{\lambda}(\Omega) < C$, for any n , by Assumptions 3, justifying an assumption on the existence of $\tau_\infty = \lim_{n \rightarrow \infty} \tau' M \Omega^{-1} M \tau$. Also, the noncentrality term is positive, implying nontrivial power of the test since the critical region is formed by $\mathcal{Q}_n(\gamma)$ being greater than equal to a critical value.

4. \mathcal{V} ROBUST TESTING

In this section we provide a detailed study of the HLV \mathcal{V} that our analysis has discovered. In particular, we present some alternative representations of \mathcal{V} that shed more light on its structure.

4.1. Discussion. We first examine the relevance of \mathcal{V} . Specifically, we analyze the ‘pre-limiting’ quantity $\mathcal{V}_n = 2\text{var} \left(n^{-1} \sum_{t=2}^n x_t' \Omega^{-1} \varepsilon_t p^{-1/2} \sum_{s=1}^{t-1} x_s \varepsilon_s \right)$. This can be

rewritten as

$$\mathcal{V}_n = 2n^{-1} \sum_{i=1}^{n-1} \left(\gamma(i, 0) (n-i)/n + 2 \sum_{j=1}^{n-i} \gamma(i, j) (n-i-j)/n \right),$$

where $\gamma(i, j) = p^{-1} E(x_t' \Omega^{-1} x_{t-i} x_t' \Omega^{-1} x_{t-i-j} \varepsilon_t^2 \varepsilon_{t-i} \varepsilon_{t-i-j})$. This is a high-order autocovariance and captures a nonlinear serial dependence in the sequence $x_t \varepsilon_t$, which disappears entirely for $j > 0$ in independent cross sectional data. We encounter $\mathcal{V}_n \rightarrow \mathcal{V} \neq 1$ when $n^{-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \gamma(i, j) (n-i-j)/n$ has a nonzero limit, with terms arising that are fourth-order cross-moments of the ε_t . Thus, the behaviour of such cross-moments, which are of the form $E(\varepsilon_t^2 \varepsilon_{t-i} \varepsilon_{t-i-j})$, is the key to obtaining non-unity \mathcal{V} . Robinson (1991), studying time series specification testing, encountered a similar term in the form of $E(\varepsilon_t^2 \varepsilon_{t-i} \varepsilon_{t-i-j})$ but imposed conditions that nullify it when $j > 0$.

A referee has pointed out that the Wald statistic is a quadratic form in the moment process. To establish the limit of the Wald statistic when the number of variables (i.e., the number of moments) grows with the sample size, we need to account for the variance of the quadratic form, hence the appearance of fourth-order dependence of a certain type in the moment process. A form of fourth-order dependence has also been encountered in HAR testing, see e.g. Lobato et al. (2002). In Section 6, we present some figures to show how \mathcal{V} can vary for various designs and deviate significantly from unity.

4.2. HLV-Robust Test Statistic. Heteroskedasticity and autocorrelation consistent (HAC) or HAR inference has been a main focus of time series inference, see e.g. Newey and West (1987), Andrews (1991b), Kiefer and Vogelsang (2002), just to name a few, and Lazarus et al. (2018) for a recent review. Although the asymptotic variance of the test statistic \mathcal{Q}_n depends on the HLV, we may still utilize the fixed-bandwidth kernel approach to obtain an asymptotically pivotal and mixed-normal test.

Introduce a kernel function $k(\cdot)$ that meets the following conditions.

Assumption 6. (1) For all $x \in \mathbb{R}$, $k(x) = k(-x)$ and $|k(x)| \leq 1$; $k(0) = 1$; $k(x)$ is continuous at zero and almost everywhere on \mathbb{R} ; $\int_{\mathbb{R}} |k(x)| dx < \infty$. (2) For any $b \in (0, 1]$ and $\rho \geq 1$, $k_b(x) = k(x/b)$ and $k^\rho(x)$ are symmetric, continuous, piecewise monotonic, and piecewise continuously differentiable on $[-1, 1]$. (3) $\int_{[0, \infty)} \bar{k}(x) < \infty$, where $\bar{k}(x) = \sup_{y \geq x} |k(y)|$.

Since ε_t and Ω are not directly observable in practice, we replace them with the least squares estimates as in Section 2 and introduce $q_t = (np)^{-1/2} x_t' \hat{\Omega}^{-1} \hat{e}_t \sum_{s=1}^{t-1} x_s \hat{e}_s$ and its demeaned version, $\bar{q}_t = q_t - n^{-1} \sum_{t=2}^n q_t$. Then, define a feasible estimate of \mathcal{V} by

$$(4.1) \quad \hat{\mathcal{V}} = \frac{2}{n} \sum_{t=2}^n \sum_{s=2}^n k\left(\frac{t-s}{nb}\right) \bar{q}_s \bar{q}_t.$$

Thus, we have a seemingly long-run variance estimate, analogous to traditional HAC/HAR inference, of a nonlinear transformation of the primitive variables.

The choice of bandwidth b has been a topic of much discussion in the HAC literature. Since \mathcal{V} captures high-order autocovariances in the growing dimensional vector $x_t \varepsilon_t$, the finite sample variation in the estimate $\hat{\mathcal{V}}$ is generally larger than in more familiar long-run variances, and the moment condition is more expensive. Motivated by this, we follow a fixed bandwidth approach, as in Sun (2014).

Our estimator is based on the weighting function $K_h(r, s) = k(h(r-s))$, where $h = 1/b$. We present numerical results in this paper with $k(u) = (1 - |u|)^h 1\{|u| < 1\}$, employing the Bartlett kernel case with $h = 1$. Sun (2014) terms this the sharp kernel estimator. Other options include the steep quadratic kernel estimator and the orthonormal series estimator with K basis function, of which Sun (2014) contains a more detailed discussion. Sun (2014) also shows that the centering in \bar{q}_t can be conveniently represented through a centered version of $K_h(\cdot)$, that is, $K_h^*(r, s) = K_h(r, s) - \int_0^1 K_h(\tau, s) d\tau - \int_0^1 K_h(r, \tau) d\tau + \int_0^1 \int_0^1 K_h(\tau_1, \tau_2) d\tau_1 d\tau_2$.

Building on the representation in Lemma 1 of Sun (2014), where the estimate $\hat{\mathcal{V}}$ is not consistent, we characterize the joint weak limit of $\hat{\mathcal{V}}$ and $\mathcal{Q}_n(\gamma)$. For real numbers

a and b , let $a \vee b$ ($a \wedge b$) denote their maximum (minimum), and introduce a process

$$(4.2) \quad \mathcal{Q}(\gamma) = \frac{W(\gamma)}{\gamma} + \frac{\bar{W}(\gamma)}{(1-\gamma)} - W(1),$$

where $(W(r), \bar{W}(r))'$, $r \in [0, 1]$, is a bivariate Gaussian process that does not depend on any model parameters including the break point γ , and has covariance kernel

$$(4.3) \quad \mathcal{C}(r_1, r_2) = \begin{pmatrix} (r_1 \wedge r_2)^2 & 1\{r_1 > r_2\}(r_1 - r_2)^2 \\ 1\{r_1 < r_2\}(r_1 - r_2)^2 & (1 - (r_1 \vee r_2))^2 \end{pmatrix}.$$

For any given $\gamma \in \Gamma$, the marginal distribution of $\mathcal{Q}(\gamma)$ is standard normal. Thus, the conclusion of Theorem 3.1 can be expressed as $\mathcal{Q}_n(\gamma) \xrightarrow{d} \mathcal{Q}(\gamma)$, pointwise in $\gamma \in \Gamma$. By taking a suitable ratio, we obtain a pivotal variable as in the following theorem, which is the basis of our test statistic.

Theorem 4.1. *Let Assumptions 1-6 hold, together with*

$$(4.4) \quad \lambda_n^{-2} p \left(v_p + \frac{p}{\sqrt{n}} \right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Under \mathcal{H}_0 , we have $\hat{\mathcal{V}} \xrightarrow{d} \mathcal{V} \int_0^1 \int_0^1 K_h^*(r, s) dW(r) dW(s)$ and

$$\mathcal{T}_n(\gamma) := \frac{\mathcal{Q}_n(\gamma)}{\sqrt{\hat{\mathcal{V}}}} \xrightarrow{d} \frac{\mathcal{Q}(\gamma)}{\sqrt{\int_0^1 \int_0^1 K_h^*(r, s) dW(r) dW(s)}}.$$

The numerator in the limit becomes $\mathcal{Q}(\gamma) + \tau_\infty \gamma(1 - \gamma)$ under \mathcal{H}_ℓ .

The asymptotic null distribution is mixed normal and pivotal. The critical values can be tabulated for each γ via Monte Carlo simulation and the authors' websites provide R code. A result when γ is unknown is given in Section 5.1. Note that the same Gaussian process $W(\cdot)$ occurs in both the limiting numerator and denominator, and this process is different from the Brownian motion in Sun (2014). In fact, it can be represented by the partial sum of $\sqrt{t/n}$ times an iid normal sequence. Since the limit also involves another variable $\bar{W}(\cdot)$, the critical values will be different from those previously tabulated in the literature.

4.3. Bias Correction. The degrees of freedom p provide a correct centering for $W_n(\gamma)$ in first order asymptotic analysis. However, in the finite sample experiments given in Section 6, e.g. Figure 2 and Figure 5, we find that the bias in $\mathcal{Q}_n(\gamma)$ gets bigger for typical values of p in nonparametric regression. Therefore, we propose a bootstrap bias correction of $\mathcal{Q}_n(\gamma)$. To estimate the bias, we implement the *null-imposed* wild bootstrap by generating

$$(4.5) \quad y_t^* = x_t' \hat{\delta}_1(\gamma) + \hat{e}_t(\gamma) \xi_t, \quad t = 1, \dots, n,$$

where ξ_t is an iid sequence of centered and normalized variables, e.g. the Rademacher variables, to compute $\mathcal{Q}_n^*(\gamma)$. It is worthwhile to note that the bootstrap DGP (4.5) imposes the null hypothesis $\delta_2 = 0$, so as not to sacrifice the power of the test. See also Gonçalves and Kilian (2007) for a thorough discussion on the wild bootstrap for infinite order autoregression. Iterating this B times, we obtain $\bar{\mathcal{Q}}_n^*(\gamma) = B^{-1} \sum_j^B \mathcal{Q}_n^{*,j}(\gamma)$, the bootstrap estimate of the bias. In our experiment, $B = 200$ suffices and thus the bootstrap is not computationally expensive. Therefore, we suggest the following bias corrected test statistic:

$$(4.6) \quad \mathcal{T}_n^b(\gamma) := \frac{\mathcal{Q}_n(\gamma) - \bar{\mathcal{Q}}_n^*(\gamma)}{\sqrt{\hat{\mathcal{V}}}}.$$

The numerical experiments in Section 6 show that the bootstrap bias corrected test controls the type I error reliably without sacrificing power unduly.² Now, with the superscript \star indicating the bootstrap analogue, we have the following result.

Theorem 4.2. *Under Assumptions 1-3 and \mathcal{H}_0 ,*

$$\sup_{\gamma \in \Gamma} |E^* W_n^*(\gamma) - p| = o_p(p^{1/2}).$$

Details of the components of $W_n^*(\gamma)$ are left to Section S.C of the online supplement.

²It is worth noting that the wild bootstrap may not be valid to approximate the quantiles of $\mathcal{Q}_n(\gamma)$ as it does not capture the high-order dependence embodied in \mathcal{V} .

5. EXTENSIONS

This section considers two useful extensions. One is a more general structural break test that does not assume a known break date, as considered by Andrews and Ploberger (1994) among many others. The other is the Wald test for general linear restrictions of growing ranks.

5.1. Testing with Unknown Break Date. The instability of an economic model over time is a major concern in empirical studies, leading to a commensurately large literature that is largely focussed on parametric models, see Perron (2006) for a review. In a parametric setting, Andrews (1993) considered testing with generalized method of moments (GMM) estimation, spawning contributions that widen the scope of this approach, such as Hidalgo and Seo (2013) and Qu and Perron (2007). Andrews and Ploberger (1994) and Elliott and Müller (2006) explore certain optimality properties of parametric tests.

In this case, the null hypothesis is specified as

$$\mathcal{H}_0 : \delta_2 = 0, \quad \text{for any } \gamma \in \Gamma,$$

where Γ is a closed interval in the interior of the unit interval. Extending Andrews and Ploberger (1994), we introduce a class of weighted exponential statistics:

$$(5.1) \quad \text{Exp}T_n(\mathfrak{c}) = \frac{\sqrt{2}}{\mathfrak{c}} \log \int_{\Gamma} \exp \left(\frac{\mathfrak{c}}{\sqrt{2}} \mathcal{T}_n^b(\gamma) \right) dJ(\gamma)$$

for a positive \mathfrak{c} and a bounded weight function J such that $\int_{\Gamma} dJ(\gamma) = 1$, where $\mathcal{T}_n^b(\gamma)$ is defined as in Section 4.3, and for which we set $\hat{\delta}_1 = \hat{\delta}_1(\hat{\gamma})$ and $\hat{e}_t = \hat{e}_t(\hat{\gamma})$ in (4.5), where $\hat{\gamma} = \arg \min \hat{\sigma}^2(\gamma)$ and $\hat{\sigma}^2(\gamma)$ is the sum of squared residuals as defined in Section 2. Since

$$(5.2) \quad \lim_{\mathfrak{c} \rightarrow 0} \text{Exp}T_n(\mathfrak{c}) = \int_{\Gamma} \mathcal{T}_n^b(\gamma) dJ(\gamma), \quad \text{and} \quad \lim_{\mathfrak{c} \rightarrow \infty} \text{Exp}T_n(\mathfrak{c}) = \sup_{\gamma \in \Gamma} \mathcal{T}_n^b(\gamma),$$

we may extend the definition of $\text{Exp}T_n(\mathfrak{c})$ for $\mathfrak{c} \in [0, +\infty]$.

The asymptotic distributions under both hypotheses follow from the continuous mapping theorem given the weak convergence of the stochastic process $\mathcal{Q}_n(\gamma)$ on Γ that is established in Theorem ST.2. The main result is stated in the following theorem.

Theorem 5.1. *Let Assumptions 1-6 hold, together with (4.4). Then,*

$$ExpT_n(\mathbf{c}) \xrightarrow{d} \frac{\sqrt{2}}{\mathbf{c}} \log \int_{\Gamma} \exp \left(\frac{\mathbf{c} \mathcal{Q}(\gamma)}{\sqrt{2 \int_0^1 \int_0^1 K_h^*(r, s) dW(r) dW(s)}} \right) dJ(\gamma),$$

under \mathcal{H}_0 , and

$$ExpT_n(\mathbf{c}) \xrightarrow{d} \frac{\sqrt{2}}{\mathbf{c}} \log \int_{\Gamma} \exp \left(\frac{\mathbf{c} \mathcal{Q}(\gamma) + \mathbf{c} \tau_{\infty} \frac{(\gamma \gamma_0 - (\gamma \wedge \gamma_0))^2}{\gamma(1-\gamma)}}{\sqrt{2 \int_0^1 \int_0^1 K_h^*(r, s) dW(r) dW(s)}} \right) dJ(\gamma),$$

under \mathcal{H}_{ℓ} as specified in (3.6), with γ_0 denoting the true break point.

The asymptotic distribution here is a natural generalization of that for the Chow test through stochastic equicontinuity of the underlying empirical process. The non-centrality term is also positive for any $\gamma \in \Gamma$ to make the test nontrivial.

Remark 1. *The test procedure can also be used if interest lies in testing for the stability of the coefficient of x_{1t} , where $x_t = (x'_{1t}, x'_{2t})'$ and the dimension p_1 of x_{1t} is less than p with $p_1 \rightarrow \infty$. Then, we can test $\delta_2 = 0$ in*

$$(5.3) \quad y_t = x'_{1t} \delta_1 + x'_{1t} \delta_2 1(t/n > \gamma) + e_t.$$

The asymptotic theory in Theorem 5.1 can still be used, as $p_1 \rightarrow \infty$ with $n \rightarrow \infty$, setting $q_t = (np_1)^{-1/2} x'_t \bar{\Omega}_1^{-1} \hat{e}_t \sum_{s=1}^{t-1} x_s \hat{e}_s$, where $\bar{\Omega}_1 = \hat{\Omega}^{-1} \bar{M}_1 (\bar{M}'_1 \hat{\Omega}^{-1} \bar{M}_1)^{-1} \bar{M}'_1 \hat{\Omega}^{-1}$ with $\bar{M}_1 = n^{-1} \sum_{t=1}^n x_t x'_{1t}$.

5.2. General Linear Restrictions. For a linear regression model $y_t = x'_t \beta + \varepsilon_t$, we consider testing a linear restriction $\mathcal{H}_0^e : R^e \beta = r$, where R^e is a matrix of rank $p \leq \dim(\beta)$. For the usual Wald statistic

$$(5.4) \quad W_n^e := n \left(R^e \hat{\beta} - r \right)' \left(R^e \hat{M}^{-1} \hat{\Omega} \hat{M}^{-1} R^e \right)^{-1} \left(R^e \hat{\beta} - r \right),$$

$\hat{M} = n^{-1} \sum_t x_t x_t'$ and $\hat{\Omega}$ is an estimator of $E \varepsilon_t^2 x_t x_t'$, define $\mathcal{Q}_n^e := (W_n^e - p) / \sqrt{2p}$. Although the test statistic appears to be very similar to the Chow test, the next theorem shows that the numerator and denominator in our corrected test statistic are related differently, calling for different critical values. Furthermore, the HLV is now obtained by replacing Ω^{-1} in Assumption 4 with $L = M^{-1} R^{e'} (R^e M^{-1} \Omega M^{-1} R^{e'})^{-1} R^e M^{-1}$, and the resulting limit denoted \mathcal{V}^e .

To estimate \mathcal{V}^e and employ bootstrap bias correction, it is convenient to reformulate the restriction as an exclusion restriction of growing dimension, without loss of generality. Indeed, let S be the orthogonal complement of R^e and $Q = (S, R^e)$. Then, let $\hat{x}_t = Q'^{-1} x_t$, $\delta = Q\beta - (0', r')'$ such that $\delta = (\delta'_1, \delta'_2)'$, with $\hat{x}_t = (\hat{x}'_{1t}, \hat{x}'_{2t})'$ conformably partitioned, $\delta_2 = R^e \beta - r$ and $\tilde{y}_t = y_t - \hat{x}'_{2t} r$. We can now test the null hypothesis $\mathcal{H}_0^e : \delta_2 = 0$ in the regression of \tilde{y}_t on \hat{x}_t .

This transformation makes it particularly convenient to impose the null in the bootstrap resampling at the bias correction stage. Let $\bar{\mathcal{Q}}_n^*$ denote the bootstrap bias correction factor for W_n^e . This yields the bias-corrected statistic $\mathcal{T}_n^{e,b} = (\mathcal{Q}_n^e - \bar{\mathcal{Q}}_n^*) / \sqrt{\hat{\mathcal{V}}^e}$, where $\hat{\mathcal{V}}^e$ is defined analogously to $\hat{\mathcal{V}}$, but now with $q_t = (np)^{-1/2} \tilde{x}'_{2t} \hat{\Omega}^{e-1} \hat{e}_t \sum_{s=1}^{t-1} \tilde{x}_{2t} \hat{e}_s$, where \tilde{x}_{2t} denotes the residuals from the regression of \hat{x}_{2t} on \hat{x}_{1t} and $\hat{\Omega}^e = n^{-1} \sum_t \tilde{x}_{2t} \tilde{x}'_{2t} \hat{e}_t^2$. Then, with $W(\cdot)$ defined in (4.3), we have the following theorem:

Theorem 5.2. *Let Assumptions 1-6 hold with the following modifications: (1) L replacing Ω^{-1} in Assumption 4 and the resulting limit denoted \mathcal{V}^e . (2) The conditions on the partial sums to $[n\gamma]$ are omitted in Assumption 3(ii). Also suppose that (4.4) holds. Then, under \mathcal{H}_0^e ,*

$$(5.5) \quad \mathcal{T}_n^{e,b} = \frac{\mathcal{Q}_n^e - \bar{\mathcal{Q}}_n^*}{\sqrt{\hat{\mathcal{V}}^e}} \xrightarrow{d} \frac{W(1)}{\sqrt{\int_0^1 \int_0^1 K_h^*(r, s) dW(r) dW(s)}}.$$

Under $\mathcal{H}_\ell^e : R^e \beta - r = 2^{1/4} \tau p^{1/4} / \sqrt{n}$, the numerator in the limit becomes $W(1) + \tau_\infty^e$, where $\tau_\infty^e = \lim_{n \rightarrow \infty} \tau' (R^e M^{-1} \Omega M^{-1} R^{e'})^{-1} \tau$.

The limiting distribution is mixed normal and pivotal but different from the limit in Theorem 3.1. This is because the Chow test considers a quadratic form in

$n^{-1/2} \sum_{t=1}^n \psi_t(\gamma) x_t \varepsilon_t$, which differs from this section by introducing a trend into the regressors via the factor $\psi_t(\gamma)$. Due to this difference, the partial sum processes converge to Gaussian processes with different covariance kernels. An R code to compute the critical values is available from the author's website.

6. MONTE CARLO

This section examines the finite sample properties of our bias corrected HLV-robust test \mathcal{T}_n^b compared to the standard chi-square test W_n , which does not account for growing p , and the unscaled \mathcal{Q}_n statistic with standard normal critical values, which does not account for \mathcal{V} , in terms of bias, size and power.

We will consider the examples below in our Monte Carlo experiments.

E1 (Multiple Regression of Growing Dimension). *Koenker (1988) found through his metastudy that it is common practice in econometrics to increase the number of regressors as the sample size n grows, at a rate of roughly $O(n^{1/4})$. In this case, the approximation error r_t is not explicitly modeled and may be set as zero. Practitioners thus adopt a flexible approach to modelling, where the assumed model becomes richer with more covariates and with more lagged terms to account for the dynamic effect in the spirit of the distributed lag model, as illustrated in e.g. Stock and Watson (2015).*

E2 (Infinite-Order Autoregression). *This model is one of the most fundamental models in time series analysis, see e.g. Brockwell et al. (1991) or Hamilton (2020). For the process to be stationary, the coefficients $\{b_j\}$ in the $AR(\infty)$ model $y_t = b_0 + \sum_{j=1}^{\infty} b_j y_{t-j} + \varepsilon_t$ are assumed to obey a certain decay rate. Specifically, the tail sum of the coefficients satisfies Assumption 2 if $\sum_{j=p}^{\infty} |b_j| = o(n^{-1/2})$. While we take p as given in our analysis, for practical purposes various methods based on information criteria are available to choose the truncation lag p , see e.g. Shibata (1980) and references therein. Wang et al. (2007) propose a lasso-based autoregressive order selection rule while Lee et al. (2018) propose a lag selection rule in an infinite order panel autoregression. For expositional ease, we assume that the observations begin from $t = 1 - p$ and $x_1 = (1, y_0, \dots, y_{1-p})$.*

E3 (Nonparametric Series Regression). *In case of the nonparametric series least squares estimation of $E(y_t|z_t)$, there exists a sequence of transformations of the covariates z_t given by $x_{nt} := x_n(z_t) : \mathbb{R}^k \mapsto \mathbb{R}^p$, and coefficients β_n such that $E(y_t|\mathcal{F}_{t-1}) = f(z_t) = x'_{nt}\beta_n + r_n(z_t)$, where $r_{nt} = r_n(z_t)$ meets Assumption 2 for a broad class of functions f , see e.g. Andrews (1991a), Newey (1997), Chen (2007) and Lee and Robinson (2016). By Lemma 1 of Lee and Robinson (2016), it is met if $|r_t|_\infty = O(p^\alpha)$ for some $\alpha < 0$ and $p^{2\alpha} \leq n^{-1}$. Depending on the smoothness of the nonparametric function $f(\cdot)$, the regressor support dimension k , and the type of basis functions used, different values of α may be implied, see e.g. Newey (1997), Chen (2007), p. 5573, for examples and further references. Often, the condition (2.2) holds under the so-called undersmoothing selection of p . Another closely related example is the partially linear regression model, e.g. Engle et al. (1986) and Robinson (1988). Again, while we do not consider data-dependent p , for practical purposes the literature proposes methods for the choice of p using cross validation or information criteria, see e.g. p. 5623 of Chen (2007) for a list of references.*

The tests are applied to the setting of the Chow test, break test with unknown break date, and testing a general linear restriction. We consider various sample sizes n and dimensions p from the three examples, E1-E3, with the error generated from a bounded ARCH process

$$(6.1) \quad \varepsilon_t = \sigma_t \xi_t, \quad \sigma_t^2 = (1 - \alpha) + \alpha \phi(\varepsilon_{t-1}),$$

where $\phi(x) = x^2 1\{|x| \leq c\} + c^2 1\{|x| > c\}$, $\xi_t = (\eta_t - E\eta_t) / \sqrt{\text{var}(\eta_t)}$, and $\{\eta_t\}$ is an iid sequence from the Marron and Wand (1992) normal mixture distributions of type 1-3, which we refer to as error 1, 2 and 3. Their error 1 is standard normal. For a standard normal vector (Z_1, \dots, Z_k) and multinomial vector (d_1, \dots, d_k) with probability $(1/5, 1/5, 3/5)$, error 2 skewed unimodal variate is $\eta_t = Z_1 d_1 + (2Z_2/3 + 1/2) d_2 + d_3 (5Z_3/9 + 13/12)$, while error 3 strongly skewed variate is $\eta_t = \sum_{l=0}^7 d_{l+1} \left(Z_{l+1} (2/3)^l + 3 \left((2/3)^l - 1 \right) \right)$ with equally likely d_i 's. We

report results using (6.1) with $\alpha \in \{0.3, 0.4, 0.5, 0.55, 0.577\}$ and $c = 2.5$. Results from $c = 3$ and ∞ are similar and omitted.

More specifically, for multiple regression, E1, the regressors x_t consist of independent AR(1) processes with coefficient α_x and ARCH innovation as in (6.1) and their lags of order up to 3. That is, we consider the distributed lag model with a growing number of variables. The first five elements of β in (2.1) are set as $d_0 (5^{-1/2}, \dots, 5^{-1/2}) p^{1/4} n^{-1/2}$ and the others as zeros. When there is a break, all the values become zero after the break so that the value d_0 controls the magnitude of the change. We vary $p \in \{5, 9, 13\}$ to examine the effect of the dimension on our tests.

For the infinite order AR regression, E2, we generate the sample from the MA(1) model $y_t = \varepsilon_t + \theta 1\{t \leq \mu\} \varepsilon_{t-1}$, $\mu = \lfloor n\gamma \rfloor$, and estimate the AR(p) model with $p = 9$ for $n = 250$ and $p = 13$ for $n = 500$. For the sieve regression, E3, we consider two variables ζ_{1t} and ζ_{2t} and their lags $\zeta_{1,t-1}$ and $\zeta_{2,t-1}$ as regressors, denoted by z_{1t}, \dots, z_{4t} , after transforming them as $2 \arctan(\zeta_{it}) / \pi$. Each ζ_{it} follows an AR(1) process with ARCH error. The regression function is set as $f(z_1, \dots, z_4) = d_0 (1, z_1, \dots, z_4, z_1^2, \dots, z_4^{p_1-1}) (1^{-2}, \dots, p_2^{-2})' + \sqrt{|z_1|/n}$ with $p_1 = \lfloor 1000^{1/4} \rfloor$ and $p_2 = 4p_1$. To estimate the regression function, we construct x_t from polynomial basis functions and its dimension p as in E2.

We first employ these DGPs to simulate values of \mathcal{V} with $n = 1000$, which are plotted in Figure 1, reporting averages from 2500 iterations. This serves as a useful illustration to observe visually that \mathcal{V} deviates from unity for various specifications. A broad observation we make is that the deviation is bigger with larger ARCH coefficients and bigger autocorrelation in x_t , although this feature is not monotone. To conclude, we observe that the nonlinear serial autocorrelation factor \mathcal{V} can induce serious distortion in inference without a suitable robustifying treatment, as we provide in section 4.2.

6.1. Chow Test. We consider three candidate break points as proportions of the sample sizes, $\gamma \in \{0.2, 0.3, 0.5\}$. We begin by examining the bias of $\mathcal{Q}_n(\gamma)$, conventionally centered by the degrees of freedom p , under the null hypothesis. Note that

a severe bias in $\mathcal{Q}_n(\gamma)$ also implies that the size of the Wald test $W_n(\gamma)$ can be distorted severely. We report the results in Figure 2, in which the line with dot markers shows the bias in $\mathcal{Q}_n(\gamma)$ for $n = 250$ and $n = 500$. For E1, (Figures 2a, 2d), each vertical partition (marked by a dotted vertical line) corresponds to a specific value of p . Within each vertical partition the DGP parameters change along the horizontal axes as (error type, α), in lexicographic order. As p grows, we observe that the $\mathcal{Q}_n(\gamma)$ statistic exhibits severe finite sample bias for all values of the DGP parameters.

A similar visualization of bias in $\mathcal{Q}_n(\gamma)$ for E2 is presented in Figures 2b, 2e. Rather than report values for different p , here we focus on the case $p = 9$ for $n = 250$ and $p = 13$ for $n = 500$ and allow the values of α and θ to vary along the horizontal axis lexicographically as (error type, α or θ), as detailed in the caption. A substantial bias in $\mathcal{Q}_n(\gamma)$ is observed for all cases, regardless of $n = 250$ or $n = 500$, albeit the biases are generally smaller in the latter case. Finally, Figures 2c, 2f show the bias in $\mathcal{Q}_n(\gamma)$ for E3, with the same p as for E2, to mimic the asymptotic regime of a sieve regression, and parameters as in E2. We observe a similar pattern of substantial bias for both sample sizes.

As discussed above, Figures 2a-2f clearly show that the biases present in $\mathcal{Q}_n(\gamma)$ are severe. In these figures we also plot the bias of the bias-corrected HLV-robust statistic $\mathcal{T}_n^b(\gamma)$, shown in black with square markers. The bootstrap bias correction seems to work well for all the cases, substantially alleviating bias. In Figure 2, we observe that $\mathcal{T}_n^b(\gamma)$ can still exhibit some bias for specific cases but for E1 and E3, unlike the bias of $\mathcal{Q}_n(\gamma)$, this is centered around zero, while for E2 it is generally smaller in absolute value. Thus we recommend the use of the bootstrap bias correction in practice especially when faced with large values of p .

We now study the finite sample rejection frequencies of four competing tests: $\mathcal{T}_n^b(\gamma)$, $\mathcal{T}_n(\gamma)$, $\mathcal{Q}_n(\gamma)$, and $W_n(\gamma)$, with specific parameter values as given in the respective figure captions. As shown earlier, the unknown HLV scaling factor \mathcal{V} varies along different ARCH parameters. This motivates our approach of experimenting with different α values and innovations. The Monte Carlo sizes resulting from the

experiment are plotted in Figure 3, wherein we place a horizontal dotted line to mark the nominal size of 5%. We report results for $\gamma = 0.3$. The vertical partitions in each panel of Figure 3 correspond, as discussed earlier, to increasing values of p from left to right in E1. We cover multiple regression (Figures 3a, 3d), AR fits (Figures 3b, 3e) and sieve regression (Figures 3c, 3f) for $n = 250, 500$.

For all DGPs, the usual Wald statistic $W_n(\gamma)$ (diamond markers) over-rejects. Simply standardizing the test statistic $W_n(\gamma)$ to $\mathcal{Q}_n(\gamma)$, hence ignoring the HLV \mathcal{V} , does not improve matters. In fact, it usually worsens the problem of over-rejection. This can be seen in the lines with triangle markers. Our HLV-robust statistic $\mathcal{T}_n(\gamma)$ does much better, as the lines with dot markers indicate. While this shows the importance of the correction for \mathcal{V} that we stress in the paper, there is still a tendency to over-reject. On the other hand, applying the bootstrap bias correction and using the bias corrected HLV-robust statistic $\mathcal{T}_n^b(\gamma)$ achieves excellent size control, as can be seen in the line with square markers. The discussion holds regardless of whether $n = 250$ or $n = 500$. Thus the importance of our proposed testing procedure is clearly visible.

We now analyze the power features of the competing test statistics for the proposed DGPs, allowing for breaks of different magnitudes and setting $\gamma = 0.5$. After the break all the coefficients become zero so that the values of d_0 govern the size of the breaks in E1 and E3, while the values of θ do so for E2. The power performance is plotted in Figure 4, where to conserve space we report results only for $n = 400$. Again, we use $p = 9$ for E2 and E3, while a range of p is employed for E1. The line marker schemes for each of the competing tests are as described earlier. Examining the figure, the power of our HLV-robust statistics $\mathcal{T}_n(\gamma)$ (dots) and $\mathcal{T}_n^b(\gamma)$ (squares) tracks that of the uncorrected ones as the break size increases for both E2 (center panel) and E3 (right panel). For E1 (left panel), we only report results for $d_0 = 2$ for clarity. We observe that $W_n(\gamma)$ tends to have the highest power but our statistics still perform reasonably well with power in excess of 80% even for large p . Recall that our size experiments earlier indicate that $W_n(\gamma)$ over-rejects, a phenomenon of which

high power is likely an artefact. Thus we conclude that our test is able to control size without sacrificing power to an undue extent.

6.2. Testing a Linear Restriction. This section presents the outcomes of bias, size and power experiments for testing general linear restrictions, analogous to those for the Chow test in the preceding discussion. We use the reparameterization of the linear restriction to the exclusion restriction $\delta_2 = 0$, as discussed in Section 5.2. We focus on E1 with $n = 400$, $p = 8, 12, 16$, $d_0 = 1$, error 1 and 2 disturbances and $\alpha = 0.4, 0.55$. The results are displayed in Figure 5, with the same marking scheme as before and three test statistics employed: W_n^e , Q_n^e and $\mathcal{T}_n^{e,b}$. In all three figures, each vertical partition marks a different value of p , increasing from left to right.

The left panel of Figure 5 shows that the bootstrap bias correction indeed improves matters, as was the case for the Chow test. The center panel again demonstrates the importance of our proposed corrections for size control. W_n^e and Q_n^e tend to over-reject, becoming worse as p increases. $\mathcal{T}_n^{e,b}$ controls size very well for medium to large p , while still outperforming W_n^e and Q_n^e for smaller p . The right panel shows that $\mathcal{T}_n^{e,b}$ sacrifices some power relative to W_n^e and Q_n^e , but not unduly so.

6.3. Sup Test for Structural Break. In this section we examine the size and power performance of the sup test for the presence of a structural break at an unknown date, which is a special case of the exp test, viz. $\lim_{c \rightarrow \infty} \text{Exp}T_n(c)$, see (5.2). Specifically, we compare $\sup_{\gamma \in \Gamma} T_n^b(\gamma)$ with $\sup_{\gamma \in \Gamma} Q_n(\gamma)$ and $\sup_{\gamma \in \Gamma} W_n(\gamma)$, where the critical values for the last test come from Andrews (1993). To conserve space, we report results for E1 and error 1 and 2 disturbances only. We also take $\alpha = 0.4, 0.55$, $p = 5, 7, 9$, $n = 400$ and $\Gamma = [0.15, 0.85]$.

The vertical partitions in Figure 6 represent $p = 5, 7, 9$, (left panel) and $d_0 = 1, 5, 10$ (right panel), increasing from left to right. Examining size control (left panel), the traditional sup test $\sup W_n(\gamma)$ (triangles) over rejects, with performance getting poorer as p increases. The correction for p immediately improves matters, with $\sup Q_n(\gamma)$ (dots) still over-rejecting but to a lesser degree. Our recommended statistic $\sup \mathcal{T}_n^b(\gamma)$

(squares) does best overall as p increases, and achieves quite acceptable size performance even when p is moderate. The right panel displays power, with break size increasing in each vertical partition from left to right as $d_0 = 1, 5, 10$. Our recommended statistic $\sup \mathcal{T}_n^b(\gamma)$ is seen to not unduly sacrifice power relative to $\sup \mathcal{Q}_n(\gamma)$.

7. EMPIRICAL EXAMPLE

We revisit structural stability in the Hamilton (2003) study of the effect of oil shocks on economic activities. The autoregressive distributed lag model, $\text{ADL}(p, p)$, with quarterly time series of outputs and several oil price measures is employed. For real output, the quarterly growth rate of chain-weighted real GDP is used, while the oil price is the nominal crude oil producer price index, seasonally unadjusted. As in Hamilton, three oil price measures were considered: the growth rate o_t from the previous quarter, the rectified linear unit, $o_t^+ = o_t 1\{o_t > 0\}$, and the net oil price increase, o_t^n , defined as the amount by which log oil prices in quarter t exceed their peak value over the previous 12 months. If it does not exceed the previous peak, then o_t^n is taken to be zero. We extend the original sample using the FRED database at the St. Louis Fed to obtain a sample from January 1949 to October 2019.

First, we reevaluate structural stability of the GDP dynamics using $\text{AR}(p)$ fits, and that of the regression function of GDP growth on oil price change using an $\text{ADL}(p, p)$ model with the three alternative measures of oil price change. As in Hamilton (2003), we do not hypothesize a predetermined break date and use $\sup_{\gamma \in \Gamma} \mathcal{T}_n^b(\gamma)$ and $\sup_{\gamma \in \Gamma} W_n(\gamma)$, $\Gamma = [0.15, 0.85]$. The p-values of the tests are reported in subtables (a) and (b) in the upper panel of Table 1, where for $\sup W(\gamma)$ these are computed using the R function of Hansen (1997). We obtain conflicting results: the standard $\sup W(\gamma)$ test supports the presence of structural break in both regression more strongly than our recommended $\sup \mathcal{T}_n^b(\gamma)$ test. Judging from our Monte Carlo simulation that illustrates the effect of the degree of the freedom (df) on finite sample properties of the two tests, $\sup W_n(\gamma)$ tends to have larger p values in the AR case ($p + 1$ df) than in the ADL case ($2p + 1$ df), while $\sup \mathcal{T}_n^b(\gamma)$ would be the opposite.

Thus, the evidence for structural instability is no longer as strong, often overturned at reasonable significance levels.

	(a) AR(p)		(b) ADL(p, p) stability						(c) ADL(p, p) exclusion					
GDP			o_t		o_t^+		o_t^n		all oil		NL			
lags p	4	6	4	6	4	6	4	6		4	6	4	6	
$\sup \mathcal{T}_n^b(\gamma)$	40.4	22.4	7.2	2.9	0.08	1.7	0	0	$\mathcal{T}_n^{e,b}$	8.7	26.9	7.6	19	
$\sup W_n(\gamma)$	7.2	4.1	0	0	2	0	0	0	W_n^e	9	4.4	2.8	4.4	
IP			o_t		o_t^+		o_t^n		all oil		NL			
lags p	12	18	12	18	12	18	12	18		12	18	12	18	
$\sup \mathcal{T}_n^b(\gamma)$	35.6	53.4	4.3	6.6	1	1.4	1.4	0.9	$\mathcal{T}_n^{e,b}$	27	30.9	11.9	18.5	
$\sup W_n(\gamma)$	0	0.5	0	0	0	0	0	0	W_n^e	1.7	0.8	0.5	0.5	

TABLE 1. $100 \times$ p-values of stability and exclusion restriction tests for full sample. $\gamma \in [0.15, 0.85]$ (a) Tests for stability of GDP or IP dynamics via AR(p) fits. (b) Tests for stability of ADL(p, p) regressions of GDP or IP on o_t , o_t^+ or o_t^n . (c) Tests for exclusion restrictions on all oil price measures (o_t , o_t^+ , o_t^n) or nonlinear oil price measures (o_t^+ , o_t^n) in ADL(p, p) regressions of GDP or IP on oil prices.

Second, we explore the relevance of the oil price measures and the relevance of the nonlinear transformations (o_t^+ , o_t^n) by testing two exclusion restrictions in the ADL(p, p) regression that include all the three oil price measures as covariates. The first exclusion restriction is to set the coefficients of all the measures zero and the second is to set those of the nonlinear transformations (o_t^+ , o_t^n) to zero. This yields 12 and 8 df, respectively, when $p = 4$, and 18 and 12 df, respectively, when $p = 6$. As shown in sub-table (c) of the upper panel of Table 1, our recommended test $T_n^{e,b}$ produces p-values bigger than 0.05 for all cases, suggesting the effect of oil price as measured by these transformations is not statistically significant, nor are the nonlinear transformations. The standard Wald test for the exclusion restriction is more supportive of their inclusion but may lack robustness with large df.

As another measure of economic activity we now consider the industrial production (IP) index. This is available at monthly frequency and thus we consider ADL(12,12) and ADL(18,18) to include lags of one year and one and a half years, respectively. With monthly data, the dimensionality becomes more important: the number of

GDP	(a) AR(p)		(b) ADL(p, p) stability						(c) ADL(p, p) exclusion					
SS1			o_t		o_t^+		o_t^n		all oil			NL		
lags p	4	6	4	6	4	6	4	6		4	6	4	6	
sup $\mathcal{T}_n^b(\gamma)$	8.7	0.9	3.8	3.1	0.2	15.8	3.9	11.4	$\mathcal{T}_n^{e,b}$	73.8	74.2	92.6	85.6	
sup $W_n(\gamma)$	0.7	0	0	0	0	0	0	0	W_n^e	65.4	46	78.5	47.5	
SS2			o_t		o_t^+		o_t^n		all oil			NL		
lags p	4	6	4	6	4	6	4	6		4	6	4	6	
sup $\mathcal{T}_n^b(\gamma)$	46.6	26.1	4.3	3.3	9.1	2	0	0	$\mathcal{T}_n^{e,b}$	11.6	24.2	23.2	34.8	
sup $W_n(\gamma)$	45.3	7.1	0	0	0.1	0	0	0	W_n^e	0.6	1.4	9.4	11	
IP	(a) AR(p)		(b) ADL(p, p) stability						(c) ADL(p, p) exclusion					
SS1			o_t		o_t^+		o_t^n		all oil			NL		
lags p	12	18	12	18	12	18	12	18		12	18	12	18	
sup $\mathcal{T}_n^b(\gamma)$	55.7	3.8	1.9	3.3	4.4	11	73.9	54.9	$\mathcal{T}_n^{e,b}$	25.3	20.4	7.2	6.3	
sup $W_n(\gamma)$	23.4	0.2	0	0	0	0	0	0	W_n^e	0.8	0	0.4	0	
SS2			o_t		o_t^+		o_t^n		all oil			NL		
lags p	12	18	12	18	12	18	12	18		12	18	12	18	
sup $\mathcal{T}_n^b(\gamma)$	26.4	45.6	3.6	5.6	0.6	0.4	1.6	0.3	$\mathcal{T}_n^{e,b}$	24.8	29.5	18.1	27.2	
sup $W_n(\gamma)$	0	0	0	0	0	0	0	0	W_n^e	0.3	0.1	0.6	1.5	

TABLE 2. 100×p-values of stability and exclusion restriction tests for subsamples. SS1: 1981:I-2019:IV, SS2: 1950:I-2007:II; $\gamma \in [0.15, 0.85]$. (a) Tests for stability of GDP dynamics via AR(p) fits. (b) Tests for stability of ADL(p, p) regressions of GDP or IP on o_t , o_t^+ or o_t^n . (c) Tests for exclusion restrictions on all oil price measures (o_t , o_t^+ , o_t^n) or nonlinear oil price measures (o_t^+ , o_t^n .) in ADL(p, p) regressions of GDP or IP on oil prices.

restrictions we test varies from 13 and 25 in the structural break test for the AR(12) and ADL(12,12) regressions to 36 and 48 for the exclusion tests in the ADL(18,18) regression. In fact, Andrews (1993) and Hansen (1997) do not even provide critical values when $df > 20$, giving even more reason for our method to be applied. The results in the lower panel of Table 1 are qualitatively similar to those for the GDP measure.

Building on studies on major oil shocks such as Hamilton (2009), we also consider two subsamples: SS1 starts after the 1980 oil shock while SS2 ends before the 2007 oil shock. As for the lag order p , we try both $p = 4$ and $p = 6$ for robustness.

Furthermore, we also test subsample time series and regression stability as well as exclusion restrictions for the dependent variable IP, using $p = 12$ and $p = 18$ as before. The p-values are reported in Table 2, with the each panel covering GDP and IP and containing subpanels for SS1 and SS2. Again, we note that evidence against the null becomes weaker when using our recommended tests $\sup \mathcal{T}_n^b(\gamma)$ or $\mathcal{T}_n^{e,b}$, as compared to the standard Wald statistic approach. Very often the conclusion of the test is changed when using our approach, but even when this is not the case there can be large differences in p-values.

REFERENCES

- Anatolyev, S. (2012). Inference in regression models with many regressors. *Journal of Econometrics* 170, 368–382.
- Anatolyev, S. (2019). Many instruments and/or regressors: A friendly guide. *Journal of Economic Surveys* 33, 689–726.
- Andrews, D. W. K. (1991a). Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* 59, 307–345.
- Andrews, D. W. K. (1991b). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.
- Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.
- Brockwell, P. J., R. A. Davis, and S. E. Fienberg (1991). *Time Series: Theory and Methods*. Springer Science & Business Media.
- Calhoun, G. (2011). Hypothesis testing in linear regression when k/n is large. *Journal of Econometrics* 165, 163–174.
- Cattaneo, M. D., M. Jansson, and W. K. Newey (2018). Inference in linear regression models with many covariates and heteroscedasticity. *Journal of the American Statistical Association* 113, 1350–1361.
- Chen, B. and Y. Hong (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80, 1157–1183.
- Chen, G. and R. A. Lockhart (2001). Weak convergence of the empirical process of residuals in linear models with many parameters. *Annals of Statistics* 29, 748–762.
- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 76, pp. 5549–5632. North Holland.
- Chen, X. and T. M. Christensen (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics* 188, 447–465.
- Cho, C. K. and T. J. Vogelsang (2017). Fixed- b inference for testing structural change in a time series regression. *Econometrics* 5, 1–26.

- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28, 591–605.
- de Jong, R. M. and H. J. Bierens (1994). On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity. *Econometric Theory* 10, 70–90.
- Elliott, G. and U. K. Müller (2006). Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies* 73, 907–940.
- Engle, R. F., C. W. Granger, J. Rice, and A. Weiss (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association* 81, 310–320.
- Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *Journal of Finance* 75, 1327–1370.
- Gonçalves, S. and L. Kilian (2007). Asymptotic and bootstrap inference for $AR(\infty)$ processes with conditional heteroskedasticity. *Econometric Reviews* 26, 609–641.
- Gupta, A. (2018). Nonparametric specification testing via the trinity of tests. *Journal of Econometrics* 203, 169–185.
- Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics* 113, 363–398.
- Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007-08. *Brookings Papers on Economic Activity SPRING*, 215–261.
- Hamilton, J. D. (2020). *Time Series Analysis*. Princeton University Press.
- Hansen, B. E. (1997). Approximate asymptotic p values for structural-change tests. *Journal of Business & Economic Statistics* 15, 60–67.
- Hidalgo, J. and M. H. Seo (2013). Testing for structural stability in the whole sample. *Journal of Econometrics* 175, 84–93.
- Hong, Y. and H. White (1995). Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.
- Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.
- Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review* 95, 161–182.
- Kiefer, N. M. and T. J. Vogelsang (2002). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica* 70, 2093–2095.
- Kline, P., R. Saggio, and M. Sølvesten (2020). Leave-out estimation of variance components. *Econometrica* 88, 1859–1898.
- Koenker, R. (1988). Asymptotic theory and econometric practice. *Journal of Applied Econometrics* 3, 139–147.
- Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018). HAR inference: Recommendations for practice. *Journal of Business & Economic Statistics* 36, 541–559.
- Lee, J. and P. M. Robinson (2016). Series estimation under cross-sectional dependence. *Journal of Econometrics* 190, 1–17.
- Lee, Y.-J., R. Okui, and M. Shintani (2018). Asymptotic inference for dynamic panel estimators of infinite order autoregressive processes. *Journal of Econometrics* 204, 147–158.
- Lobato, I. N., J. C. Nankervis, and N. E. Savin (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory* 18, 730–743.
- Marron, J. S. and M. P. Wand (1992). Exact mean integrated squared error. *Annals of Statistics* 20, 712–736.

- Mohr, M. and N. Neumeyer (2020). Consistent nonparametric change point detection combining CUSUM and marked empirical processes. *Electronic Journal of Statistics* 14, 2238–2271.
- Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Peligrad, M. (1982). Invariance principles for mixing sequences of random variables. *Annals of Probability* 10, 968–981.
- Perron, P. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics* 1, 278–352.
- Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica* 75, 459–502.
- Robinson, P. M. (1988). Root- n -consistent semiparametric regression. *Econometrica* 56, 931–954.
- Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.
- Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* 8, 147–164.
- Smith, R. L. (1992). Some interlacing properties of the Schur complement of a Hermitian matrix. *Linear Algebra and its Applications* 177, 137–144.
- Stock, J. H. and M. W. Watson (2015). *Introduction to Econometrics*. Pearson.
- Sun, Y. (2014). Fixed-smoothing asymptotics in a two-step GMM framework. *Econometrica* 82, 2327–2370.
- Sun, Y. and X. Wang (2021). An asymptotically F-distributed Chow test in the presence of heteroscedasticity and autocorrelation. *Econometric Reviews forthcoming*, 1–35.
- Wang, H., G. Li, and C.-L. Tsai (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69, 63–78.
- Wu, J. S. and C. K. Chu (1993). Kernel-type estimators of jump points and values of a regression function. *Annals of Statistics* 21, 1545–1566.

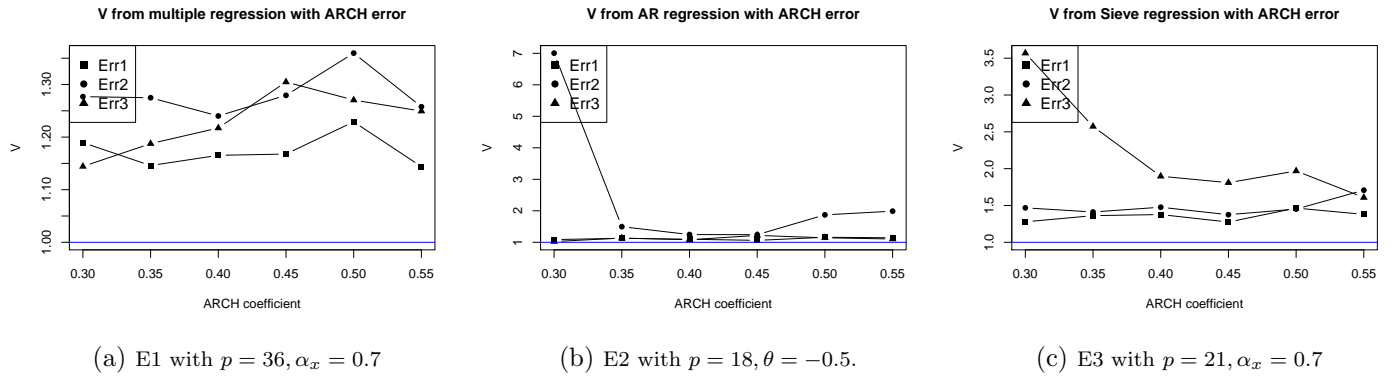


FIGURE 1. Simulated \mathcal{V} from $n = 1000$. Error 1: square; Error 2: dot; Error 3: triangle.

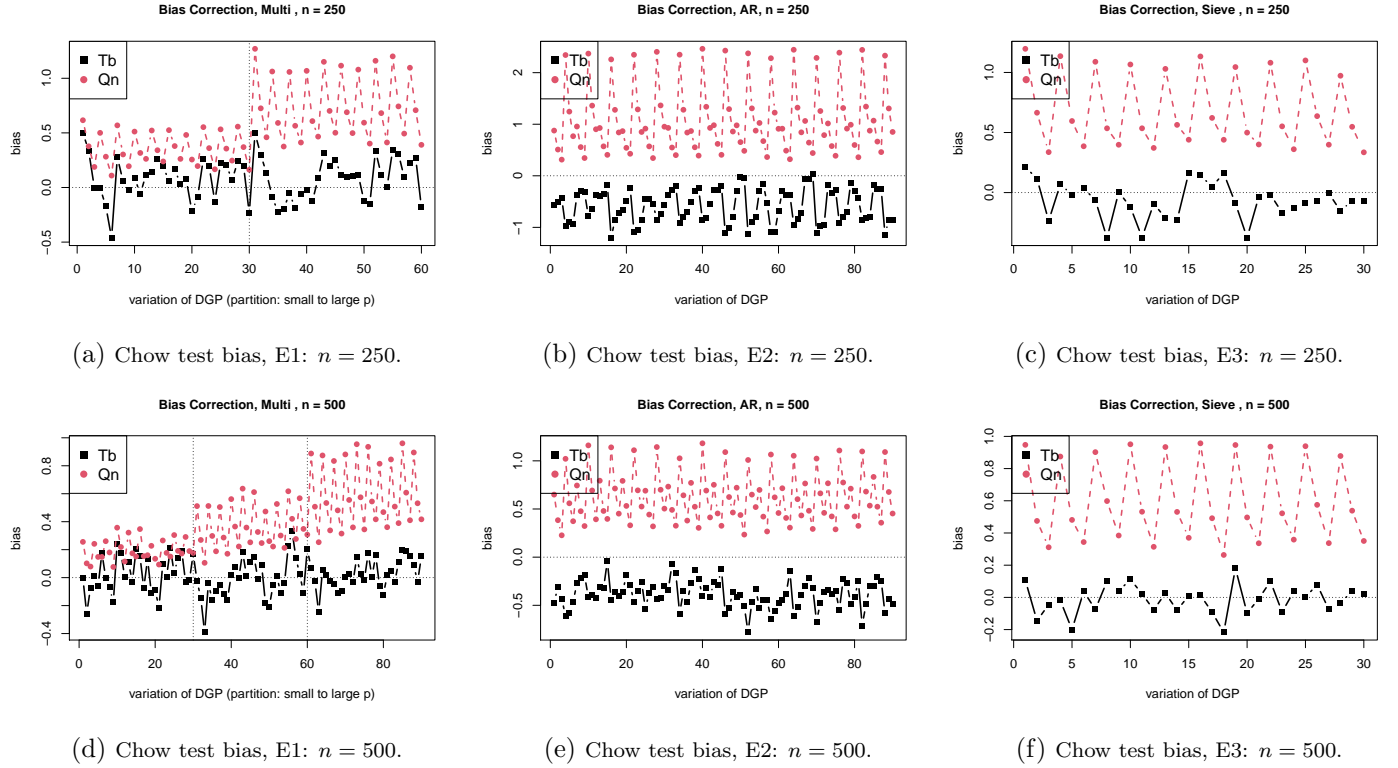


FIGURE 2. Bias in $\mathcal{Q}_n(\gamma)$ (dot) and $\mathcal{T}_n^b(\gamma)$ (square, black). For E1, the vertical partitions in (a) and (d) correspond to $p = 5, 9$ and $p = 5, 9, 13$, respectively. Within each vertical partition results are ordered lexicographically as $(\gamma \in \{0.2, 0.3, 0.5\}, \text{error} \in \{1, 2\}, \alpha \in \{0.3, 0.57\})$. E2 and E3: $p = 9$ for $n = 250$ and $p = 13$ for $n = 500$. Results horizontally ordered lexicographically as $(\gamma, \text{error}, \alpha \text{ or } \theta)$.

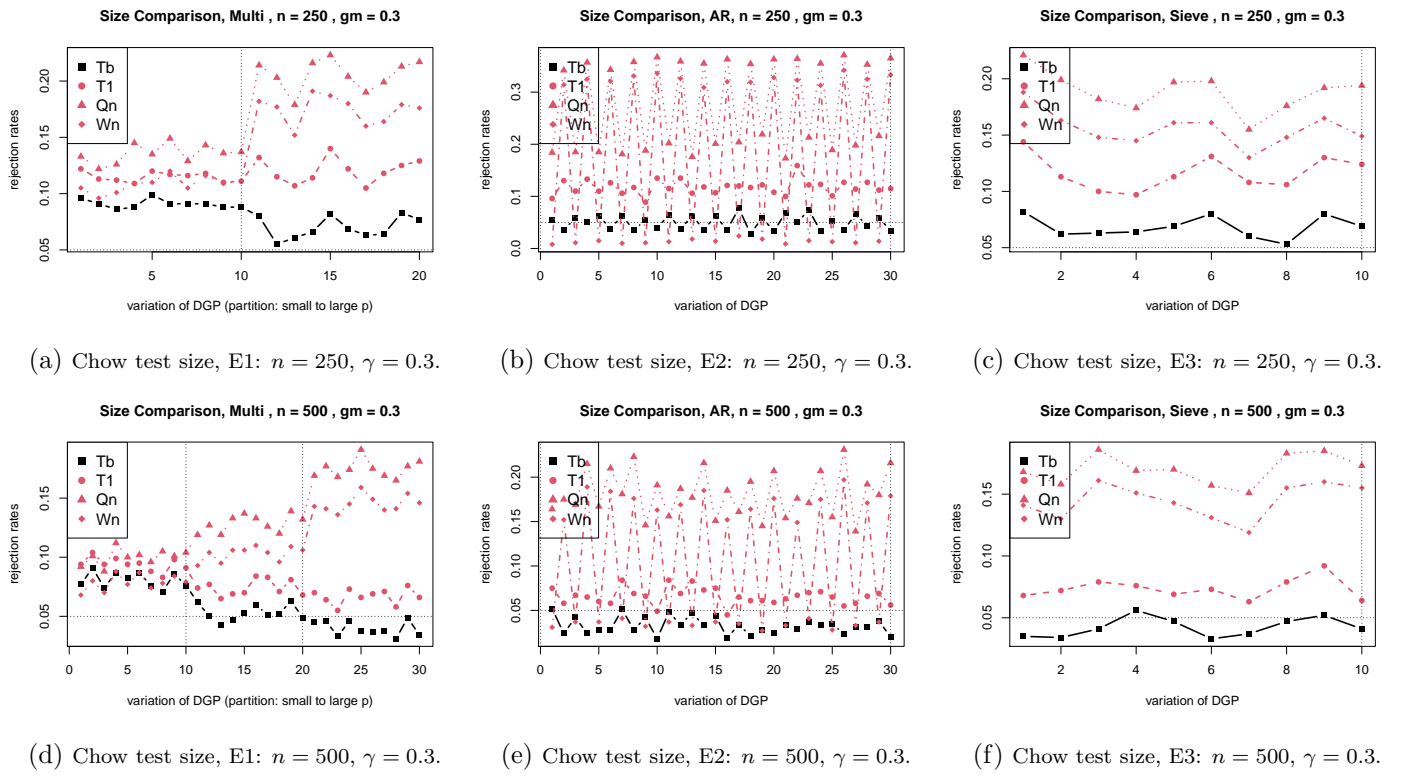


FIGURE 3. Size of Chow tests with $\gamma = 0.3$: $W_n(\gamma)$ (diamond), $Q_n(\gamma)$ (triangle), $T_n(\gamma)$ (dot) and $T_n^b(\gamma)$ (square, black). Nominal size is 5%. For E1, vertical partitions in (a) correspond to $n = 250$ and $p = 5, 9$ and those in (d) correspond to $n = 500$ and $p = 5, 9, 13$. Within each vertical partition results are ordered lexicographically as (error $\in \{1, 2\}, \alpha \in \{0.3, 0.57\}$). For E2, $p = 9$ for $n = 250$ and $p = 13$ for $n = 500$. Results horizontally ordered lexicographically as (error $\in \{1, 2\}, \alpha \in \{0.3, 0.57\}$ or $\theta \in \{-0.5, -0.1, 0.5\}$). For E3, $p = 9$ for $n = 250$ and $p = 13$ for $n = 500$. Results horizontally ordered lexicographically as (error $\in \{1, 2\}, \alpha \in \{0.3, \dots, 0.57\}$).

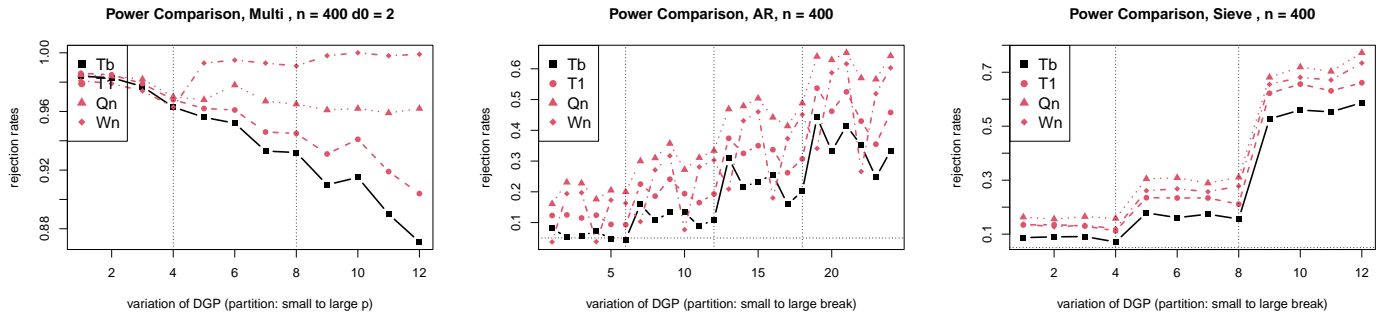


FIGURE 4. Power of Chow tests, $W_n(\gamma)$ (diamond), $Q_n(\gamma)$ (triangle), $T_n(\gamma)$ (dot) and $T_n^b(\gamma)$ (square, black): E1 (left), E2 (center) and E3 (right), $n = 400$, $\gamma = 0.5$. Vertical partitions correspond to $p = 5, 9, 13$ (left), $\theta = 0.2, 0.4, 0.6, 0.8$ (center) and $\delta = 0.5p^{1/4}/n^{1/2}(1, 5, 10)$ (right). Within each vertical partition results are ordered lexicographically as (error $\in \{1, 2, 3\}, \alpha \in \{0.3, 0.5\}$) for E1 and (error $\in \{1, 2, 3\}, \alpha \in \{0.3, 0.5\}$ or θ) for E2 and E3.

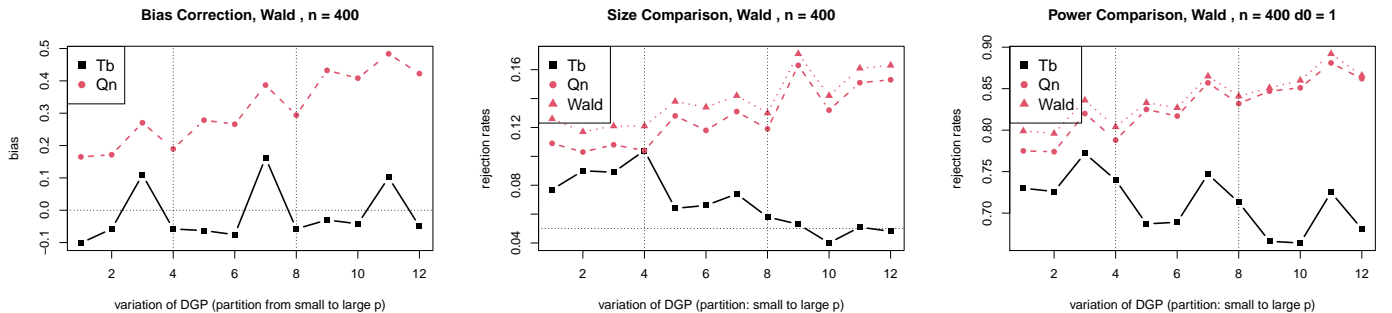


FIGURE 5. Bias, Size, and Power of Exclusion Tests in E1: Q_n^e (dot) and $T_n^{e,b}$ (square, black). Left to right: vertical partitions correspond to $p = 8, 12, 16$. Within each vertical partition results are ordered lexicographically as (error $\in \{1, 2\}, \alpha \in \{0.4, 0.55\}$).

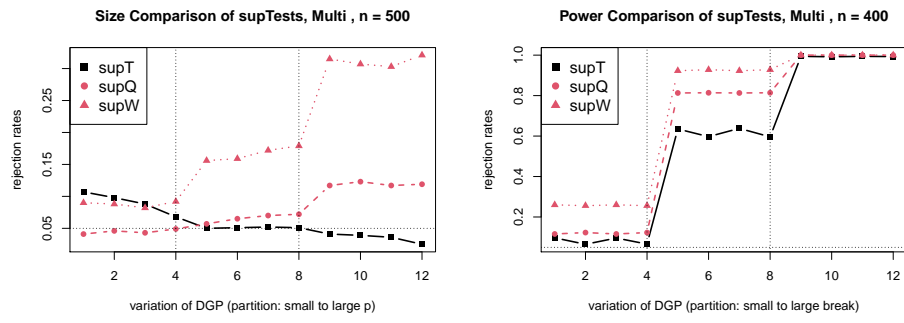


FIGURE 6. E1: Size and Power of Structural break sup tests, $\sup_{\gamma \in \Gamma} W_n(\gamma)$ (triangle), $\sup_{\gamma \in \Gamma} Q_n(\gamma)$ (dot) and $\sup_{\gamma \in \Gamma} T_n(\gamma)$ (square, black) in E1: $\Gamma = [0.15, 0.85]$. Nominal size is 5%. Vertical partitions correspond to $p = 5, 7, 9$ (left) and to break size $d_0 = 1, 5, 10$ (right). Within each vertical partition results are ordered lexicographically as (error $\in \{1, 2\}, \alpha \in \{0.4, 0.55\}$).

APPENDIX A. PROOFS OF THEOREMS

We begin with some notation. Let

$$A(\gamma) = (X^*(\gamma)' M_X X^*(\gamma))^{-1} X^*(\gamma)' M_X$$

with $X^*(\gamma)$ having t -th row $x_t^*(\gamma)' = x_t' 1\{t/n > \gamma\}$, M_X is the residual maker for the matrix X with t -th row x_t' , and

$$B(\gamma) = R M(\gamma)^{-1} \Omega(\gamma) M(\gamma)^{-1} R'.$$

Also, let $\bar{\Omega}(\gamma) = n^{-1} \sum_{t=1}^n x_t(\gamma) x_t'(\gamma) \sigma_t^2$ and $x_t(\gamma) = (x_t', x_t'^*(\gamma))'$. It is also convenient to recall that $\hat{M} = n^{-1} X' X$ and define $\hat{S}(\gamma) = n^{-1} X'^*(\gamma) X(\gamma)$. Recall that cross-referenced items prefixed with ‘S’ can be found in the online supplementary appendix.

A.1. Proofs for Section 3.

Proof of Theorem 3.1 and 3.2. These theorems are marginal convergences of the weak convergences established under \mathcal{H}_0 and \mathcal{H}_ℓ in the proof of Theorem 4.1. ■

A.2. Proof of Theorem 4.1: For the weak convergence under the null, Section A.2.1 first establishes the weak convergence of $\mathcal{Q}_n(\gamma)$, and then Section A.2.2 proves $\hat{\mathcal{V}} \xrightarrow{d} \int_0^1 \int_0^1 K_h^*(r, s) dW(r) dW(s)$, where $W(r)$ denotes the same limit Gaussian process as in Theorem ST.2. Then, the claim follows by Theorem ST.2 and the continuous mapping theorem. After completing the weak convergence under the null, we prove convergence under the local alternative in Section A.2.3.

A.2.1. Weak convergence of $\mathcal{Q}_n(\gamma)$ under \mathcal{H}_0 .

Proof. This step is quite involved and we delegate proofs of many intermediate steps to Section S.B. Summarizing these steps, Theorem ST.1 therein develops the initial approximation $\mathcal{Q}_n(\gamma) = (\mathcal{R}_n(\gamma) - p) / \sqrt{2p} + o_p(1)$, where $\mathcal{R}_n(\gamma)$ is defined in (3.4). Then, (SB.44) and Lemma SL.8 yield the second approximation

$$\frac{\mathcal{R}_n(\gamma) - p}{\sqrt{2p}} = \mathcal{S}_n(\gamma) + o_p(1),$$

uniformly in $\gamma \in \Gamma$, where

$$(A.1) \quad \mathcal{S}_n(\gamma) = \frac{n^{-1} \sum_{s \neq t} g_t(\gamma)' \Omega^{-1} g_s(\gamma) \varepsilon_t \varepsilon_s}{\gamma(1-\gamma) \sqrt{2p}},$$

and $g_t(\gamma) = x_t' 1\{t/n \leq \gamma\} - \gamma x_t$. The claim now follows by a functional CLT for $\mathcal{S}_n(\gamma)$ established in Theorem ST.2. ■

A.2.2. Weak convergence of $\hat{\mathcal{V}}$.

Proof. To establish this convergence, by Lemma 1 (c) of Sun (2014) it suffices to show that

$$(A.2) \quad \hat{\mathcal{V}} - \tilde{\mathcal{V}} = o_p(1),$$

where $\tilde{\mathcal{V}} = \frac{2}{n} \sum_{t=2}^n \sum_{s=2}^n k\left(\frac{t-s}{n/h}\right) \bar{q}_s^* \bar{q}_t^*$, $q_t^* = (np)^{-1/2} x_t' \Omega^{-1} \varepsilon_t \sum_{s=1}^{t-1} x_s \varepsilon_s$, $\bar{q}_t^* = q_t^* - n^{-1} \sum_{t=2}^n q_t^*$. Strictly speaking, Sun's Lemma 1 (c) is stated for the case where the partial sums of q_t are approximated by the partial sums of e_t , which is iid normal, but it also holds when it is approximated by the partial sums of $a_{nt}e_t$ for any real bounded array a_{nt} by repeating the same argument in the proof. In our case, $a_{nt} = \sqrt{t/n}$.

Let $\varsigma = n/h$, $\hat{\zeta}_t = \hat{h}_t' \sum_{s < t} \hat{h}_s / \sqrt{p}$, $\hat{h}_t = \hat{\Omega}^{-1} x_t \hat{e}_t$, $\bar{\zeta} / \sqrt{n} = n^{-1} \sum_{t=2}^n q_t = n^{-1} \sum_{t=2}^n \hat{\zeta}_t / \sqrt{n}$, with analogous definitions using Ω and ε_t for ζ_t , h_t and $\bar{\zeta}$. Then

$$\begin{aligned} \hat{\mathcal{V}} - \tilde{\mathcal{V}} &= n^{-2} \sum_{j=-(n-1)}^{n-1} k(j/\varsigma) n^{-1} \sum_{t=1+(j \vee 0)}^{n-|j \wedge 0|} \left\{ \left(\hat{\zeta}_t \hat{\zeta}_{t+|j|} - \zeta_t \zeta_{t+|j|} \right) + 2\bar{\zeta} \left(\hat{\zeta}_t - \zeta_t \right) \right. \\ (A.3) \quad &+ \left. \left(\bar{\zeta} - \bar{\zeta} \right) \hat{\zeta}_t + \left(\bar{\zeta}^2 - \bar{\zeta}^2 \right) \right\}. \end{aligned}$$

We obtain a bound for

$$(A.4) \quad \hat{\zeta}_t \hat{\zeta}_{t+|j|} - \zeta_t \zeta_{t+|j|} = \left(\hat{\zeta}_t - \zeta_t \right) \hat{\zeta}_{t+|j|} + \left(\hat{\zeta}_{t+|j|} - \zeta_{t+|j|} \right) \hat{\zeta}_t,$$

while omitting similar details for the other three terms. To find a bound for (A.4), first note that $\hat{h}_t = O_p(\|x_t\|) = O_p(\sqrt{p})$, by Assumption 3(ii) and finite fourth moments of x_t components (Assumption 3(i)), and because

$$(A.5) \quad \hat{e}_t = y_t - x_t' \hat{\delta}_1(\gamma) = x_t' \left(\hat{\delta}_1(\gamma) - \delta_1 \right) + x_t' 1(t/n > \gamma) \delta_{2\ell} + r_t + \varepsilon_t = O_p(1).$$

Hence

$$(A.6) \quad \hat{\zeta}_t = \hat{h}_t' \sum_{s < t} \hat{h}_s / \sqrt{p} = O_p(n\sqrt{p}).$$

By the same argument, $h_t = O_p(\sqrt{p})$ and $\zeta_t = O_p(n\sqrt{p})$ as well.

Now recall that $\hat{e}_t - \varepsilon_t = x_t' \left(\hat{\delta}_1(\gamma) - \delta_1 \right) + x_t' 1(t/n > \gamma) \delta_{2\ell} + r_t$ and $\left\| \hat{\delta}_1(\gamma) - \delta_1 \right\| = O_p\left(\left\| \hat{\delta}(\gamma) - \delta \right\|\right) = O_p\left(\lambda_n^{-1} \sqrt{p}/\sqrt{n}\right)$ implying that

$$(A.7) \quad \hat{e}_t - \varepsilon_t = O_p\left(\max\left\{\lambda_n^{-1} p/\sqrt{n}, p^{3/4}/\sqrt{n}\right\}\right).$$

Thus we obtain

$$(A.8) \quad \hat{h}_t - h_t = \Omega^{-1} \left(\Omega - \hat{\Omega} \right) \hat{\Omega}^{-1} x_t \hat{e}_t + \Omega^{-1} x_t (\hat{e}_t - \varepsilon_t) = O_p\left(\lambda_n^{-2} \sqrt{p} \max\{v_p, p/\sqrt{n}\}\right),$$

using Assumption 3(iii). Using (A.8), we get

$$(A.9) \quad \hat{\zeta}_t - \zeta_t = \left(\hat{h}_t - h_t \right)' \sum_{s < t} \hat{h}_s / \sqrt{p} + \hat{h}_t' \sum_{s < t} \left(\hat{h}_s - h_s \right) / \sqrt{p} = O_p\left(\lambda_n^{-2} n \sqrt{p} \max\{v_p, p/\sqrt{n}\}\right).$$

Using (A.6) and (A.9) in (A.4), we obtain $\hat{\zeta}_t \hat{\zeta}_{t+|j|} - \zeta_t \zeta_{t+|j|} = O_p\left(\lambda_n^{-2} n^2 p \max\{v_p, p/\sqrt{n}\}\right)$. This, along with similarly obtained bounds for

the remaining terms in (A.3) and Lemma 1 of Jansson (2002), yield

$$\hat{\mathcal{V}} - \tilde{\mathcal{V}} = O_p \left(\varsigma \left(\int_{\mathbb{R}} |k(x)| dx \right) \lambda_n^{-2} p \max \{v_p, p/\sqrt{n}\} \right) = O_p \left(\lambda_n^{-2} \max \{pv_p, p^2/\sqrt{n}\} \right),$$

which is negligible by (4.4). ■

A.2.3. *Under the alternative.*

Proof. Under \mathcal{H}_ℓ , we have $\hat{\delta}_2(\gamma) = A(\gamma)X^*(\gamma_0)\delta_{2\ell} + A(\gamma)\varepsilon + A(\gamma)r$, so that, writing $D(\gamma, \gamma_0) = A(\gamma)X^*(\gamma_0)$, similar algebra to that used in the online appendix and Lemmas SL.6-SL.10 yields

$$\begin{aligned} \mathcal{Q}_n(\gamma) &= \mathcal{S}_n(\gamma) + \frac{2n\delta'_{2\ell}D(\gamma, \gamma_0)' \hat{B}(\gamma)^{-1}A(\gamma)\varepsilon}{\sqrt{2p}} + \frac{2n\delta'_{2\ell}D(\gamma, \gamma_0)' \hat{B}(\gamma)^{-1}A(\gamma)r}{\sqrt{2p}} \\ (A.10) \quad &+ \frac{n\delta'_{2\ell}D(\gamma, \gamma_0)' \left(\hat{B}(\gamma)^{-1} - B(\gamma)^{-1} \right) D(\gamma, \gamma_0)\delta_{2\ell}}{\sqrt{2p}} \\ &+ \frac{n\delta'_{2\ell}D(\gamma, \gamma_0)' \hat{B}(\gamma)^{-1}D(\gamma, \gamma_0)\delta_{2\ell}}{\sqrt{2p}} + o_p(1). \end{aligned}$$

For the second term on the RHS of (A.10), note that this equals

$$\begin{aligned} &\frac{2n\delta'_{2\ell}D(\gamma, \gamma_0)' B(\gamma)^{-1}A(\gamma)\varepsilon}{\sqrt{2p}} + \frac{2n\delta'_{2\ell}D(\gamma, \gamma_0)' \left(\hat{B}(\gamma)^{-1} - B(\gamma)^{-1} \right) A(\gamma)\varepsilon}{\sqrt{2p}} \\ &= \frac{2n\delta'_{2\ell}D(\gamma, \gamma_0)' B(\gamma)^{-1}A(\gamma)\varepsilon}{\sqrt{2p}} + O_p \left(\lambda_n^{-2} n \|\delta_{2\ell}\| \|n^{-1}X'\varepsilon\| \left\| \hat{B}(\gamma) - B(\gamma) \right\| / \sqrt{p} \right) \\ &= \frac{2n\delta'_{2\ell}D(\gamma, \gamma_0)' B(\gamma)^{-1}A(\gamma)\varepsilon}{\sqrt{2p}} + O_p \left(\lambda_n^{-4} p^{1/4} \max \{ \lambda_n^{-1} \varkappa_p, v_p \} \right), \end{aligned}$$

proceeding like (SB.38), the second stochastic order above being negligible by (3.1). By Assumption 1, the first term has mean zero and variance equal to a constant times

$$\frac{\tau' D(\gamma, \gamma_0)' B(\gamma)^{-1} A(\gamma) A(\gamma)' B(\gamma)^{-1} D(\gamma, \gamma_0) \tau}{\sqrt{p}} = O_p(1/\sqrt{p}),$$

uniformly in γ by Lemmas SL.4 and SL.5 and the calculations therein.

By Assumption 2, the third term on the RHS of (A.10) is

$$O_p \left(n \|\delta_{2\ell}\| \|n^{-1}X'r\| / \sqrt{p} \right) = O_p \left(p^{-1/4} \right).$$

The fourth term on the RHS of (A.10) is readily seen to be $O_p \left(\left\| \hat{B}(\gamma)^{-1} - B(\gamma)^{-1} \right\| \right) = O_p \left(\lambda_n^{-4} \max \{ \lambda_n^{-1} \varkappa_p, v_p \} \right)$, which is negligible by (3.1). Thus, by (SB.28), (A.10) becomes

$$\begin{aligned} \mathcal{Q}_n(\gamma) &= \mathcal{S}_n(\gamma) + \frac{n\delta'_{2\ell}D(\gamma, \gamma_0)' B(\gamma)^{-1}D(\gamma, \gamma_0)\delta_{2\ell}}{\sqrt{2p}} + o_p(1) \\ &= \mathcal{S}_n(\gamma) + \gamma(1-\gamma)\tau' D(\gamma, \gamma_0)' M\Omega^{-1}MD(\gamma, \gamma_0)\tau + o_p(1). \end{aligned}$$

Now, by the definition of its components and steps similar to those elsewhere in the paper, it is readily seen that $\|D(\gamma, \gamma_0) - \{(\gamma + \gamma_0(1 - \gamma) - (\gamma \vee \gamma_0))/\gamma(1 - \gamma)\} I_p\| = o_p(1)$, uniformly on Γ and that $\gamma + \gamma_0(1 - \gamma) - (\gamma \vee \gamma_0) = -\gamma\gamma_0 + (\gamma \wedge \gamma_0)$ as $\gamma + \gamma_0 - (\gamma \vee \gamma_0) = (\gamma \wedge \gamma_0)$. Thus,

$$(A.11) \quad \mathcal{Q}_n(\gamma) \Rightarrow Q(\gamma) + \frac{(\gamma\gamma_0 - (\gamma \wedge \gamma_0))^2}{\gamma(1 - \gamma)} \lim_{n \rightarrow \infty} \tau' M \Omega^{-1} M \tau,$$

on Γ by Theorem ST.2, which gives the distribution of $\mathcal{Q}_n(\gamma)$ under \mathcal{H}_ℓ . ■

Proof of Theorem 4.2. In Section S.C of the online supplement. ■

A.3. Proofs for Section 5.

Proof of Theorem 5.1. Follows by Theorem 4.1 and the continuous mapping theorem. ■

Proof of Theorem 5.2. The proof proceeds exactly as that of Theorem 4.1, but is simpler due to the absence of γ . We give a brief summary and omit the details. Because $R^e \hat{\beta} - r = Rn^{-1} \hat{M}^{-1} \sum_{t=1}^n x_t \varepsilon_t$ under \mathcal{H}_0^e , we can obtain the approximation

$$\mathcal{Q}_n^e = \frac{n^{-1} (\sum_{t=1}^n x_t \varepsilon_t)' L (\sum_{t=1}^n x_t \varepsilon_t) - p}{\sqrt{2p}} + o_p(1).$$

Then, the proof of asymptotic normality follows with $w_{ns} = \xi'_s \sum_{t=1}^{s-1} \xi_t / \sqrt{np}$ as in Theorem ST.2, but now defining $\xi_t = L^{1/2} x_t \varepsilon_t$. From this it is readily seen that $E(\sum_{s=1}^n w_{ns})^2 = \mathcal{V}^e + o(1)$. ■

Online supplement to “Robust Inference on Infinite and Growing Dimensional Time Series Regression”

Abhimanyu Gupta and Myung Hwan ‘Matt’ Seo

S.A. AN EXPONENTIAL INEQUALITY FOR PARTIAL SUMS OF WEAKLY DEPENDENT RANDOM MATRICES

We develop a stochastic order for a matrix partial sum. Closely related results can be found in Theorems 4.1 and 4.2 of Chen and Christensen (2015), who establish such bounds for full matrix sums as opposed to partial sums. Our first theorem is a Fuk-Nagaev type inequality, using a coupling approach similar to Dedecker and Prieur (2004), Chen and Christensen (2015) and Rio (2017).

Theorem ST.1. *Let $\{\xi_i\}_{i \in \mathbb{Z}}$ be a β -mixing sequence with support \mathfrak{X} and r -th mixing coefficient $\beta(r)$ and let $\Xi_{i,n} = \Xi_n(\xi_i)$, for each i , where $\Xi_n : \mathfrak{X} \rightarrow \mathbb{R}^{d_1 \times d_2}$ is a sequence of measurable $d_1 \times d_2$ matrix-valued functions. Assume $E(\Xi_{i,n}) = 0$ and $\|\Xi_{i,n}\| \leq R_n$, for each i , set*

$$s_n^2 = \max_{1 \leq i, j \leq n} \max \{ \|E(\Xi_{i,n} \Xi'_{j,n})\|, \|E(\Xi'_{i,n} \Xi_{j,n})\| \},$$

and define $S_k = \sum_{l=1}^k \Xi_{l,n}$. Then, for any integer q such that $1 < q \leq n/2$ and $\varrho \geq qR_n$,

$$P \left(\sup_{1 \leq k \leq n} \|S_k\| > 4\varrho \right) \leq \left(\left\lceil \frac{n}{q} \right\rceil + 1 \right) \beta(q) + 2(d_1 + d_2) \exp \left(\frac{-\varrho^2/2}{nqs_n^2 + qR_n\varrho/3} \right).$$

The required stochastic order now follows by a choice of ϱ in Theorem ST.1:

Corollary SC.1. *Under the conditions of Theorem ST.1, if q is chosen as a function of n such that $(n/q)\beta(q) = o(1)$ and $R_n\sqrt{q \log(d_1 + d_2)} = o(s_n\sqrt{n})$ then*

$$\sup_{1 \leq k \leq n} \|S_k\| = O_p \left(s_n \sqrt{nq \log(d_1 + d_2)} \right)$$

Proof of Theorem ST.1. For $i = 1, \dots, [n/q]$, define $U_i = \sum_{j=iq-q+1}^{iq} \Xi_{j,n}$ and $U_{[n/q]+1} = \sum_{j=[n/q]q}^n \Xi_{j,n}$. Now, for an integer j that differs from an integer multiple of q by at most $[q/2]$, we have $\sup_{1 \leq k \leq n} \|S_k\| \leq 2[q/2]R_n + \sup_{j>0} \left\| \sum_{i=1}^j U_i \right\|$. If q is even (respectively odd) then $q = 2k$ (resp. $q = 2k + 1$) for some positive integer k , implying $[q/2] = [2k/2] = k$ (resp. $[q/2] = [(2k+1)/2] = k$) whence $2[q/2]R_n \leq qR_n$ (resp. $2[q/2]R_n \leq (q-1)R_n$). Thus, because $\varrho \geq qR_n$,

$$\begin{aligned} P \left(\sup_{1 \leq k \leq n} \|S_k\| > 4\varrho \right) &\leq P(2[q/2]R_n > \varrho) + P \left(\sup_{j>0} \left\| \sum_{i=1}^j U_i \right\| \geq 3\varrho \right) \\ (SA.1) \qquad &= P \left(\sup_{j>0} \left\| \sum_{i=1}^j U_i \right\| \geq 3\varrho \right), \end{aligned}$$

so it suffices to prove that

$$P \left(\sup_{j>0} \left\| \sum_{i=1}^j U_i \right\| \geq 3\varrho \right) \leq \left(\left\lfloor \frac{n}{q} \right\rfloor + 1 \right) \beta(q) + 2(d_1 + d_2) \exp \left(\frac{-\varrho^2/2}{nqs_n^2 + qR_n\varrho/3} \right).$$

Enlarging the probability space as needed, by Lemma 5.1 (Berbee's Lemma) of Rio (2017) there is a sequence ξ_i^* , $1 \leq i \leq \lfloor n/q \rfloor + 1$, such that

- (a) The random variable x_i^* is distributed as x_i for each $1 \leq i \leq \lfloor n/q \rfloor + 1$.
- (b) The sequences ξ_{2i}^* , $1 \leq 2i \leq \lfloor n/q \rfloor + 1$, and ξ_{2i-1}^* , $1 \leq 2i-1 \leq \lfloor n/q \rfloor + 1$, comprise of independent random variables.
- (c) $P(\xi_i \neq \xi_i^*) \leq \beta(q+p)$ for $1 \leq i \leq \lfloor n/q \rfloor + 1$.

Denote $\Xi_{i,n}^* = \Xi_n(\xi_i^*)$, and define U_i^* in the obvious manner. Then, we have

$$(SA.2) \quad \sup_{j>0} \left\| \sum_{i=1}^j U_i \right\| \leq \sum_{i=1}^{\lfloor n/q \rfloor + 1} \|U_i - U_i^*\| + \sup_{j>0} \left\| \sum_{i=1}^j U_{2i}^* \right\| + \sup_{j>0} \left\| \sum_{i=1}^j U_{2i-1}^* \right\|.$$

Now, by (c), we have

$$\begin{aligned} P \left(\sum_{i=1}^{\lfloor n/q \rfloor + 1} \|U_i - U_i^*\| \geq \varrho \right) &= P \left(\sum_{i=1}^{\lfloor n/q \rfloor + 1} \|U_i - U_i^*\| \geq \left\lfloor \sum_{i=1}^{\lfloor n/q \rfloor + 1} \varrho / (\lfloor n/q \rfloor + 1) \right\rfloor \right) \\ &\leq \sum_{i=1}^{\lfloor n/q \rfloor + 1} P(\|U_i - U_i^*\| \geq \varrho / (\lfloor n/q \rfloor + 1)) \\ &\leq (\lfloor n/q \rfloor + 1) \beta(q+p), \end{aligned}$$

while for all $1 \leq i \leq \lfloor n/q \rfloor + 1$ the matrices $U_i^* = \sum_{j=iq-q+1}^{iq} \Xi_{j,n}^*$ satisfy $\|U_i^*\| \leq qR_n$ and

$$\max_{1 \leq j \leq n} \max \left\{ \left\| E \left(\sum_{i=1}^j U_i U_i^{*'} \right) \right\|, \left\| E \left(\sum_{i=1}^j U_i^{*'} U_i^* \right) \right\| \right\} \leq nqs_n^2.$$

Furthermore, the sequence $\mathcal{U}_j = \sum_{i=1}^j U_{2i}^*$ is a matrix martingale (because U_{2i}^* is an independent sequence and $E\mathcal{U}_j = 0$) with difference sequence $\mathcal{U}_j - \mathcal{U}_{j-1} = U_{2j}^*$. Thus, by Corollary 1.3 of Tropp (2011),

$$(SA.3) \quad P \left(\sup_{j>0} \left\| \sum_{i=1}^j U_{2i}^* \right\| \geq \varrho \right) \leq (d_1 + d_2) \exp \left(\frac{-\varrho^2/2}{nqs_n^2 + qR_n\varrho/3} \right).$$

The third term on the RHS of (SA.2) is bounded similarly, whence the claim follows. ■

Proof of Corollary SC.1. In Theorem ST.1, take $\varrho = Cs_n \sqrt{nq \log(d_1 + d_2)}$ for a sufficiently large constant C . Then the claim follows by the condition $(n/q) \beta(q) = o(1)$ and because $R_n \sqrt{q \log(d_1 + d_2)} = o(s_n \sqrt{n})$. To verify that ϱ satisfies that requirement of Theorem ST.1, note that the latter condition implies $Cs_n \sqrt{n} \geq$

$R_n \sqrt{q \log(d_1 + d_2)}$ for sufficiently large n , so $\varrho \geq qR_n \log(d_1 + d_2) \geq qR_n$ for sufficiently large n , assuming $d_1 + d_2 \geq e \approx 2.72$. The latter condition fails only if the $\Xi_{i,n}$ are scalar. ■

S.B. FOR SECTION 3

We first present an initial approximation of $\mathcal{Q}_n(\gamma)$.

Theorem ST.1. *Let Assumptions 1-3 hold, and*

$$(SB.1) \quad \lambda_n^{-4} \sqrt{p} (\lambda_n^{-1} \varkappa_p + v_p) + \lambda_n^{-4} p^{-1} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

Then, $\sup_{\gamma \in \Gamma} |\mathcal{Q}_n(\gamma) - (\mathcal{R}_n(\gamma) - p) / \sqrt{2p}| = o_p(1)$.

Proof. Much of the details are delegated to Lemmas SL.2-SL.10. In particular, we show in Lemma SL.6 that

$$(SB.2) \quad \mathcal{Q}_n(\gamma) = \frac{n\varepsilon' A(\gamma)' B(\gamma)^{-1} A(\gamma) \varepsilon - p}{\sqrt{2p}} + o_p(1)$$

uniformly in γ .

Then, note that

$$(SB.3) \quad (X^*(\gamma)' M_X X^*(\gamma))^{-1} = n^{-1} \left(I - \hat{M}^{-1} \hat{S}(\gamma) \right)^{-1} \hat{S}(\gamma)^{-1},$$

and

$$(SB.4) \quad X^*(\gamma)' M_X \varepsilon = X^*(\gamma)' \varepsilon - \hat{S}(\gamma) \hat{M}^{-1} X' \varepsilon,$$

because $n^{-1} X^*(\gamma)' X = \hat{S}(\gamma)$. Using (SB.3) and (SB.4), we may write $n\varepsilon' A(\gamma)' B(\gamma)^{-1} A(\gamma) \varepsilon / \sqrt{2p}$ as

$$(SB.5) \quad \frac{n^{-1} R_1(\gamma)' R_2(\gamma)' B(\gamma)^{-1} R_2(\gamma) R_1(\gamma)}{\sqrt{2p}}$$

where $R_1(\gamma) = \hat{S}(\gamma)^{-1} X'^*(\gamma) \varepsilon - \gamma(1 - \gamma)^{-1} \hat{M}^{-1} X' \varepsilon$ and $R_2(\gamma) = \left(I - \hat{M}^{-1} \hat{S}(\gamma) \right)^{-1}$.

By adding and subtracting terms we can decompose (SB.5) as $\sum_{i=1}^4 \Delta_i(\gamma) + \overline{\mathcal{R}}_n(\gamma)$, with

$$\begin{aligned} \Delta_1(\gamma) &= \frac{(R_1(\gamma) - \overline{R}_1(\gamma))' R_2(\gamma)' B(\gamma)^{-1} R_2(\gamma) R_1(\gamma)}{n\sqrt{2p}}, \\ \Delta_2(\gamma) &= \frac{\overline{R}_1(\gamma)' R_2(\gamma)' B(\gamma)^{-1} R_2(\gamma) (R_1(\gamma) - \overline{R}_1(\gamma))}{n\sqrt{2p}}, \\ \Delta_3(\gamma) &= \frac{\overline{R}_1(\gamma)' (R_2(\gamma) - \gamma^{-1} I)' B(\gamma)^{-1} R_2(\gamma) \overline{R}_1(\gamma)}{n\sqrt{2p}}, \\ \Delta_4(\gamma) &= \frac{\overline{R}_1(\gamma)' B(\gamma)^{-1} (R_2(\gamma) - \gamma^{-1} I) \overline{R}_1(\gamma)}{\gamma n\sqrt{2p}}, \end{aligned}$$

where we write $\bar{R}_1(\gamma) = (1 - \gamma)^{-1} \hat{M}^{-1} \left(\gamma \sum_{t=1}^n \varepsilon_t x_t - \sum_{t=1}^{[n\gamma]} \varepsilon_t x_t \right)$ and
(SB.6)

$$\bar{\mathcal{R}}_n(\gamma) = \frac{\left(\sum_{t=1}^{[n\gamma]} \varepsilon_t x_t - \gamma \sum_{t=1}^n \varepsilon_t x_t \right)' \hat{M}^{-1} B(\gamma)^{-1} \hat{M}^{-1} \left(\sum_{t=1}^{[n\gamma]} \varepsilon_t x_t - \gamma \sum_{t=1}^n \varepsilon_t x_t \right)}{\gamma^2 (1 - \gamma)^2 n \sqrt{2p}}.$$

By (SB.28), the term sandwiched between the parentheses in the numerator of (SB.6) is

$$(SB.7) \quad \left(\hat{M}^{-1} - M^{-1} \right) B(\gamma)^{-1} \hat{M}^{-1} + M^{-1} B(\gamma)^{-1} \left(\hat{M}^{-1} - M^{-1} \right) + \gamma(1 - \gamma) \Omega^{-1}.$$

Substituting (SB.7) into (SB.6) yields three terms corresponding to the three terms in (SB.7). The first of these, multiplied by the outside terms in the sandwich formula in (SB.6), has modulus bounded by a constant times

$$\sup_{\gamma \in \Gamma} \frac{n^{-1} (\|X' \varepsilon\|^2 + \|X^*(\gamma)' \varepsilon\|^2) \left\| \hat{M} - M \right\| \|B(\gamma)^{-1}\| \|M^{-1}\| \left\| \hat{M}^{-1} \right\|^2}{\sqrt{p}} = O_p \left(\lambda_n^{-4} \sqrt{p} \varkappa_p \right),$$

by Assumption 3 and Lemmas SL.2, SL.4, and also (SB.24), while the second is similarly shown to be uniformly $O_p(\sqrt{p} \varkappa_p)$ also. By (SB.1), we conclude that

$$\bar{\mathcal{R}}_n(\gamma) = \mathcal{R}_n(\gamma) + o_p(1),$$

indicating that the theorem is proved if $\Delta_i(\gamma) = o_p(1)$, $i = 1, 2, 3, 4$. But by previously used techniques and Lemmas SL.9 and SL.10, we readily conclude that

$$\sup_{\gamma \in \Gamma} (\Delta_1(\gamma), \Delta_2(\gamma), \Delta_3(\gamma), \Delta_4(\gamma)) = O_p \left(\lambda_n^{-5} \sqrt{p} \varkappa_p \right)$$

which are all negligible by (SB.1), proving the theorem. ■

Write $\tilde{\Omega}(\gamma) = n^{-1} \sum_{t=1}^n x_t(\gamma) x_t'(\gamma) \varepsilon_t^2$.

Lemma SL.1. *Under Assumptions 1-3, and the conditions of Propositions SP.2 or SP.1 as applicable,*

$$(SB.8) \quad \sup_{\gamma \in \Gamma} \left\| \hat{\Omega}(\gamma) - \tilde{\Omega}(\gamma) \right\| = O_p \left(\lambda_n^{-2} \min \left\{ \frac{p^3}{n}, \frac{\vartheta_p^2 p}{n} \right\} \right),$$

$$(SB.9) \quad \sup_{\gamma \in \Gamma} \left\| \tilde{\Omega}(\gamma) - \bar{\Omega}(\gamma) \right\| = O_p \left(\frac{p}{\sqrt{n}} \right).$$

Proof of Lemma SL.1. The matrix inside the norm on the LHS of (SB.8) can be decomposed as $\sum_{i=1}^5 U_i(\gamma)$, with

$$U_1(\gamma) = n^{-1} \sum_{t=1}^n x_t(\gamma) x_t'(\gamma) \left[x_t'(\gamma) \left(\delta - \hat{\delta}(\gamma) \right) \right]^2,$$

$$U_2(\gamma) = n^{-1} \sum_{t=1}^n x_t(\gamma) x_t'(\gamma) r_t^2,$$

$$\begin{aligned}
U_3(\gamma) &= 2n^{-1} \sum_{t=1}^n x_t(\gamma) x'_t(\gamma) \left[x'_t(\gamma) (\delta - \hat{\delta}(\gamma)) \right] \varepsilon_t, \\
U_4(\gamma) &= 2n^{-1} \sum_{t=1}^n x_t(\gamma) x'_t(\gamma) \left[x'_t(\gamma) (\delta - \hat{\delta}(\gamma)) \right] r_t, \\
U_5(\gamma) &= n^{-1} \sum_{t=1}^n x_t(\gamma) x'_t(\gamma) r_t \varepsilon_t.
\end{aligned}$$

Recall Lemma SL.3 for $\sup_{\gamma \in \Gamma} \|\delta - \hat{\delta}(\gamma)\| = O_p(\lambda_n^{-1} \sqrt{p/n})$. Now, since the maximum eigenvalue of a non-negative definite symmetric matrix is less than equal to the trace,

$$\begin{aligned}
\|U_1(\gamma)\| &\leq n^{-1} \sum_{t=1}^n (x'_t(\gamma) x_t(\gamma))^2 (\delta - \hat{\delta}(\gamma))' (\delta - \hat{\delta}(\gamma)) \\
&\leq 2pn^{-1} \sum_{t=1}^n \sum_{j=1}^p x_{tj}^4 \|\delta - \hat{\delta}(\gamma)\|^2 = O_p(\lambda_n^{-2} p^2) O_p(p/n),
\end{aligned}$$

uniformly in γ , by the fact that $\sup_{t,j} E x_{tj}^4 < \infty$ and (SB.26). In a similar fashion,

$$E \sup_{\gamma \in \Gamma} \|U_2(\gamma)\| \leq 2En^{-1} \sum_{t=1}^n x'_t x_t r_t^2 \leq 2 \left(E (x'_t x_t)^2 E r_t^4 \right)^{1/2} = O(\lambda_n^{-2} p / \sqrt{n}).$$

Similarly and using the fact that $E(|\varepsilon_t| |x_t|) \leq \sqrt{E(\varepsilon_t^2 |x_t|)} = O(1)$, we obtain

$$\begin{aligned}
\|U_3(\gamma)\| &\leq 4n^{-1} \sum_{t=1}^n (x'_t x_t)^2 |\varepsilon_t| \|\delta - \hat{\delta}(\gamma)\|^2 = O_p(\lambda_n^{-2} p^3 / n), \\
\|U_4(\gamma)\| &\leq 4n^{-1} \sum_{t=1}^n (x'_t x_t)^{3/2} \|\delta - \hat{\delta}(\gamma)\| |r_t| \\
&\leq 4 \|\delta - \hat{\delta}(\gamma)\| \left(n^{-1} \sum_{t=1}^n (x'_t x_t)^2 \right)^{3/4} \left(n^{-1} \sum_{t=1}^n r_t^4 \right)^{1/4} \leq O\left(\sqrt{\frac{p}{n}} p^{3/2} \frac{\lambda_n^{-1}}{n^{1/4}}\right), \\
\|U_5(\gamma)\| &= 2n^{-1} \sum_{t=1}^n (x'_t x_t) |r_t \varepsilon_t| = O_p(p / \sqrt{n}),
\end{aligned}$$

all uniformly in Γ . Thus (SB.8) is established.

To show (SB.9), let x_{it} , $i = 1, \dots, p$, be a typical element of x_t . Then any element of $\tilde{\Omega}(\gamma) - \bar{\Omega}(\gamma)$ is of the form $n^{-1} \sum_{t=1}^n x_{it}(\gamma) x_{jt}(\gamma) (\varepsilon_t^2 - \sigma_t^2)$, $i, j = 1, \dots, p$, and $\varepsilon_t^2 - \sigma_t^2$ is an MDS by construction. Thus, it has mean zero and variance $n^{-2} \sum_{t=1}^n E x_{it}^2(\gamma) x_{jt}^2(\gamma) E((\varepsilon_t^2 - \sigma_t^2)^2 | \mathcal{F}_{t-1}) = O_p(n^{-1})$, by Assumption 1 and the boundedness of $E x_{it}^4$. Thus, $E \|\tilde{\Omega}(\gamma) - \bar{\Omega}(\gamma)\|^2 = O(p^2/n)$, and the claim in (SB.9) follows by Markov's inequality. ■

We establish the weak convergence of

$$(SB.10) \quad \mathcal{S}_n(\gamma) = \frac{n^{-1} \sum_{s \neq t} g_t(\gamma)' \Omega^{-1} g_s(\gamma) \varepsilon_t \varepsilon_s}{\gamma(1-\gamma) \sqrt{2p}},$$

as a process indexed by γ on any compact subset Γ in $(0, 1)$ under the uniform metric. Let ‘ \Rightarrow ’ denote weak convergence in $\ell^\infty(\Gamma)$.

Theorem ST.2. *Under Assumptions 1-5 and (SB.1), $\mathcal{S}_n(\gamma) \Rightarrow \sqrt{\mathcal{V}} \mathcal{Q}(\gamma)$, as $n \rightarrow \infty$, on any compact subset Γ in $(0, 1)$.*

Proof of Theorem ST.2. First, note that $\mathcal{S}_n(\gamma)$ equals $[\gamma(1-\gamma) \sqrt{2p}]^{-1}$ times

$$\frac{1}{n} \sum_{\substack{s,t=1 \\ s \neq t}}^{[n\gamma]} x_t' \Omega^{-1} x_s \varepsilon_t \varepsilon_s - \frac{2\gamma}{n} \sum_{s=1}^n \sum_{\substack{t=1 \\ s \neq t}}^{[n\gamma]} x_t' \Omega^{-1} x_s \varepsilon_t \varepsilon_s + \frac{\gamma^2}{n} \sum_{\substack{s,t=1 \\ s \neq t}}^n x_t' \Omega^{-1} x_s \varepsilon_t \varepsilon_s$$

and thus

$$\begin{aligned} \mathcal{S}_n(\gamma) &= \frac{\sqrt{2}}{\gamma(1-\gamma)} [\mathcal{A}_n(\gamma) - \gamma [\mathcal{A}_n(1) + \mathcal{A}_n(\gamma) - \bar{\mathcal{A}}_n(\gamma)] + \gamma^2 \mathcal{A}_n(1)], \\ &= \sqrt{2} \left(\frac{\mathcal{A}_n(\gamma)}{\gamma} + \frac{\bar{\mathcal{A}}_n(\gamma)}{(1-\gamma)} - \mathcal{A}_n(1) \right), \end{aligned}$$

where

$$\begin{aligned} \mathcal{A}_n(\gamma) &= \frac{1}{n\sqrt{p}} \sum_{s=2}^{[n\gamma]} \sum_{t=1}^{s-1} \xi_t' \xi_s, \\ \bar{\mathcal{A}}_n(\gamma) &= \frac{1}{n\sqrt{p}} \sum_{s=[n\gamma]+1}^n \sum_{t=[n\gamma]+1}^{s-1} \xi_t' \xi_s, \end{aligned}$$

and $\xi_t = \{\xi_{ti}\}_{i=1}^p = \Omega^{-1/2} x_t \varepsilon_t$ being an mds.

Due to their symmetric nature, the tightness proof is almost the same for both processes. We elaborate the tightness of $\mathcal{A}_n(\gamma)$, for which we note that $\mathcal{A}_n(\gamma)$ is a partial sum process of a heterogeneous martingale difference array $w_{ns} = \xi_s' \sum_{t=1}^{s-1} \xi_t / \sqrt{np}$, and thus it is sufficient to show

$$\begin{aligned} E |\mathcal{A}_n(\gamma_1) - \mathcal{A}_n(\gamma_2)|^4 &= E \left| \frac{1}{\sqrt{n}} \sum_{s=[n\gamma_1]+1}^{[n\gamma_2]} w_{ns} \right|^4 \\ (SB.11) \quad &\leq E \left(\sum_s E(w_{ns}^2 | \mathcal{F}_{s-1}) / n \right)^2 + n^{-1} \max_s E |w_{ns}|^4 O(|\gamma_2 - \gamma_1|) \\ &= O(|\gamma_2 - \gamma_1|), \end{aligned}$$

where we apply the Rosenthal inequality, e.g. Hall and Heyde (1980), for the inequality and a calculation similar to (SB.14) and (SB.17) for the last equality. Specifically,

$$n^{-1} \max_s E |w_{ns}|^4 \leq \max_s E \left(E((\xi'_s \xi_s)^2 | \mathcal{F}_{s-1}) \left(\sum_{t_1, t_2 < s} \xi'_{t_1} \xi_{t_2} \right)^2 \right) n^{-3} p^{-2} = O(n^{-1/2} p^2),$$

by Assumption 4 and the same reasoning as in (SB.17).

Next, we derive finite dimensional convergence by first checking the conditions of Corollary 3.1 in Hall and Heyde (1980). The first condition is the conditional Lindeberg condition

$$(SB.12) \quad \text{For all } \eta > 0, \quad \sum_s E \left((w_{ns}/\sqrt{n})^2 1(|w_{ns}| > \eta) \middle| \mathcal{F}_{s-1} \right) \xrightarrow{p} 0,$$

for which we check the sufficient Lyapunov condition

$$(SB.13) \quad \sum_s E \left((w_{ns}/\sqrt{n})^4 \middle| \mathcal{F}_{s-1} \right) \xrightarrow{p} 0.$$

The LHS of (SB.13) is positive and, by law of iterated expectations, has mean

$$n^{-2} \sum_s E w_{ns}^4 \leq n^{-1} \max_s E w_{ns}^4 = O(n^{-1/2} p^2),$$

the final bound coming from the tightness argument. This establishes (SB.12).

The second condition of Corollary 3.1 in Hall and Heyde (1980) is

$$(SB.14) \quad \sum_s E (w_{ns}^2 | \mathcal{F}_{s-1}) / n - \mathcal{V}/2 \xrightarrow{p} 0.$$

Let $\Delta_s = \Upsilon_s \Xi_s$. Then we want to show $n^{-2} p^{-1} \sum_s \text{tr} \Delta_s - \mathcal{V}/2 \xrightarrow{p} 0$ but, because $n^{-2} p^{-1} \sum_s E \text{tr} \Delta_s \rightarrow \mathcal{V}/2$, it suffices to show

$$(SB.15) \quad n^{-2} p^{-1} \sum_s (\text{tr} \Delta_s - E \text{tr} \Delta_s) \xrightarrow{p} 0.$$

The LHS of (SB.15) has variance

$$(SB.16) \quad n^{-4} p^{-2} \sum_s E (\text{tr} \Delta_s - E \text{tr} \Delta_s)^2 + 2n^{-4} p^{-2} \sum_{s_1 < s_2} E ((\text{tr} \Delta_{s_1} - E \text{tr} \Delta_{s_1}) (\text{tr} \Delta_{s_2} - E \text{tr} \Delta_{s_2})).$$

The first term in (SB.16) is bounded by $n^{-4} p^{-2} \sum_s E (\text{tr}^2 \Delta_s)$, and observe that

$$(SB.17) \quad \begin{aligned} \sum_s E (\text{tr}^2 \Delta_s) &= \sum_s E (\text{tr}^2 (\Upsilon_s \Xi_s)) \leq \sum_s E \left\{ \bar{\lambda}^2 (\Upsilon_s) \text{tr}^2 (\Xi_s) \right\} \\ &\leq E \left\{ \left(\sum_s \bar{\lambda}^4 (\Upsilon_s) \right)^{1/2} \left(\sum_s \text{tr}^4 (\Xi_s) \right)^{1/2} \right\} \\ &\leq C n^{1/2} E \left(\sum_s \text{tr}^2 (\Xi_s) \right). \end{aligned}$$

The above inequalities are obtained as follows: first, the matrix $\Xi_s = \sum_{t_1, t_2 < s} \xi_{t_1} \xi'_{t_2}$ is symmetric and positive semidefinite as it equals $(\sum_{t < s} \Omega^{-1/2} x_t \varepsilon_t) (\sum_{t < s} \Omega^{-1/2} x_t \varepsilon_t)'$. Because Υ_s is also symmetric psd, Theorem 1 of Fang et al. (1994) yields $\text{tr} (\Upsilon_s \Xi_s) \leq$

$\bar{\lambda}(\Upsilon_s) \text{tr}(\Xi_s)$, whence the remaining inequalities follow by the Cauchy Schwarz inequality, $(\sum_s a_s^4)^{1/4} \leq (\sum_s a_s^2)^{1/2}$ and Assumption 4.

Because $\text{tr}(\Omega^{-1/2} x_{t_1}' x_{t_2}' \Omega^{-1/2}) = x_{t_1}' \Omega^{-1} x_{t_2}$, the right side of (SB.17) is

$$(SB.18) \quad Cn^{1/2} \sum_s \sum_{t_1, t_2 < s; t_3, t_4 < s} E(x_{t_1}' \Omega^{-1} x_{t_2} \varepsilon_{t_1} \varepsilon_{t_2} x_{t_3}' \Omega^{-1} x_{t_4} \varepsilon_{t_3} \varepsilon_{t_4}).$$

The contribution to (SB.18) when $t_1 = t_2 = t_3 = t_4$ is

$$Cn^{1/2} \sum_s \sum_{t < s} E\left[(x_t' \Omega^{-1} x_t)^2 \varepsilon_t^4\right] = O(n^{5/2} p^2),$$

by Assumptions 1 and 3. Thus, this case contributes $O(n^{-3/2})$ to (SB.16). Next, the contribution to (SB.18) from the case $(t_1 = t_2) \neq (t_3 = t_4)$ is

$$\begin{aligned} & Cn^{1/2} \sum_s \sum_{t_1 < t_2 < s} E(x_{t_1}' \Omega^{-1} x_{t_1} \varepsilon_{t_1}^2 E(x_{t_2}' \Omega^{-1} x_{t_2} \varepsilon_{t_2}^2 | \mathcal{F}_{t_2-1})) \\ & \leq Cn^{1/2} \sum_s \sum_{t_1 < s} E\left(x_{t_1}' \Omega^{-1} x_{t_1} \varepsilon_{t_1}^2 \sum_{t_2 < s} \text{tr} \Upsilon_{t_2}\right) \\ & \leq Cn^{3/2} p \sum_{t_1 \leq n} E\left(x_{t_1}' \Omega^{-1} x_{t_1} \varepsilon_{t_1}^2 \sum_{t_2 \leq n} \bar{\lambda}(\Upsilon_{t_2})\right) \\ & \leq Cn^{5/2} p \sum_{t_1 \leq n} \text{tr}[E(x_{t_1} x_{t_1}' \varepsilon_{t_1}^2) \Omega^{-1}] \\ & = O(n^{7/2} p^2), \end{aligned}$$

by Assumption 4, and because $\text{tr} \Upsilon_{t_2} \leq p \bar{\lambda}(\Upsilon_{t_2})$. Thus, this case contributes $O(n^{7/2} p^2)$ to SB.17, and therefore $O(n^{-1/2})$ to (SB.16).

The cases $(t_1 = t_3) \neq (t_2 = t_4)$ and $(t_1 = t_4) \neq (t_2 = t_3)$ similarly contribute a constant times

$$n^{1/2} \sum_s \sum_{t_1 \neq t_2} E(x_{t_1}' \Omega^{-1} x_{t_2} \varepsilon_{t_1} \varepsilon_{t_2})^2 \leq n^{1/2} \sum_s \sum_{t_1 \neq t_2} \left(E(x_{t_1}' \Omega^{-1} x_{t_2} \varepsilon_{t_1} \varepsilon_{t_2})^4\right)^{1/2} = O(n^{7/2} p)$$

to (SB.18), by Assumption 4 and Jensen's inequality. This ensures a negligible contribution to (SB.16). Finally,

$$\begin{aligned} & Cn^{1/2} \sum_s \sum_{t_1, t_2 < s; t_3, t_4 < s}^{\neq} E(x_{t_1}' \Omega^{-1} x_{t_2} \varepsilon_{t_1} \varepsilon_{t_2} x_{t_3}' \Omega^{-1} x_{t_4} \varepsilon_{t_3} \varepsilon_{t_4}) \\ & = O\left(n^{1/2} \sum_s \sum_{t_1, t_2 < s; t_3, t_4 < s}^{\neq} \sum_{i, j, k, l=1}^p |E(x_{t_1, i} \varepsilon_{t_1} x_{t_2, j} \varepsilon_{t_2} x_{t_3, k} \varepsilon_{t_3} x_{t_4, l} \varepsilon_{t_4})|\right) \\ & = O\left(n^{3/2} p^4 \max_s \sum_{t_1, t_2 < s; t_3, t_4 < s}^{\neq} \max_{i, j, k, l=1, \dots, p} |E(x_{t_1, i} \varepsilon_{t_1} x_{t_2, j} \varepsilon_{t_2} x_{t_3, k} \varepsilon_{t_3} x_{t_4, l} \varepsilon_{t_4})|\right), \end{aligned}$$

(SB.19)

where $\sum_{t_1, t_2 < s; t_3, t_4 < s}^{\neq}$ excludes all cases which were considered before. Therefore, in view of (SB.17) and (SB.19), to establish negligibility of the first term in (SB.16) it suffices to show

$$(SB.20) \quad n^{-5/2} p^2 \max_s \max_{i,j,k,l=1,\dots,p} \sum_{t_1, t_2 < s; t_3, t_4 < s}^{\neq} |E(x_{t_1, i} \varepsilon_{t_1} x_{t_2, j} \varepsilon_{t_2} x_{t_3, k} \varepsilon_{t_3} x_{t_4, l} \varepsilon_{t_4})| = o(1).$$

The summand on the LHS above is bounded by

$$\begin{aligned} & |E(x_{t_1, i} \varepsilon_{t_1} x_{t_2, j} \varepsilon_{t_2})| |E(x_{t_3, k} \varepsilon_{t_3} x_{t_4, l} \varepsilon_{t_4})| + |E(x_{t_1, i} \varepsilon_{t_1} x_{t_3, k} \varepsilon_{t_3})| |E(x_{t_2, j} \varepsilon_{t_2} x_{t_4, l} \varepsilon_{t_4})| \\ & + |E(x_{t_1, i} \varepsilon_{t_1} x_{t_4, l} \varepsilon_{t_4})| |E(x_{t_2, j} \varepsilon_{t_2} x_{t_3, k} \varepsilon_{t_3})| \\ & + |\text{cum}_{ijkl}(x_{t_1, i} \varepsilon_{t_1}, x_{t_2, j} \varepsilon_{t_2}, x_{t_3, k} \varepsilon_{t_3}, x_{t_4, l} \varepsilon_{t_4})| \\ & = |c_{ij}(t_1 - t_2)| |c_{kl}(t_3 - t_4)| + |c_{ik}(t_1 - t_3)| |c_{jl}(t_2 - t_4)| \\ & + |c_{il}(t_1 - t_4)| |c_{kj}(t_2 - t_3)| \\ & + |\text{cum}_{ijkl}(x_{0, i} \varepsilon_0, x_{t_2-t_1, j} \varepsilon_{t_2-t_1}, x_{t_3-t_1, k} \varepsilon_{t_3-t_1}, x_{t_4-t_1, l} \varepsilon_{t_4-t_1})|. \end{aligned}$$

(SB.21)

Because $\sum_{t_1, t_2} |c_{ij}(t_1 - t_2)| \leq n \sum_{t=-\infty}^{\infty} |c_{ij}(t)|$, by Assumption 5 and (SB.21) the LHS of (SB.20) is $O(n^2 p^2 / n^{5/2}) = o(1)$, as desired. Thus the first term in (SB.16) is negligible, and by Assumption 4 we conclude the proof of (SB.15).

We finally derive the limit of the covariance kernel of the process $(\mathcal{A}_n(\gamma), \bar{\mathcal{A}}_n(\gamma))'$. Note that $E(\mathcal{A}_n(\gamma_2) - \mathcal{A}_n(\gamma_1)) \mathcal{A}_n(\gamma_1) = 0$ for any $\gamma_1 < \gamma_2$. Thus, we compute

$$\begin{aligned} E|\mathcal{A}_n(\gamma)|^2 &= \frac{1}{n} \sum_{s=1}^{[n\gamma]} E w_s^2 \\ &= \frac{1}{n^2} \sum_{s=1}^{[n\gamma]} s \left(\frac{1}{sp} \text{tr} \sum_{t_1, t_2=1}^{s-1} E[\Omega^{-1} x_s x'_s \varepsilon_s^2 \Omega^{-1} x_{t_1} x'_{t_2} \varepsilon_{t_1} \varepsilon_{t_2}] \right) \\ &= \frac{([n\gamma] + 1) [n\gamma]}{2n^2} \lim_{s, p \rightarrow \infty} \left(\frac{1}{sp} \text{tr} \sum_{t_1, t_2=1}^{s-1} E[\Omega^{-1} x_s x'_s \varepsilon_s^2 \Omega^{-1} x_{t_1} x'_{t_2} \varepsilon_{t_1} \varepsilon_{t_2}] \right) + o(1) \\ &= \frac{\gamma^2 \mathcal{V}}{2} + o(1), \end{aligned}$$

where \mathcal{V} is given in (3.2). Thus,

$$E(\mathcal{A}_n(\gamma_1) \mathcal{A}_n(\gamma_2)) \rightarrow \frac{(\gamma_1 \wedge \gamma_2)^2}{2} \mathcal{V}.$$

Similarly, note $E(\bar{\mathcal{A}}_n(\gamma_2) - \bar{\mathcal{A}}_n(\gamma_1)) \bar{\mathcal{A}}_n(\gamma_1) = 0$ for any $\gamma_1 > \gamma_2$.

$$E|\bar{\mathcal{A}}_n(\gamma)|^2 = E \left| \frac{1}{n\sqrt{p}} \sum_{s=[n\gamma]+1}^n \sum_{t=[n\gamma]+1}^{s-1} \xi'_t \xi_s \right|^2$$

$$\begin{aligned}
&= E \left| \frac{1}{n\sqrt{p}} \sum_{s=1}^{n-[n\gamma]} \sum_{t=1}^{s-1} \xi'_{t+[n\gamma]} \xi_{s+[n\gamma]} \right|^2 \\
&= \frac{(1-\gamma)^2 \mathcal{V}}{2} + o(1),
\end{aligned}$$

implying that

$$E(\bar{\mathcal{A}}_n(\gamma_1) \bar{\mathcal{A}}_n(\gamma_2)) \rightarrow \frac{(1 - (\gamma_1 \vee \gamma_2))^2}{2} \mathcal{V}$$

And

$$\begin{aligned}
E(\mathcal{A}_n(\gamma_1) \bar{\mathcal{A}}_n(\gamma_2)) &= \frac{1\{\gamma_1 > \gamma_2\}}{n^2 p} \sum_{s=[n\gamma_2]+1}^{[n\gamma_1]} \text{tr} E \left[\xi_s \xi'_s \sum_{t=1}^{s-1} \sum_{u=[n\gamma_2]+1}^{s-1} \xi_t \xi'_u \right] \\
&= \frac{1\{\gamma_1 > \gamma_2\}}{n^2} \sum_{s=[n\gamma_2]+1}^{[n\gamma_1]} (s-1-[n\gamma_2]) \mathcal{V} + o(1) \\
&= \frac{1\{\gamma_1 > \gamma_2\}}{2} (\gamma_1 - \gamma_2)^2 \mathcal{V} + o(1),
\end{aligned}$$

under Assumption 4 that

$$\lim_{s,p \rightarrow \infty} \left(\frac{1}{sp} \text{tr} \sum_{t_1=1}^s \sum_{t_2=1}^{[s\gamma]} E(\Omega^{-1} x_s x'_s \varepsilon_s^2) (\Omega^{-1} x_{t_1} x'_{t_2} \varepsilon_{t_1} \varepsilon_{t_2}) \right) = \gamma \mathcal{V}.$$

Therefore, we conclude that

$$\begin{pmatrix} \mathcal{A}_n(\gamma) \\ \bar{\mathcal{A}}_n(\gamma) \end{pmatrix} \Rightarrow \sqrt{\frac{\mathcal{V}}{2}} \begin{pmatrix} W(\gamma) \\ \bar{W}(\gamma) \end{pmatrix}.$$

Finally, apply the continuous mapping theorem to get

$$\begin{aligned}
\mathcal{S}_n(\gamma) &= \sqrt{2} \left(\frac{\mathcal{A}_n(\gamma)}{\gamma} + \frac{\bar{\mathcal{A}}_n(\gamma)}{(1-\gamma)} - \mathcal{A}_n(1) \right) \\
&\Rightarrow \sqrt{\mathcal{V}} \left(\frac{W(\gamma)}{\gamma} + \frac{\bar{W}(\gamma)}{(1-\gamma)} - W(1) \right) = \sqrt{\mathcal{V}} \mathcal{Q}(\gamma),
\end{aligned}$$

on any compact subset Γ in $(0, 1)$. ■

We note some preliminary calculations useful for the sequel. Note that

$$\hat{\delta}_2(\gamma) = A(\gamma)y = \delta_2 + A(\gamma)e = \delta_2 + A(\gamma)\varepsilon + A(\gamma)r.$$

Because $\delta_2 = 0$ under \mathcal{H}_0 , we have

$$(SB.22) \quad W_n(\gamma) = n(\varepsilon + r)' A'(\gamma) \hat{B}(\gamma)^{-1} A(\gamma) (\varepsilon + r),$$

where $\hat{B}(\gamma) = R \hat{M}(\gamma)^{-1} \hat{\Omega}(\gamma) \hat{M}(\gamma)^{-1} R'$.

Lemma SL.2. *Under the conditions of Theorem ST.1, for all sufficiently large n ,*

$$\sup_{\gamma \in \Gamma} \|\hat{M}(\gamma)\| = O_p(1), \sup_{\gamma \in \Gamma} \|\hat{M}(\gamma)^{-1}\| = O_p(\lambda_n^{-1}).$$

Proof. Note that, by the triangle inequality,

$$\|\hat{M}(\gamma)^{-1}\| \leq \|\hat{M}(\gamma)^{-1}\| \|\hat{M}(\gamma) - M(\gamma)\| \|M(\gamma)^{-1}\| + \|M(\gamma)^{-1}\|,$$

so

$$\|\hat{M}(\gamma)^{-1}\| \left(1 - \|\hat{M}(\gamma) - M(\gamma)\| \|M(\gamma)^{-1}\|\right) \leq \|M(\gamma)^{-1}\|,$$

using the triangle inequality. Taking limits of the last displayed expression as $n \rightarrow \infty$ and using Assumption 3, the rate condition (SB.1) yields $\|\hat{M}(\gamma)^{-1}\| = O_p(\lambda_n^{-1})$. Next, noting that

$$\|\hat{M}(\gamma)\| \leq \|\hat{M}(\gamma) - M(\gamma)\| + \|M(\gamma)\|,$$

the lemma follows by using Assumption 3. ■

It is useful to first establish the stochastic order of $\|\delta - \hat{\delta}(\gamma)\|$.

Lemma SL.3. *Under the conditions of Theorem ST.1, $\sup_{\gamma \in \Gamma} \|\delta - \hat{\delta}(\gamma)\| = O_p(\lambda_n^{-1} \sqrt{p/n})$.*

Proof. Note that $\delta - \hat{\delta}(\gamma) = \hat{M}(\gamma)^{-1} n^{-1} \sum_{t=1}^n x_t(\gamma) e_t$ and that

$$\begin{aligned} \|\delta - \hat{\delta}(\gamma)\|^2 &= O_p \left(\|\hat{M}(\gamma)^{-1}\|^2 n^{-2} \left\| \sum_{t=1}^n x_t(\gamma) e_t \right\|^2 \right) = \lambda_n^{-2} O_p \left(n^{-2} \left\| \sum_{t=1}^n x_t(\gamma) e_t \right\|^2 \right) \\ &= \lambda_n^{-2} O_p \left(n^{-2} \left\| \sum_{t=1}^n x_t(\gamma) \varepsilon_t \right\|^2 + n^{-2} \|X(\gamma)' r\|^2 \right), \end{aligned}$$

uniformly in γ , by Lemma SL.2. Next, $E \left(n^{-2} \left\| \sum_{t=1}^n x_t(\gamma) \varepsilon_t \right\|^2 \right)$ equals

$$(SB.23) \quad E \left(n^{-2} \sum_{s,t=1}^n x'_t(\gamma) x_s(\gamma) \varepsilon_s \varepsilon_t \right),$$

which is

$$(SB.24) \quad n^{-2} \sum_{t=1}^n E \|x_t(\gamma)\|^2 \sigma_t^2 + 2n^{-2} \sum_{s < t} E (x'_t(\gamma) x_s(\gamma) E(\varepsilon_s E(\varepsilon_t | \varepsilon_r, r < t))) = O_p(p/n),$$

by Assumptions 1 and $E x'_t(\gamma) x_t(\gamma) = O(p)$. Finally,

$$(SB.25) \quad n^{-2} \|X(\gamma)' r\|^2 \leq n^{-2} \|X(\gamma)\|^2 \|r\|^2 = \bar{\lambda} \left(\hat{M}(\gamma) \right) n^{-1} \|r\|^2 = O_p(1/n),$$

by (2.2) and Lemma SL.2. Therefore,

$$(SB.26) \quad \sup_{\gamma \in \Gamma} \left\| \delta - \hat{\delta}(\gamma) \right\| = \lambda_n^{-1} O_p(\sqrt{p}/\sqrt{n}),$$

by Markov's inequality. ■

Observe that because

$$(SB.27) \quad M(\gamma)^{-1} = \begin{bmatrix} (1-\gamma)^{-1} M^{-1} & (1-\gamma)^{-1} M^{-1} \\ (1-\gamma)^{-1} M^{-1} & [\gamma(1-\gamma)]^{-1} M^{-1} \end{bmatrix},$$

we have

$$(SB.28) \quad B(\gamma)^{-1} = \gamma(1-\gamma) M \Omega^{-1} M.$$

Lemma SL.4. *Under the conditions of Theorem ST.1,*

$$\sup_{\gamma \in \Gamma} \{ \underline{\lambda}(B(\gamma)) \}^{-1} = O(\lambda_n^{-1}) \quad \text{and} \quad \sup_{\gamma \in \Gamma} \bar{\lambda}(B(\gamma)) = O(\lambda_n^{-2}).$$

Proof. $\{ \underline{\lambda}(B(\gamma)) \}^{-1} = \bar{\lambda}(B(\gamma)^{-1})$, which, using (SB.28), is bounded by

$$C \bar{\lambda}(M \Omega^{-1} M) = C \|M \Omega^{-1} M\| \leq C \bar{\lambda}(M)^2 \underline{\lambda}(\Omega)^{-1} = O(\lambda_n^{-1}),$$

uniformly on the compact Γ , using Assumption 3(ii). For the second part of the claim, because (SB.28) implies $B(\gamma) = [\gamma(1-\gamma)]^{-1} M^{-1} \Omega M^{-1}$, it follows similarly that $\bar{\lambda}(B(\gamma))$ is uniformly bounded by a constant times ¹

$$\bar{\lambda}(M^{-1} \Omega M^{-1}) = \|M^{-1} \Omega M^{-1}\| \leq \underline{\lambda}(M)^{-2} \bar{\lambda}(\Omega) = O(\lambda_n^{-2}).$$

■

Lemma SL.5. *Under the conditions of Theorem ST.2,*

$$\sup_{\gamma \in \Gamma} \left\| \hat{B}(\gamma) \right\| = O_p(\lambda_n^{-2}), \sup_{\gamma \in \Gamma} \left\| \hat{B}(\gamma)^{-1} \right\| = O_p(\lambda_n^{-1}).$$

Proof. First, define $\tilde{B}(\gamma) = R \hat{M}(\gamma)^{-1} \Omega(\gamma) \hat{M}(\gamma)^{-1} R'$. We will use uniform bounds in the calculations without explicitly mentioning this in each step to simplify notation. Proceeding as in the proof of Lemma SL.2, we can write

$$(SB.29) \quad \left\| \hat{B}(\gamma)^{-1} \right\| \left(1 - \left\| \hat{B}(\gamma) - \tilde{B}(\gamma) \right\| \right) \leq \left\| \tilde{B}(\gamma)^{-1} \right\|,$$

$$(SB.30) \quad \left\| \tilde{B}(\gamma)^{-1} \right\| \left(1 - \left\| \tilde{B}(\gamma) - B(\gamma) \right\| \right) \leq \left\| B(\gamma)^{-1} \right\|.$$

Next, Lemma SL.2 implies

$$(SB.31) \quad \left\| \hat{B}(\gamma) - \tilde{B}(\gamma) \right\| \leq \|R\|^2 \left\| \hat{M}(\gamma)^{-1} \right\|^2 \left\| \hat{\Omega}(\gamma) - \Omega(\gamma) \right\| = O_p(\lambda_n^{-2} v_p).$$

On the other hand, $\tilde{B}(\gamma) - B(\gamma)$ equals

$$R \left[\hat{M}(\gamma)^{-1} \Omega(\gamma) \hat{M}(\gamma)^{-1} - M(\gamma)^{-1} \Omega(\gamma) M(\gamma)^{-1} \right] R'.$$

¹If $\underline{\lambda}(M \Omega^{-1} M) \geq \lambda_n$, the bound in this lemma becomes $O(\lambda_n^{-1})$.

By adding and subtracting terms inside the square brackets, this can be written as

$$(SB.32) \quad \begin{aligned} & R \left[M(\gamma)^{-1} \left(\hat{M}(\gamma) - M(\gamma) \right) \hat{M}(\gamma)^{-1} \Omega(\gamma) \hat{M}(\gamma)^{-1} \right] R' \\ & + RM(\gamma)^{-1} \Omega(\gamma) M(\gamma)^{-1} \left(\hat{M}(\gamma) - M(\gamma) \right) \hat{M}(\gamma)^{-1} R'. \end{aligned}$$

By this fact, Assumption 3, Lemmas SL.1 and SL.2, and (SB.1), we deduce from (SB.32) that

$$(SB.33) \quad \left\| \tilde{B}(\gamma) - B(\gamma) \right\| = O_p(\lambda_n^{-3} \varkappa_p + \lambda_n^{-2} v_p) = o_p(1).$$

The lemma now follows by taking limits of (SB.29) and (SB.30), and using (SB.31), (SB.33) and Lemma SL.4. ■

Lemma SL.6. *Under the conditions of Theorem ST.2 and \mathcal{H}_0 ,*

$$\frac{W_n(\gamma)}{\sqrt{2p}} = \frac{n\varepsilon' A(\gamma)' B(\gamma)^{-1} A(\gamma) \varepsilon}{\sqrt{2p}} + o_p(1).$$

Proof. Recall the notation $\hat{M} = n^{-1} X' X$ and $\hat{S}(\gamma) = n^{-1} X'^*(\gamma) X(\gamma)$. Notice that from (SB.22) we obtain

$$(SB.34) \quad \frac{W_n(\gamma)}{\sqrt{2p}} = \frac{n\varepsilon' A(\gamma)' \hat{B}(\gamma)^{-1} A(\gamma) \varepsilon}{\sqrt{2p}} + \frac{2n\varepsilon' A(\gamma)' \hat{B}(\gamma)^{-1} A(\gamma) r}{\sqrt{2p}} + \frac{nr' A(\gamma)' \hat{B}(\gamma)^{-1} A(\gamma) r}{\sqrt{2p}},$$

with r the $n \times 1$ vector with elements r_t . Begin with the modulus of the last term on the RHS of (SB.34). Recalling the relation in (SB.3) and (SB.4) for $A(\gamma)r$, we bound it by $Cn/\sqrt{2p}$ times

$$(SB.35) \quad \begin{aligned} & \left\| n^{-1} X' r \right\|^2 \left\| I - \hat{S}(\gamma) \hat{M}^{-1} \right\|^2 \left\| (n^{-1} X^*(\gamma)' M_X X^*(\gamma))^{-1} \right\|^2 \left\| \hat{B}(\gamma)^{-1} \right\| \\ & = O_p(\lambda_n^{-5} \varkappa_p^2 n^{-1}), \end{aligned}$$

where Assumption 2 bounds the first term, Lemma SL.9 yields a bound for the second and third terms after expanding the third term by (SB.3), and the last term is $O_p(\lambda_n^{-1})$ Lemma SL.5. Thus (SB.35) implies that the third term on the RHS of (SB.34) is $o_p(1)$.

We now show that the first term on the RHS of (SB.34) is

$$(SB.36) \quad \frac{n\varepsilon' A(\gamma)' B(\gamma)^{-1} A(\gamma) \varepsilon}{\sqrt{2p}} + o_p(1).$$

Indeed, as above,

$$(SB.37) \quad \begin{aligned} \frac{n\varepsilon' A(\gamma)' \left(\hat{B}(\gamma)^{-1} - B(\gamma)^{-1} \right) A(\gamma) \varepsilon}{\sqrt{2p}} &= \frac{n}{\sqrt{p}} O_p \left(\left\| n^{-1} X' \varepsilon \right\|^2 \left\| \hat{B}(\gamma)^{-1} - B(\gamma)^{-1} \right\| \right) \\ &= \sqrt{p} O_p \left(B(\gamma)^{-1} \left\| B(\gamma) - \hat{B}(\gamma) \right\| \left\| \hat{B}(\gamma)^{-1} \right\| \right) \\ &= \lambda_n^{-4} \sqrt{p} O_p \left(\lambda_n^{-1} \left\| \hat{M}(\gamma) - M(\gamma) \right\| \right) + \end{aligned}$$

$$\begin{aligned}
& \left\| \hat{\Omega}(\gamma) - \Omega(\gamma) \right\| \\
\text{(SB.38)} \quad &= O_p \left(\lambda_n^{-4} \sqrt{p} \left(\lambda_n^{-1} \varkappa_p + v_p \right) \right),
\end{aligned}$$

by Lemma SL.5, using equations (SB.24), (SB.31) and (SB.33) in the proofs thereof. This is negligible by (SB.1).

For the second term on the RHS of (SB.34), apply the Cauchy-Schwarz inequality and the preceding two results. Then, the second term becomes $o_p(1)$, establishing the lemma. ■

Denote, for convenience, $C(\gamma) = [\gamma(1-\gamma)]^{-1} n^{-1} \Sigma^{\frac{1}{2}} G(\gamma) \Omega^{-1} G(\gamma) \Sigma^{\frac{1}{2}}$, where $\Sigma = \text{diag}[\sigma_1^2, \dots, \sigma_n^2]$.

Lemma SL.7. *Under the conditions of Theorem ST.2, any eigenvalue λ of $C(\gamma)$ satisfies*

$$P(|\lambda(\lambda - 1)| < \eta) \rightarrow 1,$$

as $n \rightarrow \infty$, for any $\eta > 0$.

Proof. We have

$$\begin{aligned}
C(\gamma)^2 &= [\gamma(1-\gamma)]^{-1} n^{-1} \Sigma^{\frac{1}{2}} G(\gamma) \Omega^{-1} [\gamma(1-\gamma)]^{-1} n^{-1} G(\gamma)' \Sigma G(\gamma) \Omega^{-1} G(\gamma)' \Sigma^{\frac{1}{2}} \\
&= [\gamma(1-\gamma)]^{-1} n^{-1} \Sigma^{\frac{1}{2}} G(\gamma) \Omega^{-1} \Omega \Omega^{-1} G(\gamma)' \Sigma^{\frac{1}{2}} \\
&+ [\gamma(1-\gamma)]^{-1} n^{-1} \Sigma^{\frac{1}{2}} G(\gamma) \Omega^{-1} \{ [\gamma(1-\gamma)]^{-1} n^{-1} G(\gamma)' \Sigma G(\gamma) - \Omega \} \\
&\times \Omega^{-1} G(\gamma)' \Sigma^{\frac{1}{2}} \\
&= C(\gamma) + D(\gamma),
\end{aligned}$$

say. We now prove that

$$\text{(SB.39)} \quad \|D(\gamma)\| = o_p(1) \text{ as } n \rightarrow \infty.$$

In view of Assumptions 1 and 3(i), to prove (SB.39) it suffices to show that

$$\text{(SB.40)} \quad \left\| [\gamma(1-\gamma)]^{-1} n^{-1} G(\gamma)' \Sigma G(\gamma) - \Omega \right\| = o_p(1).$$

But

$$\begin{aligned}
n^{-1} G(\gamma)' \Sigma G(\gamma) &= n^{-1} (1 - 2\gamma) \sum_{t=1}^{[n\gamma]} x_t x_t' \sigma_t^2 + \gamma^2 \Omega \\
&= (1 - 2\gamma) \left(n^{-1} \sum_{t=1}^{[n\gamma]} x_t x_t' \sigma_t^2 - \gamma \Omega \right) + [\gamma(1-\gamma)] \Omega,
\end{aligned}$$

so (SB.40) follows if $\left\| n^{-1} \sum_{t=1}^{[n\gamma]} x_t x_t' \sigma_t^2 - \gamma \Omega \right\| = o_p(1)$, which is true by Assumption 3. Thus (SB.39) is established.

Let λ be any eigenvalue of $C(\gamma)$ and w be the corresponding eigenvector, normalised to $\|w\| = 1$. Because $\lambda w = C(\gamma)w$, we have $\lambda C(\gamma)w = C(\gamma)^2 w = [C(\gamma) + D(\gamma)]w = \lambda w + D(\gamma)w$, implying $\lambda(\lambda - 1)w = D(\gamma)w$. Thus

$$\text{(SB.41)} \quad |\lambda(\lambda - 1)| = \|D(\gamma)w\| \leq \|D(\gamma)\|.$$

Then, for arbitrary $\eta > 0$,

$$P(|\lambda(\lambda - 1)| < \eta) = P(\|D(\gamma)w\| < \eta) \geq P(\|D(\gamma)\| < \eta) \rightarrow 1, \text{ as } n \rightarrow \infty,$$

by (SB.39). This completes the proof. ■

We have $\mathcal{R}_n(\gamma) = [\gamma(1 - \gamma)]^{-1} n^{-1} \varepsilon' G(\gamma)' \Omega^{-1} G(\gamma)' \varepsilon$, which in turn equals

$$(SB.42) \quad [\gamma(1 - \gamma)]^{-1} n^{-1} \sum_{t,s=1}^n g_t(\gamma)' \Omega^{-1} g_s(\gamma) \varepsilon_t \varepsilon_s.$$

Note that $\text{tr}\{C(\gamma)\}$ is the sum of the eigenvalues of $C(\gamma)$, which is a symmetric matrix with rank p . Thus, in view of Lemma SL.7 it has p eigenvalues that approach 1 in probability, with the remainder approaching 0. Thus,

$$(SB.43) \quad \frac{\mathcal{R}_n(\gamma) - \text{tr}(C(\gamma))}{\sqrt{2p}} = \frac{\mathcal{R}_n(\gamma) - p}{\sqrt{2p}} + o_p(1),$$

whence using (SB.42) we deduce that (SB.43) equals

$$(SB.44) \quad \frac{n^{-1} \sum_{t=1}^n g_t(\gamma)' \Omega^{-1} g_t(\gamma) (\varepsilon_t^2 - \sigma_t^2) + n^{-1} \sum_{s \neq t} g_t(\gamma)' \Omega^{-1} g_s(\gamma) \varepsilon_t \varepsilon_s}{\gamma(1 - \gamma) \sqrt{2p}}.$$

Lemma SL.8. *Under the conditions of Theorem ST.2,*

$$(SB.45) \quad \sup_{\gamma \in \Gamma} n^{-1} \sum_{t=1}^n g_t(\gamma)' \Omega^{-1} g_t(\gamma) (\varepsilon_t^2 - \sigma_t^2) = o_p(1) \text{ as } n \rightarrow \infty.$$

Proof. Conditional on x_t , the LHS of (SB.45) has mean zero and variance

$$(SB.46) \quad n^{-2} \sum_{t=1}^n (g_t(\gamma)' \Omega^{-1} g_t(\gamma))^2 E[(\varepsilon_t^2 - \sigma_t^2)^2]$$

$$(SB.47) \quad + 2n^{-2} \sum_{s < t} g_s(\gamma)' \Omega^{-1} g_s(\gamma) g_t(\gamma)' \Omega^{-1} g_t(\gamma) E[(\varepsilon_t^2 - \sigma_t^2)(\varepsilon_s^2 - \sigma_s^2)].$$

The expectation in (SB.47) equals $E[(\varepsilon_t^2 - \sigma_t^2) E[(\varepsilon_s^2 - \sigma_s^2) | \varepsilon_s]] = 0$, by Assumption 1. Also by Assumption 1, (SB.46) is bounded by a constant times

$$n^{-2} \|\Omega^{-1}\|^2 \sum_{t=1}^n \|g_t(\gamma)\|^4 \leq n^{-2} \|\Omega^{-1}\|^2 \sum_{t=1}^n (\|x_t(\gamma)\|^4 + \gamma^4 \|x_t\|^4) = O_p\left(\lambda_n^{-2} \frac{p^2}{n}\right),$$

uniformly in γ , the last equality following by Assumption 3(i). ■

Lemma SL.9. *Under the conditions of Theorem ST.2, as $n \rightarrow \infty$,*

$$\left\| \left(I - \hat{M}^{-1} \hat{S}(\gamma) \right)^{-1} - \gamma^{-1} I \right\| = O_p(\lambda_n^{-1} \kappa_p).$$

Proof. First note that $\left\| \left(I - \hat{M}^{-1} \hat{S}(\gamma) \right) - \gamma I \right\|$ equals

$$\left\| (1 - \gamma)I - \hat{M}^{-1} \left(\hat{S}(\gamma) - (1 - \gamma)M \right) - (1 - \gamma)\hat{M}^{-1}M \right\|$$

$$\begin{aligned}
&\leq (1-\gamma) \left\| \hat{M}^{-1} \right\| \left(\left\| \hat{M} - M \right\| + \left\| \hat{S}(\gamma) - (1-\gamma)M \right\| \right) \\
&= O_p \left(\lambda_n^{-1} \varkappa_p \right),
\end{aligned}$$

by Assumptions 3. Since

$$\left(I - \hat{M}^{-1} \hat{S}(\gamma) \right)^{-1} - \gamma^{-1} I = -\gamma^{-1} \left(I - \hat{M}^{-1} \hat{S}(\gamma) \right)^{-1} \left\{ \left(I - \hat{M}^{-1} \hat{S}(\gamma) \right) - \gamma I \right\}$$

and $\left\| \left(I - \hat{M}^{-1} \hat{S}(\gamma) \right)^{-1} \right\| \leq \left\| \hat{M} \right\| \left\| \left(n^{-1} \sum_{t=1}^{[n\gamma]} x_t x_t' \right)^{-1} \right\| = O_p \left(\lambda_n^{-2} \varkappa_p \right)$, by Assumption 3, the lemma is established. ■

Lemma SL.10. *Under the conditions of Theorem ST.2, as $n \rightarrow \infty$,*

$$\begin{aligned}
&\left\| \left(\hat{S}(\gamma)^{-1} X'^*(\gamma) \varepsilon - \gamma(1-\gamma)^{-1} \hat{M}^{-1} X' \varepsilon \right) - \hat{M}^{-1} \frac{\left(\sum_{t=1}^{[n\gamma]} \varepsilon_t x_t - \gamma \sum_{t=1}^n \varepsilon_t x_t \right)}{1-\gamma} \right\| \\
&= O_p \left(\lambda_n^{-2} \sqrt{np} \varkappa_p \right).
\end{aligned}
\tag{SB.48}$$

Proof. First note that

$$(1-\gamma)^{-1} \left(\sum_{t=1}^{[n\gamma]} \varepsilon_t x_t - \gamma \sum_{t=1}^n \varepsilon_t x_t \right) = (1-\gamma)^{-1} X'^*(\gamma) \varepsilon - \gamma(1-\gamma)^{-1} X' \varepsilon,$$

so the term inside the norm in (SB.48) equals

$$\left(\hat{S}(\gamma)^{-1} - (1-\gamma)^{-1} \hat{M}^{-1} \right) X'^*(\gamma) \varepsilon = (1-\gamma)^{-1} \hat{M}^{-1} \left((1-\gamma) \hat{M} - \hat{S}(\gamma) \right) \hat{S}(\gamma)^{-1} X'^*(\gamma) \varepsilon.
\tag{SB.49}$$

The norm of the RHS of (SB.49) is bounded by a constant times

$$\left\| \hat{M}^{-1} \right\| \left\| \hat{S}(\gamma)^{-1} \right\| \left(\left\| n^{-1} \sum_{t=1}^{[n\gamma]} x_t x_t' - \gamma M \right\| + \left\| \hat{M} - M \right\| \right) \|X'^*(\gamma) \varepsilon\| = O_p \left(\lambda_n^{-2} \sqrt{np} \varkappa_p \right),$$

the last equality following from Assumptions 3, Lemma SL.2, and also (SB.24). ■

S.C. PROOF OF THEOREM 4.2

Proof. Let ε^* denote the vector collecting $\varepsilon_t^* = \hat{e}_t(\gamma) \xi_t$, where ξ_t is an iid sequence of Rademacher variables. Then,

$$\hat{\delta}_2^*(\gamma) = A(\gamma) \varepsilon^*,$$

since $\delta_2 = 0$ under \mathcal{H}_0 . Also, we have

$$W_n^*(\gamma) = n (\varepsilon^*)' A'(\gamma) \hat{B}^*(\gamma)^{-1} A(\gamma) \varepsilon^*,
\tag{SC.1}$$

where $\hat{B}^*(\gamma) = R \hat{M}(\gamma)^{-1} \hat{\Omega}^*(\gamma) \hat{M}(\gamma)^{-1} R'$ and $\hat{\Omega}^*(\gamma)$ is constructed as $\hat{\Omega}(\gamma)$ with the bootstrap sample.

We begin with

$$\begin{aligned}
E^* \bar{W}_n^*(\gamma) &= ntr A'(\gamma) \hat{B}(\gamma)^{-1} A(\gamma) E^* \varepsilon^* (\varepsilon^*)', \\
(\text{SC.2}) \quad &= ntr A'(\gamma) \hat{B}(\gamma)^{-1} A(\gamma) \text{diag} [\hat{e}_1(\gamma)^2, \dots, \hat{e}_n(\gamma)^2],
\end{aligned}$$

where $\bar{W}_n^*(\gamma) = n (\varepsilon^*)' A'(\gamma) B(\gamma)^{-1} A(\gamma) \varepsilon^*$. Note that the term in (SC.2) subtracted by p is $o_p(p^{1/2})$ uniformly in γ due to Lemma SL.7, Lemma SL.8, and Lemma SL.3.

Next, we show that the order of the difference between $E^* \bar{W}_n^*(\gamma)$ and $E^* W_n^*(\gamma)$ is $o_p(p^{1/2})$. Following (SB.38), write

$$E^* |\bar{W}_n^*(\gamma) - W_n^*(\gamma)| \leq E^* \left(\|n^{-1/2} A'(\gamma) \varepsilon^*\|^2 \left\| \hat{B}(\gamma)^{-1} - \hat{B}^*(\gamma)^{-1} \right\| \right).$$

To apply the Cauchy-Schwarz inequality, and to bound $E^* \left\| \hat{B}(\gamma)^{-1} - \hat{B}^*(\gamma)^{-1} \right\|^2$, we derive bounds for $E^* \left\| \hat{B}^*(\gamma)^{-1} \right\|^4$ and $E^* \left\| \hat{B}(\gamma) - \hat{B}^*(\gamma) \right\|^4$. Since both are similar to the derivations for the sample counterparts in Lemmas SL.1 and SL.5, we only illustrate the latter. Recall $\hat{B}(\gamma) - \hat{B}^*(\gamma) = R' \hat{M}(\gamma)^{-1} (\hat{\Omega}(\gamma) - \hat{\Omega}^*(\gamma)) \hat{M}(\gamma)^{-1} R$ and $\sup_{\gamma \in \Gamma} \left\| \hat{M}(\gamma)^{-1} \right\| = O_p(\lambda_n^{-1})$ by Lemma SL.2. Following the steps in the proof of Lemma SL.1, the term $\hat{\Omega}(\gamma) - \hat{\Omega}^*(\gamma)$ is given by the sum of $U_1^*(\gamma)$ and $U_3^*(\gamma)$ therein. Due to the triangle inequality and c_r inequality, we only show $E^* \left\| U_j^*(\gamma) \right\|^4 = O_p(\lambda_n^{-8} p^{12}/n^4)$, for $j = 1, 3$. Note that by the independence of the sequence ξ_t

$$\begin{aligned}
E^* \|U_1^*(\gamma)\|^4 &\leq \left(n^{-1} \sum_{t=1}^n (x'_t(\gamma) x_t(\gamma))^2 \right)^4 E^* \left((\delta^* - \hat{\delta}^*(\gamma))' (\delta^* - \hat{\delta}^*(\gamma)) \right)^4 \\
&\leq O_p(p^8) \left\| \hat{M}(\gamma)^{-1} \right\|^8 n^{-8} \sum_{t_1, t_2, t_3, t_4} \hat{e}_{t_1}^2 x'_{t_1} x_{t_1} \cdots \hat{e}_{t_4}^2 x'_{t_4} x_{t_4},
\end{aligned}$$

to yield the desired result and the bound for U_3^* is similarly obtained. Putting these together yields $E^* \left\| \hat{B}(\gamma)^{-1} - \hat{B}^*(\gamma)^{-1} \right\|^2 = O_p(\lambda_n^{-10} p^{12}/n^4)$.

Next, similar to the preceding bound,

$$E^* \|n^{-1/2} A'(\gamma) \varepsilon^*\|^4 = O_p(\lambda_n^{-4}) \left(n^{-1} \sum_{t=1}^n x'_t x_t \hat{e}_t^2 \right)^2 = O_p(\lambda_n^{-4} p^2),$$

as ξ_t is an iid Rademacher sequence. Then, under the condition (3.1), $\lambda_n^{-14} p^{14}/n^4 = o(p)$ and this completes the proof. ■

S.D. PRIMITIVE CONDITIONS FOR EXAMPLES E2 AND E3

For the AR model in E2, a set of primitive conditions is given in the next Proposition. Gonçalves and Kilian (2007) has emphasized the empirical relevance of allowing for conditional heteroskedasticity in autoregressive models, which is allowed below by relaxing Berk (1974)'s condition of an iid error to an mds process. Let L be the lag operator and $b(L) = \sum_{j=1}^{\infty} b_j L^j$ denote the lag polynomial.

Proposition SP.1. *Suppose that (1) $b(z) \neq 0$ for any $|z| \leq 1$ and $b^{-1}(e^{i\lambda})$ exists and is nonzero for $-\pi < \lambda \leq \pi$. (2) $\{\varepsilon_t\}$ is a stationary mds that possesses a density of bounded variation, $E\varepsilon_t = 0$ and $E|\varepsilon_t|^\kappa < C$ for some $\kappa \geq 4$ and $E(\varepsilon_t^2|\mathcal{F}_{t-1})$ is bounded and bounded away from zero. (3) $p^3 = o(n)$. (4) $\sum_{j=p}^\infty |b_j| = o(n^{-1/2})$. (5) (σ_t, y_{t-1}) is ρ -mixing with $\sum_{j=1}^\infty \rho(2^j) < \infty$. Then, Assumptions 1 - 3 are satisfied with $\varkappa_p = v_p = p^{-1/3}$.*

Proof. Lemma 3 of Berk (1974) established that the minimum eigenvalue of the limiting autocovariance matrix M is bounded away from zero. As for Ω , note that for some $c > 0$, which is an a.s. lower bound of $E(\varepsilon_t^2|\mathcal{F}_{t-1})$, and any $|a| = 1$

$$a'\Omega a = E(a'x_t)^2 \varepsilon_t^2 = E(a'x_t)^2 E(\varepsilon_t^2|\mathcal{F}_{t-1}) \geq cE(a'x_t)^2,$$

to conclude that the minimum eigenvalue of Ω is also bounded away from zero.

Lemma 3.4 of Peligrad (1982) yields $E|\sum_{t=1}^n (z_t^2 - Ez_t^2)| \leq n \sum_i \rho(2^i) Ez_t^4$ for a ρ -mixing sequence z_t . Since $z_t = \sigma_t y_{t-j}$ for $j = 1, \dots, p$ in the current case and (σ_t, y_{t-1}) is ρ -mixing, the bound may be set as $n \sum_i \rho(0 \vee (2^i - p)) Ez_t^4 \leq n(\sum_i \rho(2^i) + \log p) E\sigma_t^4 y_{t-j}^4$ for any $j \leq p$. Then, $\|\bar{\Omega}(\gamma) - \Omega(\gamma)\| = O_p(n^{-1}p^2 \log p)$. The rest of the proof is given in Lemma SL.1. ■

Let $\mathbb{Z} \subseteq \mathbb{R}^k$ denote the support of z_t in E3. The following proposition provides some more primitive conditions for E3 as given by Chen and Christensen (2015).

Proposition SP.2. *Suppose that the following hold: (1) The sequence $\{z_t\}$ is strictly stationary and β -mixing with β -mixing coefficient $\beta(\cdot)$. Let $q = q(n)$ be a sequence of integers satisfying $\beta(q)n/q \rightarrow 0$ as $n \rightarrow \infty$ and $q \leq n/2$; (2) \mathbb{Z} is compact and rectangular, and $\sup_{z \in \mathbb{Z}} \|x_{nt}(z)\| = O(\vartheta_p)$; (3) The x_t are tensor-products of power series, univariate polynomial spline, trigonometric polynomial wavelet or orthogonal polynomial bases. Then, Assumptions 1 -3 are met with $\varkappa_p = \vartheta_p \sqrt{q(\log p)}/n$ and $v_p = \min\{p^3/n, \vartheta_p^2 p/n\}$.*

Proof. We prove that Assumption 3 is met for the partial sum only, with the result for the full sum following from Corollary 4.2 of Chen and Christensen (2015). By Assumption 3, we can normalize the x_t so that $E(x_t x_t') = I_p$ without loss of generality. The result then follows by Corollary SC.1 by taking $\Xi_{t,n} = n^{-1}(x_t x_t' - I_p)$, which implies that the terms in Theorem ST.1 have bounds: $R_n \leq n^{-1}(C\vartheta_p^2 + 1)$ and $s_n^2 \leq n^{-2}(C\vartheta_p^2 + 1)$. The second claim follows similarly. The rest of the proof is given in Lemma SL.1. ■

The permissible mixing decay rate depends on the dimension p of x_t : larger p requires faster mixing decay. Both exponential and geometric decays are allowed. See the discussions of Assumption 4 and Remark 2.3 in Chen and Christensen (2015) for more detailed discussion in relation to the sieve basis functions. The sequence q depends on the mixing decay rate. For instance, if $\beta(q)$ decays at an exponential rate, q can be set as $\log n$. If all elements of $x_t(\cdot)$ are bounded, then $\vartheta_p = p^{1/2}$. Under

suitable conditions, it can be shown that $\vartheta_p = p$ for power series or orthogonal polynomials and $\vartheta_p = p^{1/2}$ for univariate polynomial splines, trigonometric polynomials or wavelets, see Newey (1997); Chen and Christensen (2015).

REFERENCES

- Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.
- Chen, X. and T. M. Christensen (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics* 188, 447–465.
- Dedecker, J. and C. Prieur (2004). Coupling for τ -dependent sequences and applications. *Journal of Theoretical Probability* 17, 861–885.
- Fang, Y., K. A. Loparo, and X. Feng (1994). Inequalities for the trace of matrix product. *IEEE Transactions on Automatic Control* 39, 2489–2490.
- Gonçalves, S. and L. Kilian (2007). Asymptotic and bootstrap inference for $AR(\infty)$ processes with conditional heteroskedasticity. *Econometric Reviews* 26, 609–641.
- Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. Academic Press.
- Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.
- Peligrad, M. (1982). Invariance principles for mixing sequences of random variables. *Annals of Probability* 10, 968–981.
- Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. Springer-Verlag, Berlin.
- Tropp, J. A. (2011). Freedman’s inequality for matrix martingales. *Electronic Communications in Probability* 16, 262–270.