2021

Three essays on predictive regression, low-frequency variation, and dynamic stochastic general equilibrium models

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

THREE ESSAYS ON PREDICTIVE REGRESSION, LOW-FREQUENCY VARIATION, AND DYNAMIC STOCHASTIC GENERAL EQUILIBRIUM MODELS

by

TAOSONG DENG

B.A., Hunan University, 2012
M.A., Boston University, 2013

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2021

Approved by

First Reader

Zhongjun Qu, Ph.D.
Professor of Economics

Second Reader

Pierre Perron, Ph.D.
Professor of Economics

Third Reader

Jean-Jacques Forneron, Ph.D.
Assistant Professor of Economics

For the love of my family

Acknowledgments

Throughout my time at Boston University, I am extremely lucky to receive a tremendous amount of help and support, without which the dissertation can not be finished.

I would first like to express my deepest gratitude to my advisors, Zhongjun Qu, Pierre Perron, and Jean-Jacques Forneron for their guidance, patience, and encouragement. It is a wonderful experience to learn from excellent researchers like them. Zhongjun, as my main advisor, played a fundamental role in the first and third chapters of my dissertation. He not only taught me everything about doing research, from seeking topics to choosing appropriate methodologies, but also helped me shape my research habits. He is always available to guide me, even during his sabbatical and vacations. I learned a lot from numerous discussions with him. Pierre inspired my interest in econometrics. In particular, when I thought econometrics was too technical for me, he showed me how to think about the subject in an intuitive way. I gained a lot of insights from his sharp comments on my research and the econometrics profession. I am also heavily indebted to him for his support in bringing me into the Ph.D. program. JJ arrived at BU when I was a senior Ph.D. student and I can still remember how brilliant his job talk was. It is truly fortunate to have a fantastic researcher like him as my advisor. He always offers to discuss research and significantly improved my job interview skills, even though he is extremely busy being a junior faculty.

My sincere thanks also go to the other members of the econometrics group, especially Iván Fernández-Val and Hiroaki Kaido. I have benefited substantially from taking their courses and hearing their discussions at the BU Econometrics Seminar. I will miss all the times I spent with my fellow econometrics students, including Undral Byambadalai, Shuowen Chen, Junwen Lu, Siyi Luo, Anlong Qin, Yuan Tian, Guang

Zhang, and Yi Zhang. Alessandro Casini has my special appriciation. We've known each other since the first day sitting in Pierre and Zhongjun's class when I was still a master student in statistics while he was already in the Ph.D. program. Common interests lead us to take many classes together, in which we often work as a group, and eventually turn into our coauthorship.

Besides, I gratefully acknowledge the gigantic help of other BU faculties in Economics as well as in Mathematics and Statistics. My special attitude goes to David Fried, who introduced me to proof-based mathematics, and Murad Taqqu, who showed me how to rigorously think about probability.

Last but not least, I would like to express my deepest gratitude to my family and especially my wife Buqu Gao, who has been taking care of our dear daughter Qiqi in my absence for almost 18 months. This dissertation would not have been possible without their unlimited indulgence, warm love, and endless support.

THREE ESSAYS ON PREDICTIVE REGRESSION, LOW-FREQUENCY VARIATION, AND DYNAMIC STOCHASTIC GENERAL EQUILIBRIUM MODELS

TAOSONG DENG

Boston University, Graduate School of Arts and Sciences, 2021

Major Professor: Zhongjun Qu
Professor of Economics

ABSTRACT

This dissertation investigates several important issues related to filtering, estimation, and inference in time series econometrics. The applied focus is on financial and macroeconomic models that include predictive regressions and dynamic stochastic general equilibrium models as prominent examples.

Chapter 1 studies inference in predictive regression with a nearly integrated predictor. Conventional tests for predictive regressions exhibit substantial size distortions while existing valid inference procedures usually require multiple steps for their implementation. I propose a simple procedure using an augmented regression that requires only one step to test the coefficients in a predictive regression with a nearly integrated predictor. I prove that the usual t-test using conventional standard normal critical values is conservative. Furthermore, to address the situation where the predictive test is uninformative because of possible outlying events or regime changes, I propose a class of robust tests and study their asymptotic properties. In the empirical application, I find considerable evidence of the predictability of NYSE/AMEX returns using nearly integrated predictors, such as the log dividend-price ratio or the

log earning-price ratio.

Chapter 2 (joint with Alessandro Casini and Pierre Perron) establishes theoretical results about the low frequency contamination induced by general nonstationarity for estimates such as the sample autocovariance and the periodogram, and hence deduces consequences for heteroskedasticity and autocorrelation robust (HAR) inference. We show that for short memory nonstationarity data these estimates exhibit features akin to long memory due to low frequency contamination, to which, however, estimates based on nonparametric smoothing over time are robust. The theoretical findings are further confirmed by simulations. Since inconsistent long-run variance (LRV) estimation tends to be inflated when the data are nonstationary, HAR tests based on LRV can suffer from low frequency contamination, being more undersized with lower power than those based on HAC, whereas tests based on the recently introduced double kernel HAC estimator do not.

The last chapter (joint with Zhongjun Qu) develops a new particle filter for dynamic stochastic general equilibrium (DSGE) models by mapping the state vector into two subvectors: a subvector whose components are observed and a subvector whose components are latent. By only sampling and propagating particles of the latent variables, we avoid the need to introduce measurement errors, a convenient but questionable practice. For implementation, we propose to approximate the observables' density conditional on the latent variables using series expansions. As an important feature, the new filter also allows us to study singular DSGE models using the composite likelihood, therefore providing a unified treatment of both singular and nonlinear DSGE models.

Contents

List of Tables

List of Figures

List of Abbreviations

ACF	.....	Autocorrelation Function
AMEX	.....	American Stock Exchange
AR	.....	Autoregressive
ARCH	.....	Autoregressive Conditional Heteroskedasticity
ARFIMA	.....	Autoregressive Fractionally Integrated Moving Average
CRSP	.....	Center for Research in Security Prices
DGP	.....	Data Generating Process
DK-HAC	.....	Double Kernel Heteroskedasticity and Autocorrelation Consistent
DSGE	.....	Dynamic and Stochastic General Equilibrium
GLS	.....	Generalized Least Squares
HAC	.....	Heteroskedasticity and Autocorrelation Consistent
HAR	.....	Heteroskedasticity and Autocorrelation Robust
LRV	.....	Long Run Variance
MCMC	.....	Monte Carlo Markov Chain
NBER	.....	National Bureau of Economic Research
NYSE	.....	New York Stock Exchange
OLS	.....	Ordinary Least Squares
SLS	.....	Segmented Locally Stationary
S&P	.....	Standard and Poor
SV	.....	Stochastic Volatility
WWII	.....	World War II

Chapter 1

Inference in Predictive Regression with a Nearly Integrated Predictor

1.1 Introduction

It is well known that traditional predictive regression tests often produce spurious predictability because their empirical sizes far exceed their nominal sizes: the resulting t-statistic values are often significant, typically greater than two or even three. This might lead researchers to mistakenly conclude that an outcome variable in the next period could be forecasted by a predictor in the current period. Mankiw and Shapiro (1986) and Stambaugh (1986) are the first to indicate that the usual t-value produced by such regressions would reject the null hypothesis too often. The inference error is mainly caused by two reasons: first, when a predictor is highly persistent, the first-order approximation can fail. Second, when the innovations of a predictor and outcome variables are highly correlated, the endogeneity problem is severe. See Elliott and Stock (1994) and Cavanagh et al. (1995) for theoretical results and Campbell and Yogo (2006) for simulation results.

A common approach to deal with the endogeneity problem is by projecting the disturbance term in the outcome variable onto the innovation in the predictor variable. Using this approach, Stambaugh (1986) observes that there exists a close relationship between the downward bias in the ordinary least squares (OLS) estimator for the autoregressive coefficient of a predictor and the upward bias in the OLS es-

timator for the predictive coefficient of a predictive regression. Relying upon this relationship, Stambaugh (1999) achieves first-order bias-correction by drawing on previous work on bias-corrected estimators for autoregressive regressions. Still employing the connection between the two OLS estimators, Lewellen (2004) suggests a conservative second-order bias-corrected estimator by restricting the degree of the predictor's persistence to be less than unity. Thereafter, he verifies the validity of the corresponding t-test procedure using conventional asymptotics. Rather than exploiting the relationship between the two OLS estimators, Amihud and Hurvich (2004) achieves second-order bias-correction for the estimator of the predictive coefficient by augmenting the original equation with the innovations in the predictor. They derive the asymptotic distribution of their test using the conventional approach. Later, Cai and Wang (2014) provided the asymptotic distribution for their method using the local-to-unity approximation. Finally, Campbell and Yogo (2006) propose a robust inference procedure based on the Bonferroni correction, and it has become popular among financial practitioners.

Motivated by these methods, I propose a simple procedure by directly estimating an augmented regression. Different from the existing methods that involve estimation and inference in multiple steps, my method only requires a single step to conservatively estimate the predictive parameter, probably representing the easiest way to produce valid inferences. Nevertheless, my method has good power in practice when a nearly integrated predictor is used. In my application, the inference results are generally similar to, if not stronger than, those yielded by the efficient but more complicated Q-test provided by Campbell and Yogo (2006). To address the situation whereby the predictive test becomes uninformative due to possible outlying events or regime changes, I propose a class of robust tests by splitting the full sample into admissible subsamples over which the original procedure are applied. Their asymptotic

and finite sample properties are studied.

Employing a common setting in the previous literature, I assume that the standard predictive regression has the following linear structural form:

$$y_t = \alpha + \beta x_{t-1} + u_t, \tag{1.1}$$

$$x_t = \gamma + \rho x_{t-1} + v_t, \tag{1.2}$$

for $t = 1, 2, \ldots, T$. It has the usual state space form, where the first and second equations describe the forecast and transition equations, respectively. For the forecast equation (1.1), let y_t be the outcome variable in period t, x_{t-1} a predictor observed in period $t-1$, and u_t the remaining forecast error. As specified in the transition equation (1.2), the predictor x_t is commonly formulated by an autoregressive model with order 1, or AR(1). Notably, v_t is the innovation in x_t. The unknown parameter ρ captures the degree of persistence in x_t. If $|\rho| < 1$ and fixed, x_t is integrated of order zero, or $I(0)$. If $\rho = 1$, x_t is integrated of order one, or $I(1)$. If $|\rho| > 1$, x_t is explosive. Herein, I exclude the explosive case and assume $|\rho| \leq 1$. I am particularly interested in the case where $\rho \approx 1$, i.e., when a predictor is nearly integrated.

Generally, I can allow the innovation term v_t in (1.2) to have short-run dynamics that are captured by an AR(p) process, writing

$$\Phi(L)v_t = e_t,$$

where $\Phi(L)v_t = \sum_{j=0}^{p} \Phi_j L^j$, with $\Phi_0 = 1$ and $\Phi(1) \neq 0$. Moreover, the vector $(u_t, e_t)'$ is a martingale difference sequence with the conditional covariance matrix Σ_t, which can be time-varying in principle. Recently, Casini (2020) has proposed double-kernel heteroskedasticity and autocorrelation (DK-HAC) estimators using a segmented locally stationary framework, thereby showing that the use of nonparametric smoothing over time can substantially improve power for tests based on the HAC estimators un-

der such a nonstantionary environment. Further, Casini et al. (2020) studies the consequences for heteroskedasticity and autocorrelation robust (HAR) inference under such a setting. However, to facilitate my analysis, I restrict that $(u_t, v_t)'$ is a martingale difference sequence with a time-invariant conditional covariance matrix. Throughout the paper, the following conditions are assumed for innovations driving the data generating processes (1.1) and (1.2), unless otherwise stated.

Assumption 1.1.1 (Martingle Difference) Denote the two dimensional process $w_t = (u_t, v_t)'$. Let $\mathcal{F}_t = \{w_s | s \leq t\}$ be the induced filtration. Assume w_t is a martingale difference sequence with finite fourth moments satisfying

1. $E[w_t|\mathcal{F}_{t-1}] = 0$,
2. $\sup_t E[u_t^4], \sup_t E[v_t^4], E[x_0^2] < \infty$,
3. $E[w_t w_t'] = \Sigma$, with

$$\Sigma = Cov(u_t, v_t) = \begin{bmatrix} \sigma_u^2 & \theta\sigma_u\sigma_v \\ \theta\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix},$$

where θ denotes the correlation coefficient $Cor(u_t, v_t)$, which can be further assumed to be negative for the sake of practical relevance.

The assumption that the correlation coefficient θ between u_t and v_t is negative is without loss of generality because redefining the predictor variable x_t as $-x_t$ flips the sign. In the empirical finance literature, the outcome variable y_t is usually the stock excess return in the t-th period. The predictor x_{t-1} could be a financial ratio, such as the log of the dividend-price ratio (d-p) or log of earning-price ratio (e-p) in the $(t-1)$-th period. Under this setting, as price enters both the outcome variable $y_t = \log p_t - \log p_{t-1}$ and the predictor $x_t = \log d_t - \log p_t$ and further affects their values in an opposite way, the correlation coefficient θ is very negative (see Table 1.3).

I am mainly interested in whether the lagged regressor x_{t-1} can predict the outcome variable y_t. Therefore, the parameter β is the unknown parameter of interest.

I say that x_{t-1} is able to predict y_t if $\beta \neq 0$. I focus on the null hypothesis $\beta = 0$ versus the one-sided composite alternative $\beta > 0$, following the previous literature.

The rest of the paper proceeds as follows: the existing popular estimation and inference procedures for the predictive regression model are discussed in the Section 2. In Section 3, I review the so-called local-to-unity asymptotics used here. My main inference procedure for testing predictability, using the framework of augmented regressions, is proposed in Section 4. In Section 5, concerning possible outlying events and regime changes, I propose a family of robust tests under model instability. The finite sample properties of my proposed tests are evaluated in Section 6. An empirical application of my methods, along with a comparison of the popular Campbell and Yogo (2006) Bonferroni procedure is presented in Section 7, while the conclusions are given in Section 8.

1.2 Inference with a Nearly Integrated Predictor

Direct runs of the first regression (1.1) produce large t-values, typically greater than two and sometimes greater than three (see Table 1.3 in Section 7). New researchers unfamiliar with this setting might mistakenly conclude that certain predictors, such as financial ratios could indeed strongly forecast outcome variables such as stock returns. However, Mankiw and Shapiro (1986) and Stambaugh (1986) show that when the contemporaneous correlation coefficient θ is close to -1 and the autoregressive coefficient ρ is close to 1, the usual t-test of the null hypothesis that $\beta = 0$ leads to rejections too often. Figure 1·1 displays time series plots of the log dividend-price ratio for the NYSE/AMEX value-weighted index and the log smoothed earnings-ratio for the S&P 500 index at monthly frequency, from 1926 to 2018. (The shaded regions are the recession periods dated by the NBER.) Note that both valuation ratios are very persistent. Heuristically, the predictor's variance $\sigma_x^2 = \sigma_v^2/(1-\rho^2)$ diverges to infinity

as $\rho \to 1$; hence, the traditional first-order asymptotics provides poor approximations in finite sample. Consequently, the usual tests can have large size distortions.

However, recall that the martingale difference assumption only specifies that u_t is serially uncorrelated and $E(u_t|x_{t-1}, x_{t-2}, \ldots) = 0$. In fact, the common assumption $E(u_t|\ldots, x_{t+1}, x_t, x_{t-1,\ldots}) = 0$ typically fails to hold in predictive regression settings. This can be easily seen by setting the autoregressive coefficient $\rho = 0$. In this case x_t is highly correlated with v_t and thus u_t. As ρ increases, given the correlation coefficient θ, the correlation between x_t and $u_{t-s}(s > 0)$ through v_{t-s} becomes larger. In practice, if y_t is the excess stock return reflecting price change in period t while x_t is the log dividend-price ratio inversely affected by the price change at the same period, it must be true that $E(u_t|x_t, x_{t-1}) \neq 0$. Therefore, the Gauss-Markov theorem does not apply, and the OLS estimator of β, though consistent, is biased in finite sample.

In fact, the finite sample bias of the OLS estimator $\hat{\beta}$ is proportional to the bias of $\hat{\rho}$, which has been extensively analyzed by, for example, Kendall (1954), White (1961), Dickey and Fuller (1979), Dickey and Fuller (1981), and Evans and Savin (1984). Denote the $T \times 1$ regressor vectors $\iota = (1, 1, \ldots, 1)'$ and $x_{-1} = (x_1, \ldots, x_T)'$, the regressor matrix $X = (\iota, x_{-1})$, and parameters $b = (\alpha, \beta)', p = (\gamma, \rho)'$. For OLS estimators $\hat{b}$ and $\hat{p}$, we have

$$\hat{b} = (X'X)^{-1}X'y = b + (X'X)^{-1}X'u$$
$$\hat{p} = (X'X)^{-1}X'x = p + (X'X)^{-1}X'v.$$

We can decompose the forecast error u by projecting it onto the innovation v, obtaining

$$u_t = \delta v_t + e_t, \tag{1.3}$$

where $\delta = Cov(u, v)/Var(v)$. It follows that

$$\hat{b} - b = \delta(\hat{p} - p) + (X'X)^{-1}X'e,$$

or

$$\hat{\beta} - \beta = \delta(\hat{\rho} - \rho) + \eta, \tag{1.4}$$

where η is the second entry of $(X'X)^{-1}X'e$. That is,

$$\eta = \frac{\sum x_{t-1}^{\mu} e_t}{\sum x_{t-1}^{\mu 2}} \quad \text{with} \quad x_{t-1}^{\mu} = x_{t-1} - T^{-1}\sum_{t=1}^{T} x_{t-1}.$$

Taking expectation gives

$$E\hat{\beta} - \beta = \delta(E\hat{\rho} - \rho). \tag{1.5}$$

The correlation induces an upward bias for $\hat{\beta}$ if $\delta < 0$ (and downward bias if $\delta > 0$) because $\hat{\rho}$ is downward biased in finite samples and can be approximated by

$$E\hat{\rho} - \rho = -\frac{1 + 3\rho}{T} + O(T^{-2}). \tag{1.6}$$

By plugging (1.6) into relationship (1.5), we obtain the first-order bias-corrected estimator given in Stambaugh (1986)

$$\hat{\beta}^s = \hat{\beta} + \frac{\hat{\delta}(1 + 3\hat{\rho})}{T}.$$

However, Stambaugh (1999) shows that the approximation error can be non-trivial even when the sample size is large.

Instead of using the approximation (1.6), Lewellen (2004) directly refers to (1.4) to construct his own bias-corrected estimator. Heuristically, ignoring the error term

η and then computing the parameter value β gives the estimator

$$\hat{\beta}^l = \hat{\beta} - \hat{\delta}(\hat{\rho} - \rho).$$

The term $\hat{\delta}(\hat{\rho} - \rho)$ is interpreted as finite sample bias. Particularly, setting $\rho = 1$ gives the most conservative estimator. Under the assumption that $\rho < 1$, he demonstrates that one could use standard normal distribution critical values to construct a conservative test using the classic asymptotics. Unsurprisingly, when ρ is close to unity, the test has more power than the test proposed in Stambaugh (1999).

To see the nature of this type of estimators, I plug the error decomposition (1.3) into the original forecasting equation (1.1) and obtain an alternative expression

$$y_t = \alpha + \beta x_{t-1} + \delta v_t + e_t. \tag{1.7}$$

Suppose we know v_t. Subtract v_t from both sides to further obtain

$$y_t - \delta v_t = \alpha + \beta x_{t-1} + e_t. \tag{1.8}$$

Then, the infeasible estimator $\hat{\beta}^i$ is the second element of $\hat{b}^i$ with

$$\hat{b}^i - b = (X'X)^{-1}X'e = \hat{b} - b - \delta(\hat{p} - p).$$

This shows that Stambaugh (1999) and Campbell and Yogo (2006)'s estimators are deeply connected to regression (1.8) because the infeasible Q-statistic

$$Q(\beta) = \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}[y_t - \delta(x_t - \rho x_{t-1}) - \beta x_{t-1}]}{\sigma_e(\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}} = \frac{\sum_{t=1}^{T} x_{t-1}^{\mu} e_t}{\sigma_e(\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}$$

proposed in Campbell and Yogo (2006) could be interpreted as the t-statistic obtained from regression (1.8). Therefore, the t-statistic from this equation is optimal in the sense that it is the uniformly most powerful conditional on the ancillary statistic

$\sum_{t=1}^{T} x_{t-1}^{\mu 2}$, where $x_{t-1}^{\mu} = x_{t-1} - T^{-1}\sum_{t=1}^{T} x_{t-1}$. In practice, v_t can be obtained from (1.2) once we know the autoregressive coefficient ρ. Thus, Campbell and Yogo (2006) make the infeasible test feasible by applying the Bonferroni method, following Cavanagh et al. (1995). The confidence interval for the nuisance parameter ρ is obtained by inverting a unit root test for the predictors, following Stock (1999), using the GLS detrending procedure from Elliott et al. (1996). Particularly, by restricting $\rho = 1$, this procedure nests the conservative estimation and inference procedures conducted by Lewellen (2004).

Finally, inspired by the expression (1.7), Amihud and Hurvich (2004) obtain their estimator $\hat{\beta}^a$ by running the augmented regression

$$y_t = \alpha + \beta x_{t-1} + \delta \hat{v}_t + \epsilon_t,$$

where $\hat{v}_t$ is the residual from regression (1.2). Cai and Wang (2014) then establish the asymptotic distribution for $\hat{\beta}^a$ using local-to-unity asymptotics, which is not pivotal and depends on nuisance parameters. Consequently, one is unable to directly use it to conduct inference, and one needs to estimate the unknown parameters first.

The procedure proposed here follows this prior literature, and I directly work with the augmented equation (1.7). Together with a restriction procedure for ρ, I am able to conduct valid inferences for predictability. In the simplest case restricting $\rho = 1$, I provide a fairly simple method that only requires one step to conduct conservative but valid inferences. Moreover, the t-statistic proves to be asymptotically equivalent to Campbell and Yogo (2006) infeasible Q-statistic due to the nearly integrated nature of the predictor.

1.3 Local-to-Unity Asymptotics

Instead of resorting to the exact finite-sample theory under the normality assumption to study estimators in the setting of predictive regression, e.g., Evans and Savin (1981), Evans and Savin (1984), and Stambaugh (1999), I employ the local-to-unity asymptotic theory, following, e.g., Phillips (1987), Perron (1991), Elliott and Stock (1994), and Cavanagh et al. (1995).

There are two practical advantages of using local-to-unity asymptotics over the exact Gaussian theory. The first is that the asymptotic distribution of test statistics does not depend on sample size; thus, we do not need to recompute the critical values for relevant test statistics according to different sample sizes. The second is that large sample theory does not require the strict normality assumption, allowing for short-run dynamics in the predictors and heteroskedasticity in innovations.

Local-to-unity asymptotics is an asymptotic framework where the largest autoregressive root is modeled as $\rho = \exp(c/T) \approx 1 + c/T$, with $c = T(\rho - 1)$ is a fixed constant as T becomes arbitrarily large. For example, if $c = -10$, then $\rho = 0.80$ for $T = 50$, $\rho = -0.9$ for $T = 100$, and $\rho = 0.98$ for $T = 500$. Within this framework, the asymptotic distribution theory is continuous even when x_t is $I(1)(c = 0)$, thereby allowing x_t to be nearly integrated ($c < 0$) or even nearly explosive ($c > 0$). An important feature of the nearly integrated case is that the sample moments (e.g., mean and variance) of the process x_t do not converge to constant probability limits but to functionals of a diffusion process when appropriately scaled.

Let $W(r)$ be the standard Brownian motion. Let $J_c(r)$ be the diffusion process induced by the stochastic differential equation $dJ_c(r) = cJ_c(r)dr + dW(r)$ with initial condition $J_c(0) = 0$. Denote its centered counterpart $J_c^{\mu}(r) = J_c(r) - \int J_c(r)dr$, where the integration is over $[0, 1]$ unless otherwise noted. Let $\Rightarrow$ denote weak convergence in the space of $D[0, 1]$ of cadlag functions.

To obtain some intuitions, assume $\gamma = x_0 = 0$. Since $\rho = \exp(c/T)$,

$$x_t = e^{c/T} x_{t-1} + v_t = \sum_{j=1}^{t} e^{c(t-j)/T} v_t,$$

the invariance principle gives

$$J_c(r) = \int_0^r e^{c(r-s)} dW(s) = W(r) + c \int_0^r e^{c(r-s)} W(s) ds.$$

Alternatively, since $\rho \approx 1 + c/T$,

$$\Delta x_t = c \frac{x_{t-1}}{T} + v_t.$$

yielding

$$dJ_c(r) = cJ_c(r)dr + dW(r). \tag{1.9}$$

Solving this stochastic differential equation gives the same result. Now, I list some weak convergence results obtained from Phillips (1987) and Cavanagh et al. (1995) that will be used later.

Lemma 1.3.1 (Weak Convergence) *Suppose the Martingale Difference Assumption 1.1.1 holds and $\rho = 1 + c/T$. Let*

$$x_{t-1}^{\mu} = x_{t-1} - T^{-1} \sum_{t=1}^{T} x_{t-1}, \quad J_c^{\mu} = J_c(r) - \int_0^1 J_c(s) ds.$$

The following limits hold jointly.

1. $T^{-3/2} \sum_{t=1}^{T} x_{t-1}^{\mu} \Rightarrow \sigma_v \int_0^1 J_c^{\mu}(r) dr.$
2. $T^{-2} \sum_{t=1}^{T} x_{t-1}^{\mu 2} \Rightarrow \sigma_v^2 \int_0^1 J_c^{\mu}(r)^2 dr.$
3. $T^{-1} \sum_{t=1}^{T} x_{t-1}^{\mu} v_t \Rightarrow \sigma_v^2 \int_0^1 J_c^{\mu}(r) dW(r).$

Recall the decomposition $u_t = \delta v_t + e_t$ such that $Cov(v_t, e_t) = 0$. Then, $u_t =$

$\theta\frac{\sigma_u}{\sigma_v}v_t + e_t$ and $\sigma_e^2 = (1-\theta^2)\sigma_u^2$. Under the null hypothesis that $\beta = \beta_0$, simple algebra gives the t-statistic obtained from regression (1.1) as follows:

$$\begin{aligned} t_{OLS}(\beta_0) &= \frac{\hat{\beta}_{OLS} - \beta_0}{se(\hat{\beta}_{OLS})} = \frac{T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu} u_t}{\hat{\sigma}_u \sum_{t=1}^{T}\left(\sum_{t=1}^{T} x_{t-1}^{\mu 2}\right)^{1/2}} \\ &= \theta\frac{\sigma_u}{\hat{\sigma}_u}\frac{T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu} v_t}{\sigma_v \sum_{t=1}^{T}\left(\sum_{t=1}^{T} x_{t-1}^{\mu 2}\right)^{1/2}} + (1-\theta^2)^{1/2}\frac{\sigma_u}{\hat{\sigma}_u}\frac{T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu} e_t}{\sigma_e \sum_{t=1}^{T}\left(\sum_{t=1}^{T} x_{t-1}^{\mu 2}\right)^{1/2}}. \end{aligned}$$

By construction,

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{[Tr]} v_t \\ T^{-1/2}\sum_{t=1}^{[Tr]} e_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_v W(r) \\ \sigma_e G(r) \end{bmatrix},$$

where W and G are two independent standard Brownian motions. Thus we have

$$T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu} u_t \Rightarrow \sigma_u\sigma_v\left(\theta\int_0^1 J_c^{\mu}(r)dW(r) + (1-\theta^2)^{1/2}\int_0^1 J_c^{\mu}(r)dG(r)\right).$$

It follows the usual t-statistic has the limiting distribution

$$\begin{aligned} t_{OLS}(\beta_0) &\Rightarrow \theta\frac{\int_0^1 J_c^{\mu}(r)dW(r)}{(\int_0^1 J_c^{\mu}(r)^2 dr)^{1/2}} + (1-\theta^2)^{1/2}\frac{\int_0^1 J_c^{\mu}(r)dG(r)}{(\int_0^1 J_c^{\mu}(r)^2 dr)^{1/2}} \\ &= \theta q(c) + (1-\theta^2)^{1/2} Z, \end{aligned}$$

where

$$q(c) = \frac{\int_0^1 J_c^{\mu}(r)dW(r)}{\left(\int_0^1 J_c^{\mu}(r)^2 dr\right)^{1/2}}$$

and Z is a standard normal random variable independent of W and thus $q(c)$ (see Elliott and Stock (1994)). Note that this usual t-statistic is not asymptotically pivotal. That is, its asymptotic distribution depends on the nuisance parameters θ and c through the nonstandard distribution $q(c)$, which makes the test infeasible.

In general, the asymptotic distribution of the t-statistic is nonstandard because of its dependence on $q(c)$. However, the t-statistic returns to the usual standard normal distribution in the special case $\theta = 0$. Likewise, the t-statistic should be approximately normal when $c \ll 0$ because first-order asymptotics are a satisfactory approximation when the predictor is stationary. Formally, Phillips (1987) shows that $q(c) \Rightarrow \tilde{Z}$ as $c \to -\infty$, where $\tilde{Z}$ is a standard normal random variable independent of Z.

The size distortion ensues from the fact that the distribution of $q(c)$ is skewed to the left, which in turn skews the distribution of the t-statistic to the right when $\theta < 0$. This causes a right-tailed t-test that uses conventional critical values to over-reject, and a left-tailed test to under-reject. The larger the absolute value of the persistence of a predictor c or the correlation coefficient θ, the larger the size distortion. Campbell and Yogo (2006) plot the asymptotic size of the nominal 5% one-sided t-test, i.e., the plot of

$$p(c, \theta; 0.05) = \Pr\left(\theta q(c) + (1 - \theta^2)^{1/2} Z > z_{0.05}\right),$$

where $z_{0.05} = 1.645$ denotes the 95th percentile of the standard normal distribution. Moreover, given the correlation coefficient θ, they tabulate the values of $c \in (c_{min}, c_{max})$ for which the size of the right-tailed t-test exceeds 7.5%; thus, it can be used to construct a pretest to decide whether inference based on the conventional t-test is sufficiently reliable.

1.4 Testing Predictability using an Augmented Regression

To motivate the proposed estimator, I consider the infeasible estimator obtained from the aforementioned infeasible augmented regression

$$y_t = \alpha + \beta x_{t-1} + \delta v_t + e_t, \tag{1.10}$$

where $v_t = x_t - \rho x_{t-1}$ and $e_t = u_t - \delta v_t$.

Theorem 1.4.1 *Suppose the errors (u_t, v_t) are serially independent and identically distributed as bivariate normal. The infeasible estimator $\tilde{\beta}$ is unbiased, i.e., $E[\tilde{\beta}] = \beta$.*

The estimation and inference would be easy if we knew v_t or ρ. However, v_t is not directly observable, and thus the estimator is infeasible. We may construct a proxy $\tilde{v}_t = x_t - \tilde{\rho} x_{t-1}$ for some specified $\tilde{\rho}$. As the nuisance parameter ρ is nearly integrated, we can reparameterize $\rho = 1 + c/T$ and, analogously, $\tilde{\rho} = 1 + \tilde{c}/T$. Then we can rewrite (1.10) as

$$y_t = \alpha + \beta x_{t-1} + \delta \tilde{v}_t + \delta(v_t - \tilde{v}_t) + e_t,$$

where $v_t - \tilde{v}_t = (\tilde{c} - c)x_{t-1}/T$. Hence, we have the following feasible augmented regression

$$y_t = \alpha + \beta x_{t-1} + \delta \tilde{v}_t + \epsilon_t. \tag{1.11}$$

Note that c is not identified. Nevertheless, in practice, one can simply set $\tilde{c} = 0$ or $\tilde{\rho} = 1$, yielding $\tilde{v}_t = x_t - x_{t-1} = \Delta x_t$. Therefore, one can use the feasible augmented regression

$$y_t = \alpha + \beta x_{t-1} + \delta \Delta x_t + \epsilon_t.$$

Theorem 1.4.2 *Suppose the innovations* $(u_t, v_t)'$ *satisfy the Martingale Difference Assumption 1.1.1. Let* (x_t, y_t) *be generated according to the predictive regression* (1.1)-(1.2)*. Suppose* $\rho = 1 + c/T$ *and* $\tilde{\rho} = 1 + \tilde{c}/T$*. Obtained from* (1.11)*,* $t(\beta_0)$ *is denoted as the t-statistic that tests* $\beta = \beta_0$*. Then*

$$\begin{aligned} t(\beta_0) &\Rightarrow \frac{\sqrt{1-\theta^2}\int_0^1 J_c^{\mu}(r)dG(r) + (\tilde{c}-c)\theta\int_0^1 J_c^{\mu}(r)^2dr}{\sqrt{1-\theta^2}(\int_0^1 J_c^{\mu}(r)^2dr)^{1/2}} \\ &= Z + \frac{(\tilde{c}-c)\theta}{(1-\theta^2)^{1/2}}\left(\int_0^1 J_c^{\mu}(r)^2dr\right)^{1/2}, \\ t(\delta_0) &\Rightarrow \tilde{Z}, \end{aligned}$$

where Z *and* $\tilde{Z}$ *are standard normal variables.*

Remark 1.4.1 For fixed c, the extent to which the t-statistic on β is located below the standard normal distribution is controlled by $\tilde{c}$. Consequently, the t-test using standard normal critical values is conservative yet valid for any predictor with $c \leq \tilde{c}$. For the special case where $c = \tilde{c}$, the distribution of the t-statistic aligns with the normal; thus, the test is correctly sized. In particular, we can set $\tilde{c} = 0$ to make the t-test conservative for any stationary or integrated predictor ($c \leq 0$ or $\rho \leq 1$).

Remark 1.4.2 The rate of convergence of $\hat{\beta}$ is $O(T^{-1})$, faster than $O(T^{-1/2})$, i.e., the rate of convergence of $\hat{\delta}$, thereby leading to the annihilation of the diagonal entry of the scaled design matrix $\Gamma_T^{-1}\mathcal{X}X\Gamma_T^{-1}$, whose expression is given in the appendix. As a consequence, the t-statistic on δ is standard normal and free from the nuisance parameter c. Furthermore, the distribution of the t-statistic is asymptotically equivalent to the infeasible Q-statistic proposed by Campbell and Yogo (2006). Hence, choosing $\tilde{c} = 0$ makes the t-test asymptotically equivalent to the procedure proposed in Lewellen (2004).

Because of the high degree of persistence, the available predictors often appear to be an $I(1)$ process, even if they are actually $I(0)$. However, the outcome variable, for example, the excess stock return, often has a very low level of autocorrelation. On the one hand, Figure 1·2 presents a time series of the excess stock return for the NYSE/AMEX value-weighted index at monthly frequency. It evidently shows a

starkly different pattern from the predictor plot of Figure 1·1. Thus, we would expect the OLS estimation to fail. On the other hand, Figure 1·3 is the time series of the first difference of the corresponding financial ratio, which commoves with the outcome variable because of the endogeneity. Taking advantage of the correlation information between the forecast error u_t and innovation v_t, we can alleviate the endogeneity problem by including the first difference of the persistent predictor in my regression.

Recall that the local-to-unity approximation crucially assumes that the value of the parameter c is close to zero, i.e., the predictor is nearly integrated. A legitimate concern is whether the proposed test could produce spurious inferences when a stationary predictor is used. Therefore, the validity of the conservative test should be evaluated when $c \to -\infty$ as well. To do this, we resort to the joint moment generating function $m(u, v)$ of the functional pairs $\left(\int_0^1 J_c(r)dW(r), \int_0^1 J_c(r)^2 dr\right)$. But the joint weak convergence result holds:

$$\left(T^{-1}\sum_{t=1}^{T} x_{t-1}u_t, T^{-2}\sum_{t=1}^{T} x_{t-1}^2\right) \Rightarrow \left(\int_0^1 J_c(r)dW(r), \int_0^1 J_c(r)^2 dr\right),$$

given that u_t satisfies the conditions in the Martingale Difference Assumption 1.1.1 and $\sigma_u = 1$. Thus I refer to the joint moment generating function $M(u, v)$ of $\left(\sum_{t=1}^{T} x_{t-1}u_t, \sum_{t=1}^{T} x_{t-1}^2\right)$ first.

Lemma 1.4.3 *Suppose $x_t = \alpha x_{t-1} + u_t$, where $\alpha < 1$ and u_t is independent identically $N(0, \sigma^2)$ and $x_0 = 0$. The joint moment generating function $M(u, v)$ of $(\sum_{t=1}^{T} x_{t-1}u_t, \sum_{t=1}^{T} x_{t-1}^2)$ has the following expression*

$$\begin{aligned} M(u, v) &= E\exp\left\{u\left(\sum_{t=1}^{T} x_{t-1}u_t\right) + v\left(\sum_{t=1}^{T} x_{t-1}^2\right)\right\} \\ &= E\exp\left\{u\left(\sum_{t=1}^{T} x_{t-1}x_t - \alpha\sum_{t=1}^{T} x_{t-1}^2\right) + v\left(\sum_{t=1}^{T} x_{t-1}^2\right)\right\} \\ &= \left\{\frac{1-s}{r-s}r^T + \frac{1-r}{s-r}s^T\right\}^{-1/2}, \end{aligned}$$

where r, s *are roots of the equation* $x^2 - px + q^2$, *with* $p = 1 + \alpha^2 + 2\alpha u - 2v$ *and* $q = -(\alpha + u)$.

To study the joint behaviors of the functional pairs $\left(\int_0^1 J_c(r)dW(r), \int_0^1 J_c(r)^2 dr\right)$, the limiting behavior of the joint moment generating function $m_T(u, v)$ of $(T^{-1}\sum_{t=1}^T x_{t-1}u_t, T^{-2}\sum_{t=1}^T x_{t-1}^2)$ should be investigated instead. Hence, correct scaling should be applied first and then let $T \to \infty$, reaching the targeting joint moment generating function

$$m(u, v) = \lim_{T\to\infty} m_T(u, v) = \lim_{T\to\infty} M(uT^{-1}, vT^{-2}).$$

Remark 1.4.3 The assumption that u_t is independently identically normal serves as an apparatus for deriving the targeting joint moment generating function $m(u, v)$ of $\left(\int_0^1 J_c(r)dW(r), \int_0^1 J_c(r)^2 dr\right)$. Once the correct expression is obtained, we can relax u_t to be martingale difference sequence by simply referring to the functional central limit theorem or invariance principle.

It turns out that

$$L(u, v) = \lim_{c\to-\infty} m[(-2c)^{1/2}u, (-2c)v] = \exp\left(\frac{u^2}{2} + v\right).$$

Thus, I have the following results, as given in Phillips (1987).

Theorem 1.4.4 *Suppose that* $J_c(r)$ *is an Ornstein-Uhlenbeck process generated by a Brownian motion according to* (1.9). *As* $c \to -\infty$*:*

1. $(-2c)\int_0^1 J_c(r)^2 dr \to 1$
2. $(-2c)^{1/2}\int_0^1 J_c(r)dW(r) \Rightarrow N(0, 1)$,

The centered counterpart of this theorem can also be easily obtained, as presented in corollary 1.4.5.

Corollary 1.4.5 *Suppose that* $J_c(r)$ *is an Ornstein-Uhlenbeck process generated by a Brownian motion according to* (1.9). *As* $c \to -\infty$*:*

1. $(-2c)\int_0^1 J_c^\mu(r)^2 dr \to 1$.
2. $(-2c)^{1/2}\int_0^1 J_c^\mu(r) dW(r) \Rightarrow N(0,1)$.

Using these stochastic bounded terms appearing in the asymptotic distribution of the proposed test, the test can, thus, be easily validated even if the predictor is far from the nearly integrated case.

Theorem 1.4.6 *Suppose innovations $(u_t, v_t)'$ satisfy the Martingale Difference Assumption 1.1.1. Let (x_t, y_t) be generated according to the predictive regression* (1.1)-(1.2). *Given $c \leq 0$ and $\theta \leq 0$, under the null hypothesis that $\beta = \beta_0$, we have* $\lim_{T\to\infty} \Pr(t(\beta_0) > x) \leq \Pr(Z > x)$.

Therefore, my test is conservative, independent of the persistence level of the predictors used. That is, even if using the test for a predictive regression with a less persistent predictor could lead to a loss of power, this approach will not produce spurious predictability. By including a proxy variable that can take away endogeneity and possibly some predictive element, which is usually the first difference of the persistent predictor, a conservative testing procedure is established. The critical values are just standard normal, which makes the testing procedure fairly easy.

The asymptotic distribution of the t-statistic is plotted against the standard normal distribution in Figure 1·4. Apparently, when $c = 0$, the two distributions align with each other. As the magnitudes of c and θ increases, the test becomes more conservative. Finally, to analyze the power of my proposed t-test, the following weak convergence result is derived under the local alternative.

Theorem 1.4.7 *Suppose innovations $(u_t, v_t)'$ satisfy the Martingale Difference Assumption 1.1.1. Let (x_t, y_t) be generated according to the predictive regression* (1.1)-(1.2). *Under the local alternative $\beta = \beta_0 + b/T$, where $b > 0$, we have*

$$t(\beta_0) \Rightarrow Z + \left(\frac{(\tilde{c} - c)\theta}{(1-\theta^2)^{1/2}} + \frac{b}{(1-\theta^2)^{1/2}}\right)\left(\int_0^1 J_c^\mu(r)^2 dr\right)^{1/2}.$$

Remark 1.4.4 The term $\frac{(\tilde{c}-c)\theta}{(1-\theta^2)^{1/2}}$ is the source of the power-loss, where $\tilde{c} = 0$ is usually chosen for nearly integrated predictors. The term involving the parameter $b = T(\beta - \beta_0)$ will need to be sufficiently large to dominate the first term so that the test can reject the null hypothesis when the alternative is true.

Remark 1.4.5 The local power comparison between the t-statistic and Q-statistic proposed by Campbell and Yogo (2006) is shown in Figure 1·5. Evidently, when the predictor is nearly integrated, my t-test generally exhibits more power than the Q-test, particularly for the unit root case $c = 0$. This is because the Q-statistic involves constructing Bonferroni intervals for the unknown parameter c, leading to a loss of power. However, as c moves toward the negative direction, the power of the Q-test starts to outperform my t-test that sets $\tilde{c} = 0$.

1.5 A Family of Robust Tests under Model Instability

In practice, as Stock and Watson (1996a) and Lettau and Van Nieuwerburgh (2008) report, economic and financial time series are subject to smooth or abrupt changes. Structural breaks in predictive regressions are formally discussed in, for example, Paye and Timmermann (2006) and Rapach and Wohar (2006), and they have found strong evidence of model instability. Such events might render the already conservative test proposed in the last section and other existing tests uninformative.

To remedy this issue, I follow Qu (2007) and propose a family of robust tests to determine whether predictability exists in any one or more parts of the sample. Specifically, I test the null hypothesis of $\beta = 0$ against the alternative hypothesis of $\beta > 0$ in some subsamples. The class of tests I aim to propose allows us to uncover the hidden predictability when the system is affected by possible structural changes. To put it in a concrete way, we will first construct a test which allows m structural changes, hence $m+1$ segments. Let $(T_1, \ldots, T_m)$ be a partition of the original sample, with the convention that $T_0 = 0$ and $T_{m+1} = T$. The partition divides the sample

into $m+1$ subsamples, which are denoted as subsample 1 to subsample $m+1$. That is, for $k = 1, \ldots, m+1$, we specify

$$y_t = \alpha_k + \beta_k x_{t-1} + u_t, \tag{1.12}$$

$$x_t = \gamma + \rho x_{t-1} + v_t. \tag{1.13}$$

The null hypothesis is $\beta_k = 0$ for $k = 1, \ldots, m+1$ while the alternative is $\beta_k > 0$ for some k. In our context, we need to evaluate the t-statistics using subsamples and compute linear combinations of the segmented t-statistics.

Let the t-statistic over a typical subsample k be

$$t_k(\beta) = \frac{\hat{\beta}_k - \beta}{se(\hat{\beta}_k)},$$

which can be constructed for different partitions of the sample and we are interested in searching for the ones that contain subsamples with the most evitable predictability. Since we need to consider all possibilities, a specification for the minimum length of each subsample is necessary. Let

$$\Pi_\epsilon = \{(T_1, \ldots, T_m) : |T_{k+1} - T_k| \geq \epsilon T, k = 1, \ldots, m\}$$

be the admissible set of dates splitting a sample and let

$$\Lambda_\epsilon = \{(\lambda_1, \ldots, \lambda_m) : |\lambda_{k+1} - \lambda_k| \geq \epsilon, k = 1, \ldots, m\}$$

be the corresponding admissible set for cut-offs partitioning the unit interval $[0, 1]$. The parameter $\epsilon > 0$ plays the fundamental role in determining the minimum length of a subsample. The partitions restricted then contain subsamples with at least $[\epsilon T]$ observations. If one has prior knowledge about it, power gain could be achieved. For example, if it is believed that breaks are rare, then a large ϵ could be used, through

which power could be improved. Otherwise, a small ϵ should be specified to avoid a situation where the test becomes inconsistent. A rule of thumb is to choose $\epsilon = 0.2$. I use this parameter value to produce critical values for my tests.

As the testing statistics will involve joint convergence results of subsamples, it is useful to first introduce a lemma implied by the property of Wiener processes. Analogous to the sample cut-offs T_k, denote $\lambda_k = [T_k/T]$. These points are used to divide the unit interval $[0, 1]$ into $m + 1$ subintervals for the sake of asymptotics.

Lemma 1.5.1 *Suppose innovations $(u_t, v_t)'$ satisfy the Martingale Difference Assumption 1.1.1 and $e_t = u_t - \delta v_t$. For segments $k = 1, \ldots, m + 1$, we have the following joint weak convergence result:*

$$\begin{bmatrix} T^{-1/2} \sum_{t=T_{k-1}+1}^{T_k} v_t \\ T^{-1/2} \sum_{t=T_{k-1}+1}^{T_k} e_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_v \left(W(\lambda_k) - W(\lambda_{k-1}) \right) \\ \sigma_e \left(G(\lambda_k) - G(\lambda_{k-1}) \right) \end{bmatrix},$$

where W and G are two independent standard Brownian motions.

The lemma is a direct consequence of invariance principle. I now evaluate the joint behavior of the t-statistics over each segment. For ease of notation, denote, given a partition dividing the sample into $m + 1$ segments, denote the demeaned series over each segment,

$$x_{t-1}^{\mu_k} = x_{t-1} - \bar{x}^k = x_{t-1} - \frac{\sum_{t=T_{k-1}+1}^{T_k} x_{t-1}}{T_k - T_{k-1}}.$$

By adjusting the series according to segments, I ensure that the test has power when structural changes in the deterministic component do occur. Denote further the two

bounds

$$s_k(c) = \frac{\sqrt{1-\theta^2}\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)dG(r) - c\theta\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2dr}{\sqrt{1-\theta^2}\left(\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2dr\right)^{1/2}},$$
$$q_k(c) = \frac{\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)dG(r)}{\left(\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2dr\right)^{1/2}},$$

where

$$J_c^{\mu_k} = J_c(r) - (\lambda_k - \lambda_{k-1})^{-1}\int_{\lambda_{k-1}}^{\lambda_k} J_c(s)ds.$$

Theorem 1.5.2 *Suppose the innovations $(u_t, v_t)'$ satisfy the Martingale Difference Assumption 1.1.1. Let (x_t, y_t) be generated according to the predictive regression* (1.12)-(1.13)*. Then, under the null hypothesis, for each segment $k = 1, 2, \ldots, m+1$, the corresponding t-statistic $t_k(\beta_0)$ has the following joint weak convergence result:*

$$t_k(\beta_0) = \frac{T^{-1}\sum_{t=T_{k-1}+1}^{T_k} x_{t-1}^{\mu_k}\epsilon_t}{\hat{\sigma}_e\left(T^{-2}\sum_{t=T_{k-1}+1}^{T_k} x_{t-1}^{\mu_k 2}\right)^{1/2}}$$
$$\Rightarrow \frac{\sqrt{1-\theta^2}\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)dG(r) - c\theta\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2dr}{\sqrt{1-\theta^2}\left(\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2dr\right)^{1/2}} = s_k(c),$$

Moreover, conditional on the path of $(W(r), G(r))$, the following inequalities hold jointly

$$s_k(c) = \frac{\sqrt{1-\theta^2}\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)dG(r) - c\theta\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2dr}{\sqrt{1-\theta^2}\left(\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2dr\right)^{1/2}}$$
$$\leq \frac{\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)dG(r)}{\left(\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2dr\right)^{1/2}} = q_k(c).$$

Remark 1.5.1 I establish a result similar to the previous Theorem 1.4.2. But at this time, the joint bound $\{q_k(c)\}_{k=1}^{m+1}$ of $\{t_k(\beta_0)\}_{k=1}^{m+1}$ is not free from the nuisance parameter c, even though, marginally, its component $q_k(c)$ still is. This is because

the Ornstein-Uhlenbeck process $J_c(r)$ affects the joint behavior of the t-statistic over each segment, say, $t_k(\beta_0)$ and $t_{k+1}(\beta_0)$, through the Brownian motion $dG(r)$.

Now, I introduce the robust tests based on the supremum operator, where the supremum is taken over the admissible partitions.

Theorem 1.5.3 *Suppose innovations $(u_t, v_t)'$ satisfy the Martingale Difference Assumption 1.1.1. Let (x_t, y_t) be generated according to the predictive regression* (1.12)-(1.13)*. For a given $\epsilon > 0$, denote*

$$\operatorname{Sup} T_k = \sup_{\Pi \in \Pi_\epsilon} t_k, \quad \operatorname{Sup} S_k = \sup_{\Lambda \in \Lambda_\epsilon} s_k, \quad \operatorname{Sup} Q_k = \sup_{\Lambda \in \Lambda_\epsilon} q_k.$$

Then under the null hypothesis, for $k = 1, \ldots, m+1$,

$$\operatorname{Sup} T_k \Rightarrow \operatorname{Sup} S_k.$$

Moreover, conditional on the path of $(W(r), G(r))$, the following inequalities hold jointly:

$$\operatorname{Sup} S_k \leq \operatorname{Sup} Q_k.$$

Remark 1.5.2 In practice, $q_k(c)$, and thus $\operatorname{Sup} Q_k$, depends on the parameter c, which is usually unknown. A natural choice is to replace $J_c(r)$ in $q_k(c)$ with $W(r)$ by setting $c = 0$. That is, I can use

$$r_k = q_k(0) = \frac{\int_{\lambda_{k-1}}^{\lambda_k} W^{\mu_k}(r) dG(r)}{\left(\int_{\lambda_{k-1}}^{\lambda_k} W^{\mu_k}(r)^2 dr \right)^{1/2}},$$

to form a feasible distribution bound $\operatorname{Sup} R_k = \sup_{\Lambda \in \Lambda_\epsilon} r_k$, because the simulations (Figure 1·6 and Table 1.1) demonstrate that the cumulative distribution of $\operatorname{Sup} R_k$ always provides an upper envelope for that of $\operatorname{Sup} Q_k(c)$ with $c \leq 0$. Testing with the critical values of $\operatorname{Sup} R_k$ is more conservative than with that of $\operatorname{Sup} Q_k(c)$, but nevertheless valid. In fact, as $c \to -\infty$ ($c \leq -50$), the cumulative distribution of $\operatorname{Sup} Q_k(c)$ converges to a lower bound which is not far from the upper bound $\operatorname{Sup} R_k$.

The joint segmental convergence results facilitates the construction of a predictability test allowing for m changes (and hence $m+1$ segments) using a simple

linear combination as well.

Corollary 1.5.4 *Suppose the innovations* $(u_t, v_t)'$ *satisfy the Martingale Difference Assumption 1.1.1. Let* (x_t, y_t) *be generated according to the predictive regression* (1.1)-(1.2)*. For a given* $\epsilon > 0$*, denote*

$$\operatorname{Sup} T^m = \sup_{\Pi \in \Pi_\epsilon} \sum_{k=1}^{m+1} t_k, \quad \operatorname{Sup} S^m = \sup_{\Lambda \in \Lambda_\epsilon} \sum_{k=1}^{m+1} s_k, \quad \operatorname{Sup} Q^m = \sup_{\Lambda \in \Lambda_\epsilon} \sum_{k=1}^{m+1} q_k.$$

Then, under the null hypothesis,

$$\operatorname{Sup} T^m \Rightarrow \operatorname{Sup} S^m.$$

Moreover, conditional on the path of $(W(r), G(r))$*, the following inequality holds pointwisely:*

$$\operatorname{Sup} S^m \leq \operatorname{Sup} Q^m.$$

Now we consider the behavior of $\operatorname{Sup} T^m$ under the alternative hypothesis. Then there will be at least one segment with predictability, i.e., $\beta_k > 0$ for some k. The k-th segment t_k will diverge to infinity, so will $\operatorname{Sup} T^m$. Therefore, the test is consistent.

In practice, we may have some knowledge about the timing of the change. Suppose we suspect that a structural change occurs at the end of the sample, then a forward recursive t-test can be used to check whether predictability exists prior to that. Let $T_1 = [\epsilon T]$ be the earliest possible date that the change happens. The limiting distribution under the null hypothesis is stated in the following corollary.

Corollary 1.5.5 (A Forward Recursive t-test) *Suppose the innovations* $(u_t, v_t)'$ *satisfy the Martingale Difference Assumption 1.1.1. Let* (x_t, y_t) *be generated according to the predictive regression* (1.12)-(1.13)*. For a given* $\epsilon > 0$*, define the admissible set for the earliest possible change date*

$$\Pi_\epsilon^F = \{T_1 : \epsilon T \leq T_1 \leq T\},$$

and its counterpart

$$\Lambda_\epsilon^F = \{\lambda : \epsilon \leq \lambda_1 \leq 1\}.$$

Let t_1 denote the t-statistic for the first subsample. Further denote

$$\operatorname{Sup} T^F = \sup_{T_1 \in \Pi_\epsilon^F} t_1, \quad \operatorname{Sup} S^F = \sup_{\lambda_1 \in \Lambda_\epsilon^F} s_1, \quad \operatorname{Sup} Q^F = \sup_{\lambda_1 \in \Lambda_\epsilon^F} q_1.$$

Then, we have

$$\operatorname{Sup} T^F \Rightarrow \operatorname{Sup} S^F.$$

Moreover, conditional on the path of $(W(r), G(r))$, the following inequality holds:

$$\operatorname{Sup} S^F \leq \operatorname{Sup} Q^F.$$

Remark 1.5.3 For implementation, we can use

$$r_1 = q_1(0) = \frac{\int_0^{\lambda_1} W^{\mu_1}(r) dG(r)}{\left(\int_0^{\lambda_1} W^{\mu_1}(r)^2 dr\right)^{1/2}},$$

to form a feasible distribution bound given by $\operatorname{Sup} R^F = \sup_{\lambda_1 \in \Lambda_\epsilon^F} r_1$ because the simulations (Subfigure (a) of Figure 1·6 and Table 1.1) demonstrate that the cumulative distribution of $\operatorname{Sup} R^F$ always provides an upper envelope for that of $\operatorname{Sup} Q^F(c)$ with $c \leq 0$.

Suppose we suspect that a structural change occurs at the beginning of the sample, a reverse recursive t-test can be used to check whether predictability exists after that. Let $T_1 = [(1-\epsilon)T]$ be the latest possible date that the change happens. The limiting distribution under the null hypothesis is stated in the following corollary:

Corollary 1.5.6 (A Reverse Recursive t-test) *Suppose the innovations $(u_t, v_t)'$ satisfy the Martingale Difference Assumption 1.1.1. Let (x_t, y_t) be generated according to the predictive regression* (1.12)-(1.13)*. For a given $\epsilon > 0$, define the admissible set*

for the change date T_1

$$\Pi_\epsilon^R = \{T_1 : 1 \leq T_1 \leq (1-\epsilon)T\},$$

and its counterpart

$$\Lambda_\epsilon^R = \{\lambda_1, 0 \leq \lambda_1 \leq 1-\epsilon\}.$$

Let t_2 *denote the* t*-statistic for the second subsample, Moreover, define*

$$\mathrm{Sup}\, T^R = \sup_{T_2 \in \Pi_\epsilon^R} t_2, \quad \mathrm{Sup}\, S^R = \sup_{\lambda_1 \in \Lambda_\epsilon^R} s_2, \quad \mathrm{Sup}\, Q^R = \sup_{\lambda_1 \in \Lambda_\epsilon^R} q_2.$$

Then, we have

$$\mathrm{Sup}\, T^R \Rightarrow \mathrm{Sup}\, S^R.$$

Moreover, conditional on the path of $(W(r), G(r))$*, the following inequality holds:*

$$\mathrm{Sup}\, S^R \leq \mathrm{Sup}\, Q^R.$$

Remark 1.5.4 For implementation, we can use

$$r_2 = q_2(0) = \frac{\int_{\lambda_1}^1 W^{\mu_2}(r) dG(r)}{\left(\int_{\lambda_1}^1 W^{\mu_2}(r)^2 dr\right)^{1/2}},$$

to form a feasible distribution bound given by $\mathrm{Sup}\, R^R = \sup_{\lambda_2 \in \Lambda_\epsilon^R} r_2$, because simulations (Subfigure (b) of Figure 1·6 and Table 1.1) demonstrate the cumulative distribution of $\mathrm{Sup}\, R^R$ always provides an upper envelope for that of $\mathrm{Sup}\, Q^R(c)$ with $c \leq 0$.

Suppose we conjecture that return predictability exists in the middle of the sample, then we can conduct the rolling window t-test as follows: let (T_1, T_2) be the cut-off for a subsample. The limiting distribution under the null hypothesis is stated in the following corollary.

Corollary 1.5.7 (A Rolling Window t-test) *Suppose the innovations* $(u_t, v_t)'$ *satisfy the Martingale Difference Assumption 1.1.1. Let* (x_t, y_t) *be generated according*

to the predictive regression (1.12)-(1.13). *For a given* $\epsilon > 0$, *define the admissible set for the subsample cut-off*

$$\Pi_\epsilon^W = \{(T_1, T_2) : T_2 - T_1 \geq \epsilon T, T_1 \geq 1, T_2 \leq T\},$$

and its counterpart

$$\Lambda_\epsilon^W = \{(\lambda_1, \lambda_2) : \lambda_2 - \lambda_1 \geq \epsilon, \lambda_1 \geq 0, \lambda_2 \leq 1\}.$$

Let the t_2 *denote the t-statistic for the selected subsample. Define*

$$\operatorname{Sup} T^W = \sup_{(T_1,T_2)\in\Pi_\epsilon^W} t_2, \quad \operatorname{Sup} S^W = \sup_{(\lambda_1,\lambda_2)\in\Lambda_\epsilon^W} s_2 \quad \operatorname{Sup} Q^W = \sup_{(\lambda_1,\lambda_2)\in\Lambda_\epsilon^W} q_2$$

Then, we have

$$\operatorname{Sup} T^W \Rightarrow \operatorname{Sup} S^W.$$

Moreover, conditional on the path of $(W(r), G(r))$, *the following inequality holds:*

$$\operatorname{Sup} S^W \leq \operatorname{Sup} Q^W.$$

Remark 1.5.5 For implementation, we can use

$$r_2 = q_2(0) = \frac{\int_{\lambda_1}^{\lambda_2} W^{\mu_2}(r) dG(r)}{\left(\int_{\lambda_1}^{\lambda_2} W^{\mu_2}(r)^2 dr\right)^{1/2}},$$

to form a feasible distribution bound given by $\operatorname{Sup} R^W = \sup_{\lambda_1\in\Lambda_\epsilon^W} r_2$ because the simulations (Subfigure (c) of Figure 1·6 and Table 1.1) demonstrate that the cumulative distribution of $\operatorname{Sup} R^W$ always provides an upper envelope for that of $\operatorname{Sup} Q^W(c)$ with $c \leq 0$.

1.6 Finite Sample Properties of the Proposed Tests

In this section, I will evaluate the finite sample properties of my proposed tests, including the usual t-test, the forward recursive t-test, the reverse recursive t-test,

and the rolling window t-test. Specifically, I use the predictive regression model (1.1)-(1.2) to simulate $N = 10,000$ samples. To evaluate large sample properties, I choose the sample size $T = 500$ to generate distributions and relevant critical values. Since we need to use bounds that are free from nuisance parameters to conduct inference, it is of interest to compare the statistics with their known bounds $c = 0$. Recall that $c = 0$ corresponds to the case when the predictor is a unit root process, and $\theta = 0$ corresponds to the case when the innovations in the predictor are uncorrelated with the outcome variable.

We first evaluate the large sample ($T = 500$) properties of the proposed tests. Figure 1·4 displays the cumulative distribution functions of the usual t-statistic and the standard normal distribution, with different parameter specifications. As evident from the graph, the t-statistic always falls to the left-hand side of the standard normal distribution for $c \leq 0$ and $\theta \leq 0$. Therefore, even though the t-statistic depends on the nuisance parameter c that is difficult to estimate, it is still possible to conduct a conservative test by using the traditional critical values given by the standard normal distribution. As the magnitude of c or θ increases, the distribution of t-statistic shifts further toward the left of the standard normal distribution. When $c = 0$ or $\theta = 0$, the t-statistic distribution returns to the standard normal distribution. Hence, the test becomes exact.

Figure 1·5 displays the comparison of the local power of my t-test with that of Q-test from Campbell and Yogo (2006), with different parameters specifications. As we can see from the graph, the t-test is more powerful than the Q-test when $c = 0$. When c decreases to $c = -2$, the t-test still has good power relative to the more complicated Q-test. As it c continues to decrease, the Q-test starts to outperform my t-test.

Figure 1·6 displays the bounds $\operatorname{Sup} Q^F$, $\operatorname{Sup} Q^R$, and $\operatorname{Sup} Q^W$ for different c. As

c becomes more negative, these bounds shift to the left until reaching some lower contours. At the same time, when $c = 0$, they become the upper contours, given by $\operatorname{Sup} R^F$, $\operatorname{Sup} R^R$, and $\operatorname{Sup} R^W$, respectively.

Figure 1·7 displays the cumulative distribution of the forward recursive t-statistic $\operatorname{Sup} T^F$ and the two proposed bounds $\operatorname{Sup} Q^F$ and $\operatorname{Sup} R^F$ with different parameters specifications. The dashed line lying in the middle is the $\operatorname{Sup} Q^F$ bound, which depends on the nuisance parameter c. The solid line lying on the right-hand side is the $\operatorname{Sup} R^F$ bound, whose critical values are used to conduct conservative inference. These two bounds are very close, even for $c = -20$. Additionally, the $\operatorname{Sup} Q^F$ bound lies strictly on the left-hand side of the $\operatorname{Sup} R^F$ bound in my simulations. Same as before, as the magnitudes of c and θ increase, the distribution of the forward recursive t-statistic shifts toward the left of the two bounds. When $c = 0$ or $\theta = 0$, the distribution of the forward recursive t-statistic coincides with the two bounds, implying that the test becomes exact.

Figure 1·8 displays the cumulative distributions of the reverse recursive t-statistic $\operatorname{Sup} T^R$ and the two proposed bounds, with different parameter specifications. The dashed line lying in the middle is the $\operatorname{Sup} Q^R$ bound, which depends on the nuisance parameter c. The solid line lying on the right-hand side is the $\operatorname{Sup} R^R$ bound, whose critical values are used to conduct conservative inference. These two bounds lie very close, even for $c = -20$. Furthermore, the $\operatorname{Sup} Q^R$ bound lies strictly on the left-hand side of the $\operatorname{Sup} Q^R$ bound in my simulations. Same as before, as the magnitude of c or θ increases, the distribution of the forward recursive t-statistic further shifts toward the left of the two bounds. When $c = 0$ or $\theta = 0$, the distribution of the reverse recursive t-statistic coincides with the two bounds. Hence, the test becomes exact.

Figure 1·9 displays the cumulative distributions of the rolling window t-statistic $\operatorname{Sup} T^W$ and the two proposed bounds, with different parameters specifications. The

dashed line lying in the middle is the $\text{Sup}\, Q^W$ bound, which depends on the nuisance parameter c. The solid line lying on the right-hand side is the $\text{Sup}\, R^W$ bound, whose critical values are used to conduct conservative inference. These two bounds lie very close, even for $c = -20$. Further, the $\text{Sup}\, Q^W$ bound lies strictly on the left-hand side of the $\text{Sup}\, Q^W$ bound in my simulations. Same as before, as the magnitude of c or θ increases, the distribution of the rolling window t-statistic further shifts toward the left of the two bounds. When $c = 0$ or $\theta = 0$, the distribution of the rolling window t-statistic coincides with the two bounds. Hence, the test becomes exact.

Figure 1·10 uses the $\text{Sup}\, T^F$ to illustrate the local power of the robust statistics versus that of the t-statistic. Specifically, I assume there is only one structural change occurring at T_b at which β changes from b/T to 0. Moreover, I consider two cases, $T_b = 400$ and $T_b = 100$, as well as the situation where there is no structural change. The results show that when there is no structural change, the vanilla t-statistic has higher power than the $\text{Sup}\, T^F$. When the change date is near the end of the sample ($T_b = 400$), the forward recursive t-statistic has larger power, despite the fact that the critical values are simulated under the assumption that the structural change could happen as early as $T_b = 100$. Finally, when $T_b = 100$, the forward recursive t-statistic displays significantly higher power. These findings show that the robust statistics are useful when we conjecture there exist structural changes.

Table 1.1 reports the critical values of all relevant quantities using samples with sample size $T = 500$. These critical values could be treated as critical values coming from their corresponding asymptotic distributions. As we can see in Panel A, in the unit root case ($c = 0$) or the contemporaneous independent innovations case ($\theta = 0$), all statistics coincide with their corresponding bounds. Regarding the other nontrivial cases, it is evident that we see the critical values of all the bounds are larger than their counterparts. Again, we see the bounds Q and the bound R are very close, even for

a large degree of $c = -20$, and the critical values of the Q bounds are always smaller than those of the R bounds. Moreover, keeping c fixed, the correlation coefficient θ also pushes the distribution of the test statistics as well as the Q bounds toward the left, hence deteriorating the approximation by the R bounds.

Finally, I evaluate the small sample properties of these inference procedures by checking their empirical rejection rate samples with sample size $T = 50$, 100, and 200 (see Table 1.2). The small size $T = 50$ is relevant to my application when using annual data. Apparently, the usual t-statistic, the forward recursive t-statistic, and the reverse recursive t-statistics all have good small sample properties when $c = 0$ or $\theta = 0$. However, the over-rejection rate of rolling window tests comes from the fact that when the trimming parameter $\epsilon = 0.2$ is chosen, the subsample has a very small sample size. For example, for a sample of size $T = 50$, the subsample only has 10 data points, which is clearly inadequate for convergence. Therefore, one should either have a bigger sample or choose a bigger trimming parameter to apply the rolling window t-test. Similarly, as the degree of c or θ increases, the tests become conservative; thus, the empirical rejection rates become lower than the nominal rates.

1.7 Empirical Application

In this section, I apply my methods of inference in predictive regressions with nearly integrated predictors on US financial data and compare my results with previous findings in the literature. Specifically, for the outcome variable, I use excess returns of stocks over a risk-free return. For the nearly integrated predictors, I use dividend-price and earnings-price ratios.

My data span from 1926 to 2018. I follow Campbell and Shiller (1988) and Campbell and Yogo (2006) to process data. The excess returns and the financial ratios are all in logs. To compute the excess returns, I collect the NYSE/AMEX value-weighted

index for the stock returns and the T-bill returns for the risk-free returns at various frequencies from the Center for Research in Security Prices (CRSP). In particular, I use the return on the one-month T-bill for the monthly series and the return on the three-month T-bill for the quarterly series. Rolling over the return on the three-month T-bill gives the annual risk-free return. Regarding the financial ratios, the dividend-price ratio is computed as dividends over the past year divided by the current price, and the earnings-price ratio is computed as a moving average of earnings over the last ten years divided by the current price. While the dividend data for the NYSE/AMEX index are available at the CRSP, the earning counterpart is missing. Hence, the earnings-price ratio from the S&P 500 is used instead. Earnings data have been available at quarterly frequency since 1935, and annual frequency before that. I follow Shiller (2016) to construct the monthly earning by linear extrapolation.

In addition to the testing predictability for the full sample, I also evaluate the results for various subsamples corresponding to previous empirical work. I consider five additional subsamples in total. For the extra beginning dates, 1946 and 1952 are important. The former date marks the end of WWII and is used by Amihud and Hurvich (2004) and Lewellen (2004) to construct their subsamples. The later date 1952 was used by Stambaugh (1999), Lewellen (2004), Campbell and Yogo (2006) to construct their subsamples. For the extra ending date, 2002 is used by Campbell and Yogo (2006) and is close to the ending date of samples used by much previous literature on testing predictability.

Table 1.3 summarizes the results from running only regression (1.1). The estimates $\hat{\beta}$ obtained from all subsamples produce large magnitude values. Not surprisingly, the corresponding t-statistics are all significant, often larger than two. However, the inference drawn from the usual practice could be invalid because the predictors are highly persistent, and their innovations are highly correlated with the outcome

variable. This table also reports the correlation coefficient estimate $\hat{\theta}$. Evidently, the estimates of the correlation coefficient are often very close to minus one. Moreover, the table displays the DF-GLS statistic and the 95% confidence for the autoregressive coefficient ρ (and its corresponding location coefficient c) for the predictors, as well as the log dividend and the log earnings-price ratios. The results suggest that all the series are highly persistent and the obtained confidence interval always coverer a unit root. Indeed, the traditional inference practice is invalid.

Therefore, alternative procedures are needed to draw the correct statistical inference. I summarize the results yielded by my proposed tests in Table 1.4. The table first reports the estimate $\hat{\delta}$ of the regression coefficient of the disturbance in outcome variable on innovations in predictors and the estimate $\hat{\beta}$ capturing the predictability, both obtained from the augmented regressions. Thereafter, I compute the t-statistic to draw statistical inference about the predictability of selected predictors at different frequencies and compare the results with those obtained from the Q-statistic proposed by Campbell and Yogo (2006). Finally, I apply my various supremum t-statistics to pick up possible predictability existing over some subsamples but hard to be detected by the usual t-statistic.

For the full sample (1927-2018), my t-statistic demonstrates that the earnings-price ratio can predict the excess stock return while the dividend-price can not. For the subsample (1927-2002), the earnings-price ratio continues to be a reliable predictor, and it is more significant than when using the full sample. Meanwhile, the t-statistic picks the dividend-price ratio as a weak predictor only at the annual frequency. Regarding the post-WWII subsample (1946-2018), I find that the time dividend-price ratio is a more reliable predictor than the earnings-price ratio and earning-price ratio, which only serves as a weak predictor at quarterly frequency. The subsample (1946-2002) produces the same quantitative result, but with a more

significant degree, and consequently, the two predictors at all frequencies display the predictive power. Next, for the 1952-2018 subsample, all the predictors lose the ability to forecast the return, according to the t-statistic. Finally, for the 1952-2002 subsample, the t-statistic reveals that the dividend-price ratio is highly capable of predicting the excess return at quarterly and monthly frequencies, but not at annual frequency. For comparison, Table 1.4 also provides the 90% confidence interval obtained by the Q-statistic of Campbell and Yogo (2006), which corresponds to a 95% one tailed-test. To make the comparison fair, I only compare my results with significant levels above 95% with the Q-test. As the outcome shows, the inference results are similar but generally weaker than those drawn from the proposed t-statistic, except for two cases: For e-p at annual frequency over the 1927-2018 sample, the resulting Q-statistic is more significant than my t-statistic. For the 1946-2018 subsample, the Q-statistic picks the predictability of e-p at annual frequency, while the t-statistic can not. This is because e-p in this case lacks sufficient persistency and my test becomes too conservative. For the other few cases, i.e., d-p at quarterly and monthly frequencies over the 1952-2002 subsample, as well as e-p at monthly frequency over the 1927-2018, 1946-2018, and 1946-2002 subsamples, my t-statistic demonstrates more power, not only due to the nearly integrated nature of the predictors I employ, but also can be due to the fact that Q-test is more vulnerable to model instability because of the Bonferroni procedure. Generally speaking, the proposed t-statistic performs satisfactorily when a predictor is highly persistent, usually corresponding to a predictor at high frequency.

Lastly, as the testing results vary over different subsamples, it is likely that the t-statistic and the Q-statistic may become uninformative due to possible structural changes. Hence, I apply the proposed supremum tests to the available data. Recall that the $\mathrm{Sup}\, T^F$ test is robust with respect to structural changes at the end of the

sample but not at the beginning, the Sup T^R test is robust to structural changes at the beginning of the sample but not at the end, and the Sup T^W test is able to detect predictability in the middle of the sample. Therefore, Sup T^W should be the most robust statistic against possible structural changes, and we look at the corresponding results first. Unsurprisingly, the Sup T^W test suggests that the dividend and earnings-price ratios are almost always able to predict the returns. The special case is that when it comes to the full sample, the dividend-price ratio at the monthly frequency cannot significantly predict the return at the 5% significance level, but it can at the 10% significance level. Furthermore, the Sup T^F statistic produces similar results to the vanilla t-statistic for the full sample (1927-2018) and the 1927-2002 subsample, but more significant. For the other subsamples, the Sup T^F test yields the inference that all the predictors are able to predict the outcome variable at all frequencies, although for the 1952-2018 and 1952-2002 subsamples, we need to withhold a 10% significance level to conclude that the dividend-price ratio is able to predict the returns at the annual frequency. Lastly, the Sup T^R statistic shows dividend-price and earnings-price ratio at quarterly and monthly frequencies display predictability for the full sample (1927-2018), the 1927-2002 and 1946-2002 subsamples, with two exceptions: the dividend-price ratio for the 1927-2018 subsample and the earnings-price ratio for the 1946-2002 subsample, both at quarterly frequency, can only weakly predict the excess returns, meaning that one needs to endure the 10% significance level. Additionally, the dividend-price at monthly frequency shows predictability for the 1952 - 2002 subsample.

1.8 Conclusion

Herein this paper, I proposed a conservative t-statistic that is fairly easy to use. First, unlike the previous relatively complicated proposed procedures, the inference results

can be obtained from running a single augmented regression. Second, the simple t-statistic uses standard normal critical values to draw conservative inferences. Third, although the test is conservative, in practice, it yields similar results given by the efficient Q-test proposed by Campbell and Yogo (2006), when practitioners use a nearly integrated predictor, such as the dividend or earnings-price ratio. Notebly, my tests are sensitive to the sample periods I use, and they are particularly responsive to the starting and ending dates. This suggests that structural changes could affect the informativeness of my tests. Therefore, I also propose a class of robust tests by resorting to the supremum operator to address such concerning situations. Overall, my results show that both predictors have predictability for the returns at least over subsamples.

1.9 Figures

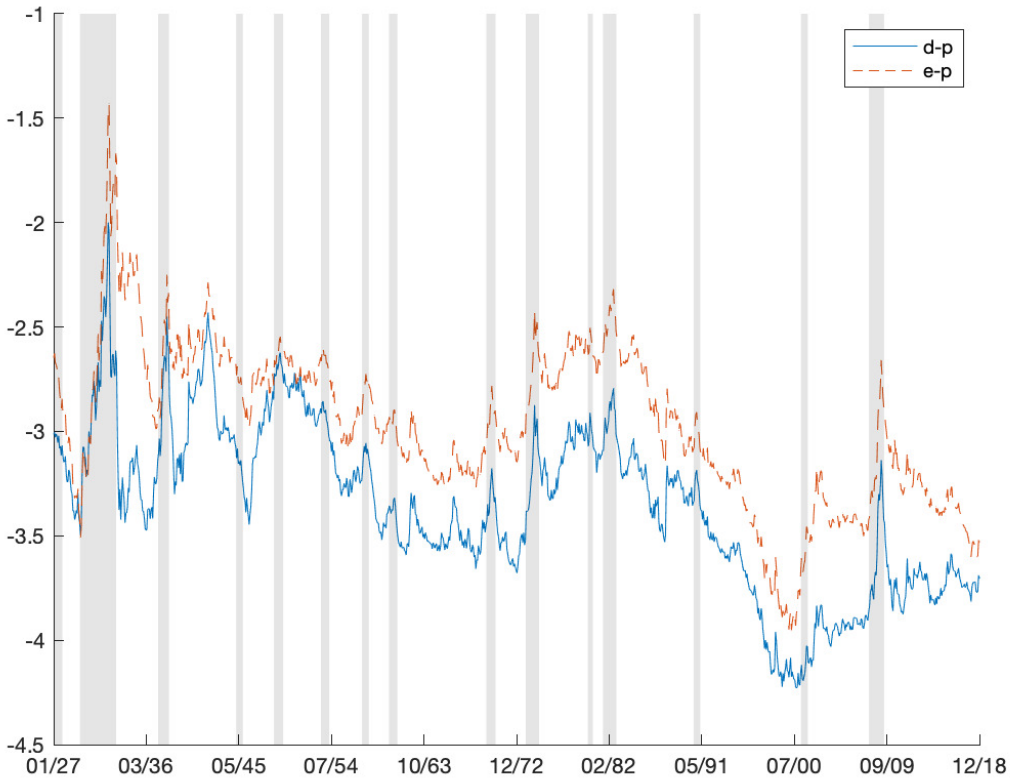


Figure 1·1: Time Series Plot of the Financial Ratios

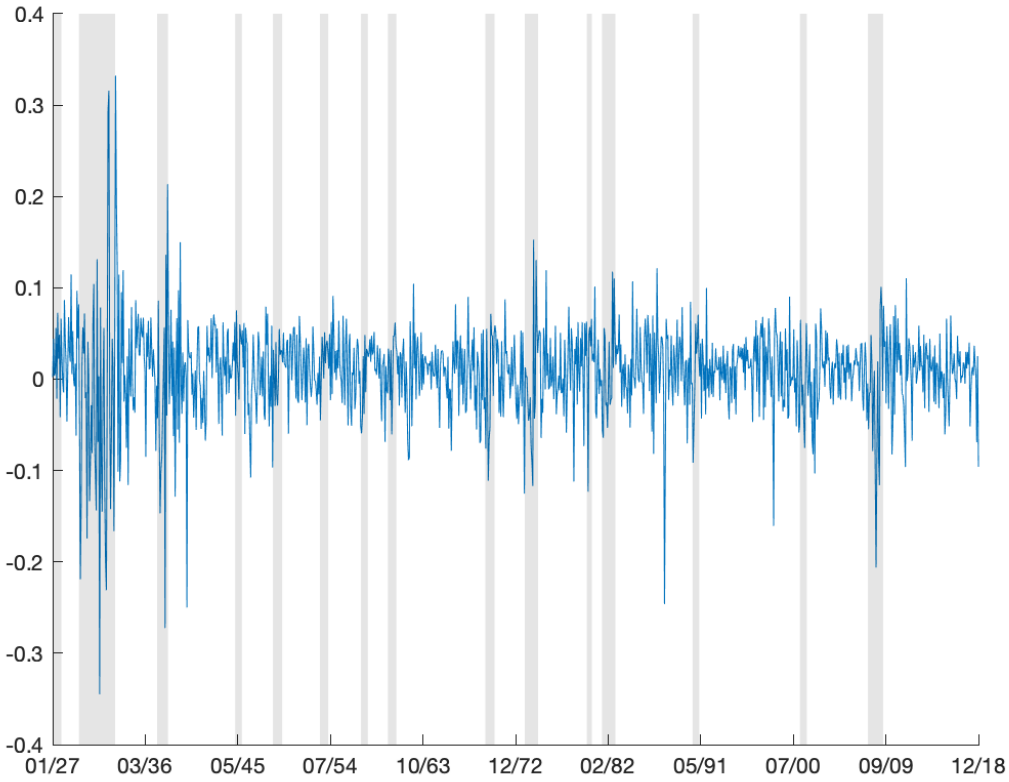


Figure 1·2: Time Series Plot of the Excess Stock Return

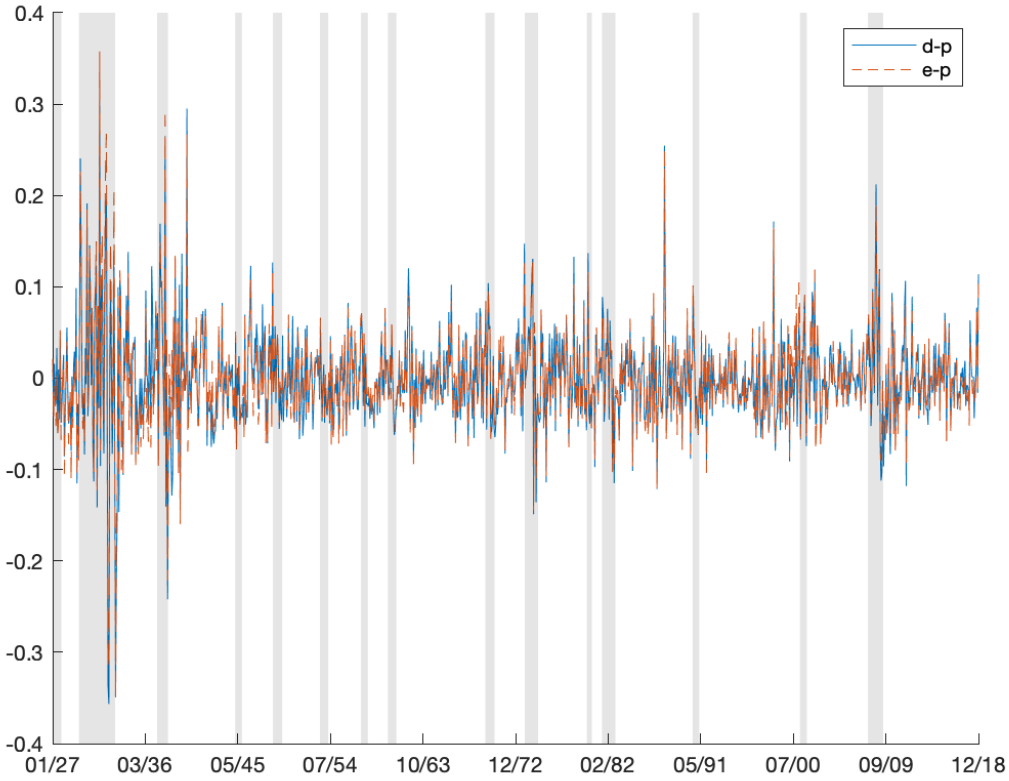
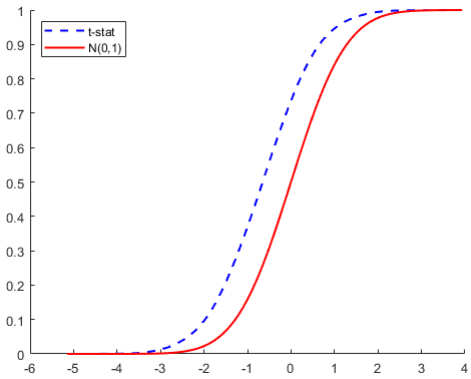
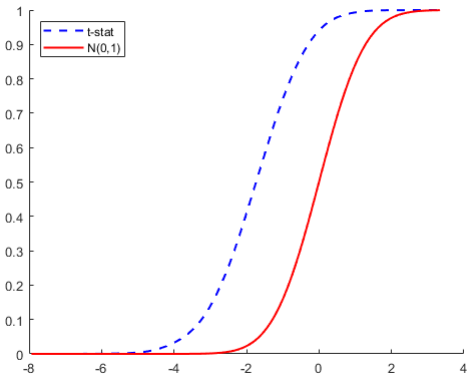


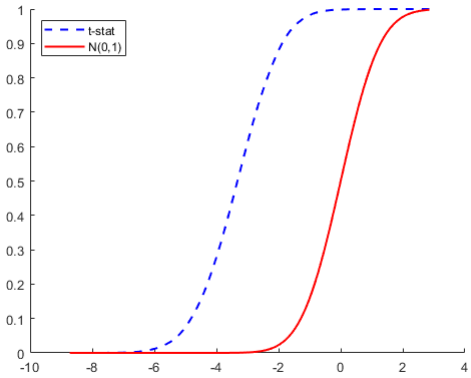
Figure 1·3: Time Series Plot of the First Differences of Financial Ratios



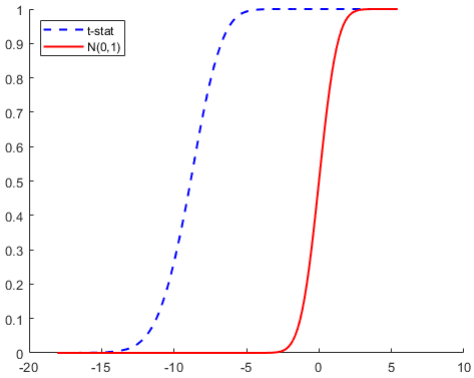
(a) $c = -2, \theta = -0.75$



(b) $c = -2, \theta = -0.95$

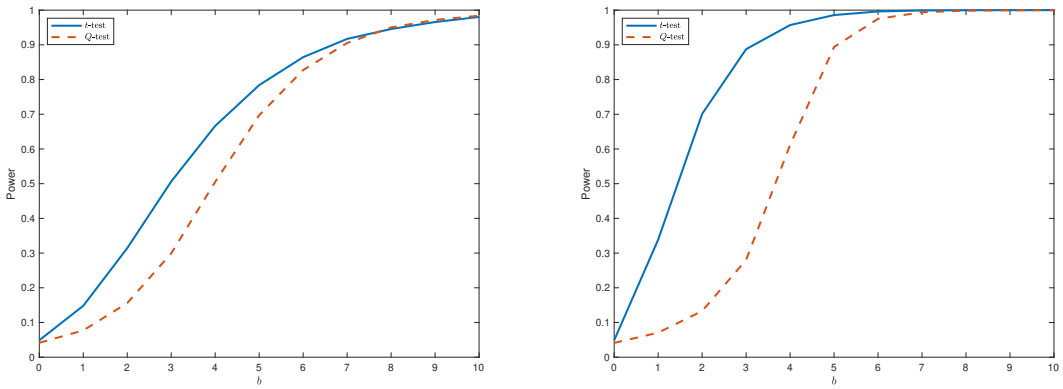


(c) $c = -20, \theta = -0.75$

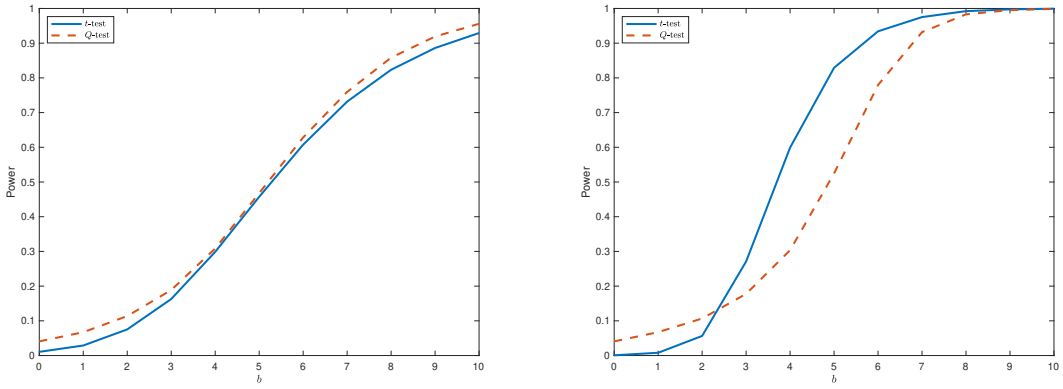


(d) $c = -20, \theta = -0.95$

Figure 1·4: t-statistic vs standard normal; $T = 500, N = 10,000$

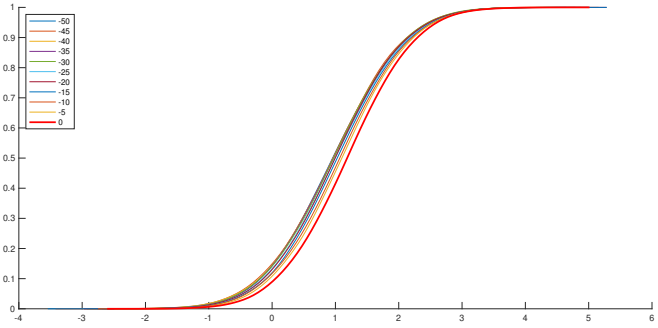


(a) $c = 0, \theta = -0.75$ **(b)** $c = 0, \theta = -0.95$

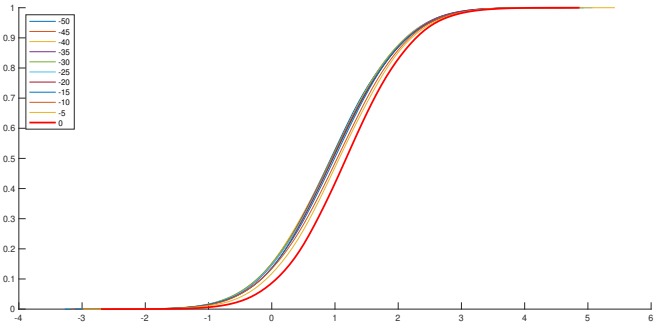


(c) $c = -2, \theta = -0.75$ **(d)** $c = -2, \theta = -0.95$

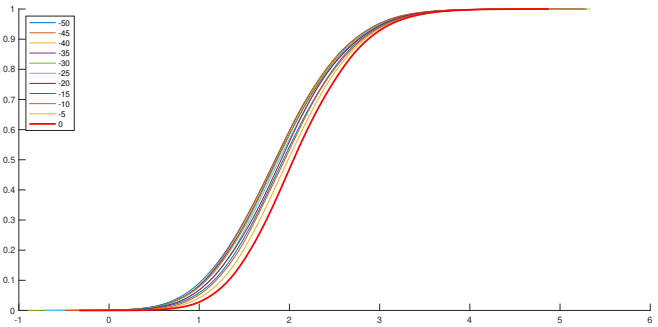
Figure 1·5: t-statistic vs Q-statistic local power; $T = 500, N = 10,000$



(a) Sup R^F as Feasible Distribution Bound for Sup Q^F

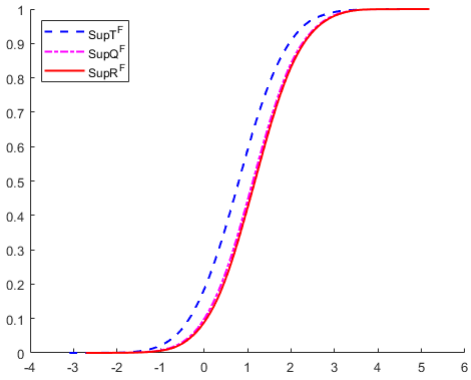


(b) Sup R^R as Feasible Distribution Bound for Sup Q^R

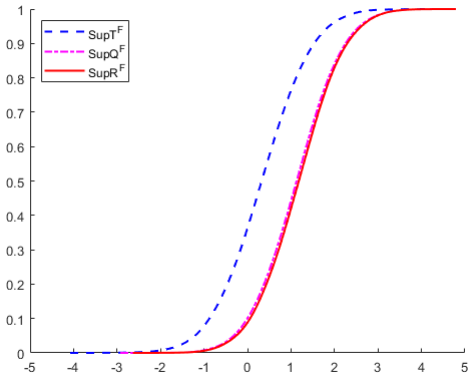


(c) Sup R^W as Feasible Distribution Bound for Sup Q^W

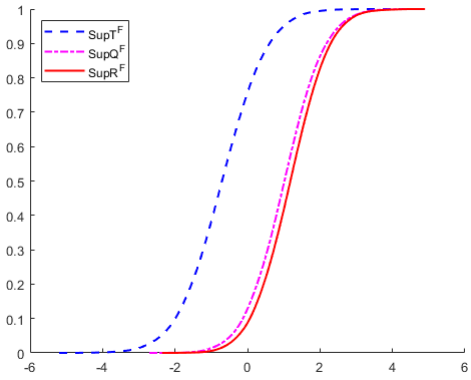
Figure 1·6: Feasible Distribution Bounds for typical cases; $c = [-50, 0]$; $T = 500, N = 10,000$



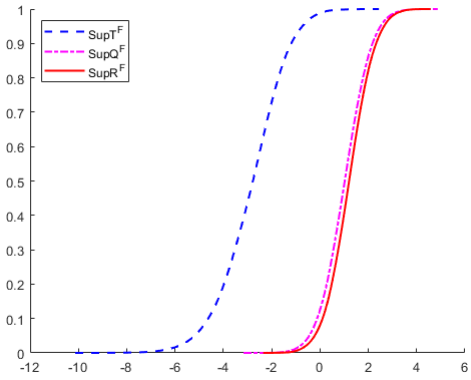
(a) $c = -2, \theta = -0.75$



(b) $c = -2, \theta = -0.95$

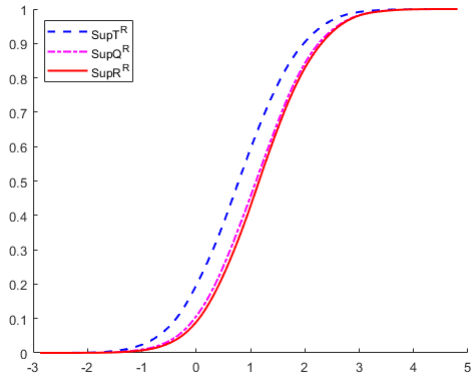


(c) $c = -20, \theta = -0.75$

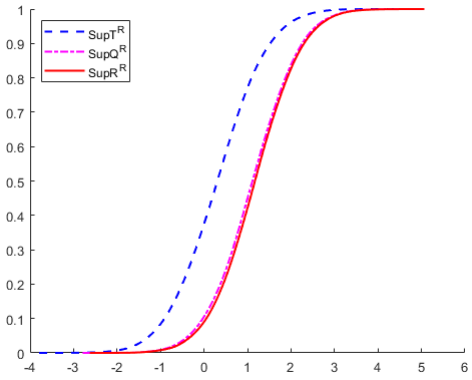


(d) $c = -20, \theta = -0.95$

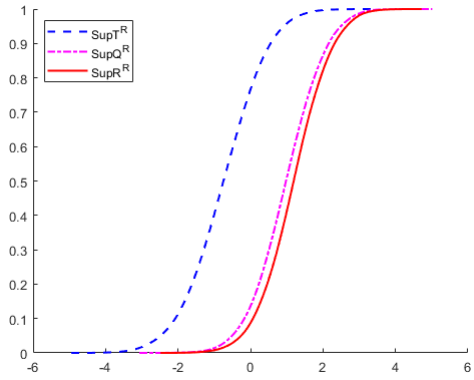
Figure 1·7: $\text{Sup}\, T^F$ vs $\text{Sup}\, Q^F$ vs $\text{Sup}\, R^F$; $T = 500, N = 10,000$



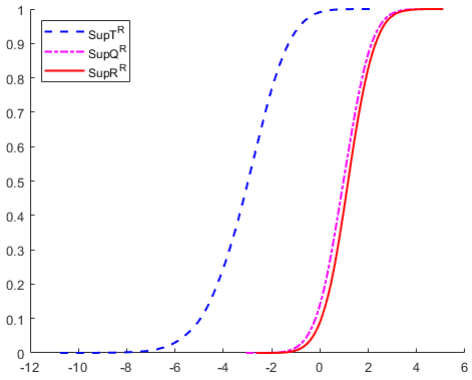
(a) $c = -2, \theta = -0.75$



(b) $c = -2, \theta = -0.95$

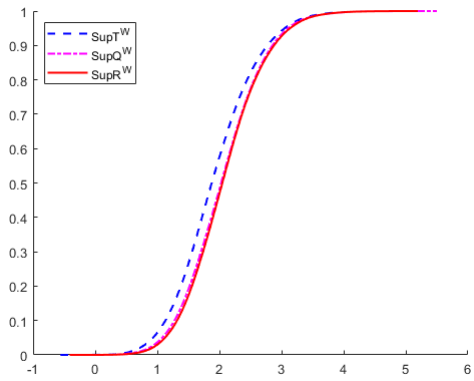


(c) $c = -20, \theta = -0.75$

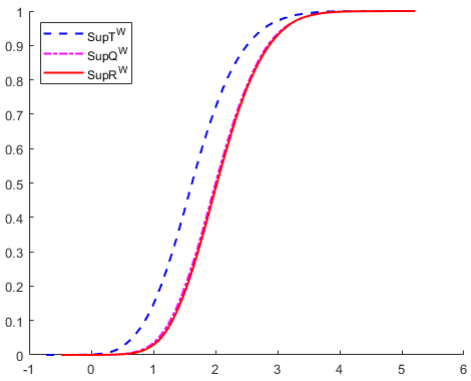


(d) $c = -20, \theta = -0.95$

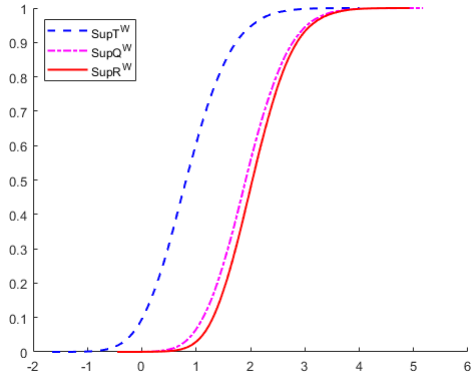
Figure 1·8: $\operatorname{Sup} T^R$ vs $\operatorname{Sup} Q^R$ vs $\operatorname{Sup} R^R$; $T = 500, N = 10,000$



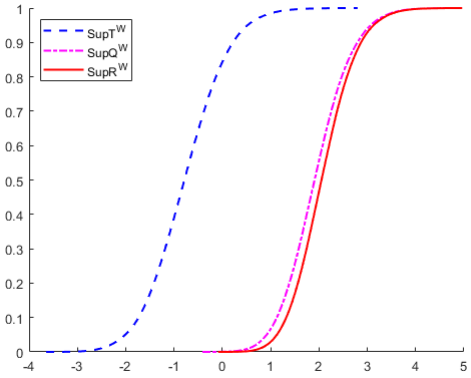
(a) $c = -2, \theta = -0.75$



(b) $c = -2, \theta = -0.95$

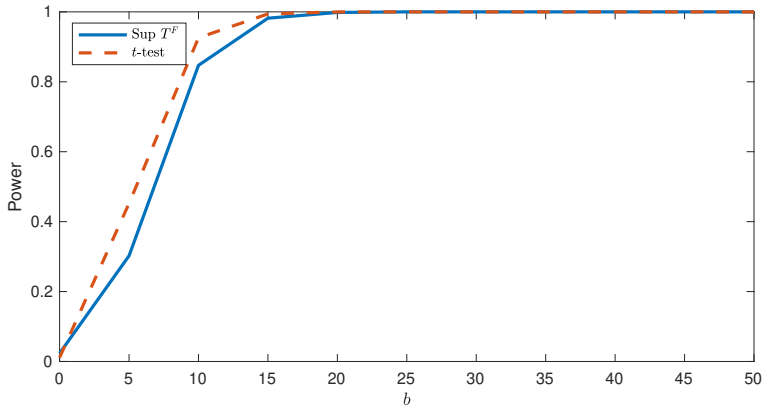


(c) $c = -20, \theta = -0.75$

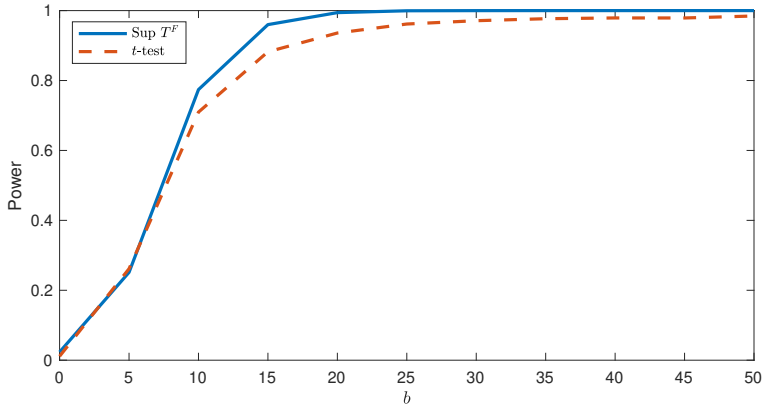


(d) $c = -20, \theta = -0.95$

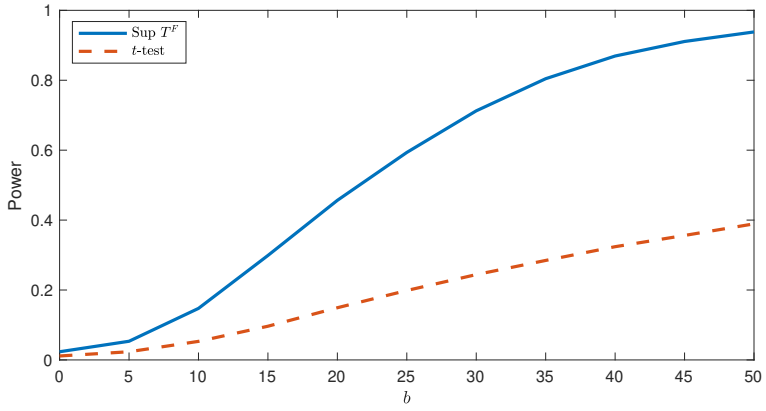
Figure 1·9: $\operatorname{Sup} T^W$ vs $\operatorname{Sup} Q^W$ vs $\operatorname{Sup} R^W$; $T = 500, N = 10,000$



(a) No Structural Change



(b) One Structural Change: After $T_b = 400$, $b = 0$



(c) One Structural Change: After $T_b = 100$, $b = 0$

Figure 1·10: Sup T^F vs t-statistic local power; $T = 500$, $N = 10,000$, $c = -2$, $\theta = -0.75$

1.10 Tables

Table 1.1: Large Sample ($T = 500$) Critical Values for the Proposed Statistics

α	t-statistic	Normal	Sup T^F	Sup T^R	Sup T^W	Sup Q^F	Sup Q^R	Sup Q^W	Sup R^F	Sup R^R	Sup R^W
Panel A: $c = 0$ or $\theta = 0$											
.10	1.283	1.282	2.297	2.292	2.950	2.298	2.281	2.862	2.298	2.281	2.862
.05	1.647	1.645	2.642	2.638	3.255	2.612	2.587	3.120	2.612	2.587	3.120
.01	2.361	2.326	3.260	3.254	3.795	3.196	3.202	3.661	3.196	3.202	3.661
Panel B: $c = -2, \theta = -0.75$											
.10	0.647	1.282	1.957	1.931	2.755	2.269	2.237	2.838	2.301	2.279	2.859
.05	1.019	1.645	2.285	2.266	3.034	2.582	2.555	3.116	2.624	2.586	3.126
.01	1.791	2.326	2.965	2.862	3.592	3.145	3.171	3.604	3.197	3.160	3.607
Panel C: $c = -2, \theta = -0.95$											
.10	-0.327	1.282	1.513	1.492	2.509	2.267	2.224	2.845	2.303	2.277	2.855
.05	0.087	1.645	1.868	1.828	2.785	2.593	2.577	3.095	2.620	2.589	3.101
.01	0.825	2.326	2.526	2.501	3.352	3.183	3.125	3.634	3.217	3.211	3.654
Panel D: $c = -20, \theta = -0.75$											
.10	-1.886	1.282	0.638	0.555	1.729	2.169	2.135	2.785	2.279	2.276	2.854
.05	-1.483	1.645	0.992	0.916	2.016	2.472	2.442	3.045	2.606	2.583	3.109
.01	-0.754	2.326	1.672	1.598	2.549	3.066	3.025	3.567	3.202	3.178	3.606
Panel E: $c = -20, \theta = -0.95$											
.10	-6.787	1.282	-1.161	-1.330	0.243	2.201	2.133	2.782	2.329	2.277	2.855
.05	-6.144	1.645	-0.753	-0.926	0.555	2.516	2.457	3.046	2.622	2.601	3.130
.01	-5.211	2.326	0.027	-0.123	1.134	3.113	3.042	3.573	3.207	3.213	3.671

Table 1.2: Small Sample ($T = 50, 100, 200$) Rejection Rates for the Proposed Statistics

α	t-statistic	Sup T^F	Sup T^R	Sup T^W
		$T = 50$		
		Panel A: $c = 0$ or $\theta = 0$		
.10	0.102	0.117	0.112	0.210
.05	0.053	0.071	0.068	0.152
.01	0.011	0.028	0.023	0.082
		Panel B: $c = -2, \theta = -0.75$		
.10	0.031	0.068	0.062	0.160
.05	0.012	0.040	0.037	0.112
.01	0.002	0.015	0.014	0.053
		Panel C: $c = -2, \theta = -0.95$		
.10	0.004	0.029	0.026	0.097
.05	0.002	0.018	0.015	0.069
.01	0.000	0.006	0.005	0.032
		$T = 100$		
		Panel A: $c = 0$ or $\theta = 0$		
.10	0.103	0.102	0.101	0.149
.05	0.057	0.055	0.054	0.093
.01	0.011	0.017	0.012	0.034
		Panel B: $c = -2, \theta = -0.75$		
.10	0.035	0.048	0.046	0.105
.05	0.014	0.026	0.025	0.063
.01	0.003	0.008	0.005	0.022
		Panel C: $c = -2, \theta = -0.95$		
.10	0.003	0.020	0.020	0.056
.05	0.001	0.011	0.010	0.029
.01	0.000	0.002	0.003	0.010
		$T = 200$		
		Panel A: $c = 0$ or $\theta = 0$		
.10	0.097	0.099	0.099	0.131
.05	0.049	0.051	0.053	0.071
.01	0.009	0.013	0.013	0.019
		Panel B: $c = -2, \theta = -0.75$		
.10	0.029	0.047	0.049	0.082
.05	0.011	0.022	0.024	0.046
.01	0.001	0.004	0.005	0.013
		Panel C: $c = -2, \theta = -0.95$		
.10	0.003	0.019	0.017	0.041
.05	0.002	0.008	0.008	0.024
.01	0.000	0.001	0.001	0.006

Table 1.3: Estimates of the Model Parameters ignoring Non-standard Features

Frequency	Variable	Observation	$\hat{\beta}$	t-statistic	$\hat{\theta}$	DF-GLS	95% CI:c	95% CI:ρ
Panel A: 1927-2018								
Annual	d-p	92	0.112	2.148**	-0.684	-1.355	[-10.199 3.132]	[0.889 1.034]
	e-p	92	0.137	2.680***	-0.963	-2.011	[-16.365 1.087]	[0.822 1.012]
Quarterly	d-p	368	0.036	2.140**	-0.943	-1.782	[-14.155 1.884]	[0.962 1.005]
	e-p	368	0.039	2.838***	-0.986	-1.838	[-14.155 1.884]	[0.962 1.005]
Monthly	d-p	1104	0.006	1.520*	-0.956	-1.627	[-12.102 2.583]	[0.989 1.002]
	e-p	1104	0.011	2.597***	-0.987	-1.500	[-11.147 2.870]	[0.990 1.003]
Panel B: 1927-2002								
Annual	d-p	77	0.136	2.231**	-0.700	-1.218	[-8.528 3.532]	[0.889 1.046]
	e-p	77	0.150	2.598***	-0.961	-1.957	[-16.365 1.087]	[0.788 1.014]
Quarterly	d-p	305	0.036	2.140**	-0.943	-1.782	[-14.155 1.884]	[0.954 1.006]
	e-p	305	0.047	2.948***	-0.987	-2.063	[-16.365 1.087]	[0.946 1.004]
Monthly	d-p	913	0.009	1.781**	-0.955	-1.525	[-11.147 2.870]	[0.988 1.003]
	e-p	913	0.013	2.700***	-0.988	-1.732	[-13.124 2.203]	[0.986 1.002]
Panel C: 1946-2018								
Annual	d-p	73	0.115	2.302**	-0.618	-1.248	[-8.528 3.532]	[0.883 1.048]
	e-p	73	0.120	2.175**	-0.949	-1.814	[-14.155 1.884]	[0.806 1.026]
Quarterly	d-p	292	0.029	2.348**	-0.953	-1.657	[-13.124 2.203]	[0.955 1.008]
	e-p	292	0.029	2.140**	-0.977	-1.575	[-12.102 2.583]	[0.959 1.009]
Monthly	d-p	876	0.009	2.346***	-0.957	-1.625	[-12.102 2.583]	[0.986 1.003]
	e-p	876	0.009	2.110**	-0.981	-1.377	[-10.199 3.132]	[0.988 1.004]
Panel D: 1946-2002								
Annual	d-p	58	0.131	2.377***	-0.641	-0.943	[-6.415 3.967]	[0.889 1.068]
	e-p	58	0.116	1.922**	-0.940	-1.503	[-11.147 2.870]	[0.808 1.050]
Quarterly	d-p	229	0.040	2.772***	-0.962	-1.272	[-9.362 3.352]	[0.959 1.015]
	e-p	229	0.034	2.166**	-0.976	-1.621	[-12.102 2.583]	[0.947 1.011]
Monthly	d-p	685	0.013	2.778***	-0.957	-1.278	[-9.362 3.352]	[0.986 1.005]
	e-p	685	0.010	2.137**	-0.980	-1.481	[-11.147 2.870]	[0.984 1.004]
Panel E: 1952-2018								
Annual	d-p	67	0.112	1.923**	-0.665	-0.875	[-6.415 3.967]	[0.904 1.059]
	e-p	67	0.114	1.927**	-0.962	-1.641	[-10.801 1.200]	[0.839 1.018]
Quarterly	d-p	268	0.028	1.948**	-0.966	-0.832	[-5.779 4.072]	[0.978 1.015]
	e-p	268	0.027	1.833**	-0.981	-1.171	[-8.528 3.532]	[0.968 1.013]
Monthly	d-p	804	0.008	1.926**	-0.965	-0.690	[-5.293 4.166]	[0.993 1.005]
	e-p	804	0.008	1.784**	-0.982	-0.915	[-6.415 3.967]	[0.992 1.005]
Panel F: 1952-2002								
Annual	d-p	52	0.124	1.929**	-0.699	-0.630	[-4.758 4.256]	[0.909 1.082]
	e-p	52	0.107	1.643**	-0.958	-1.386	[-10.199 3.132]	[0.804 1.060]
Quarterly	d-p	205	0.039	2.326**	-0.978	-0.453	[-4.289 4.316]	[0.979 1.021]
	e-p	205	0.030	1.812**	-0.980	-1.280	[-7.122 3.827]	[0.965 1.019]
Monthly	d-p	613	0.012	2.329**	-0.968	-0.286	[-3.463 4.436]	[0.994 1.007]
	e-p	613	0.009	1.771**	-0.982	-1.056	[-7.122 3.827]	[0.988 1.006]

Table 1.4: Tests of Predictability using Proposed Methods and Campbell-Yogo Method

Frequency	Predictor	Observation	$\hat{\delta}$	$\hat{\beta}$	t-statistic	Sup T^F	Sup T^R	Sup T^W	90% CY-CI:β
Panel A: 1927-2018									
Annual	d-p	92	-0.909	0.046	1.181	2.214	2.021	3.408**	[-0.010 0.189]
	e-p	92	-0.951	0.019	1.303*	3.175**	1.895	5.129***	**[0.011 0.194]**
Quarterly	d-p	368	-0.962	-0.007	-1.598	-0.078	2.578*	3.452**	[-0.012 0.034]
	e-p	368	-0.974	0.005	2.024**	5.163***	2.737**	6.588***	**[0.002 0.043]**
Monthly	d-p	1104	-0.917	-0.003	-2.569	-0.821	2.746**	3.047*	[-0.005 0.008]
	e-p	1104	-0.961	0.002	2.228**	5.628***	3.512***	6.907***	[-0.000 0.011]
Panel B: 1927-2002									
Annual	d-p	77	-0.891	0.070	1.573*	2.214	2.500*	3.221**	[-0.002 0.223]
	e-p	77	-0.942	0.029	1.762**	5.724***	2.177	5.470***	**[0.023 0.228]**
Quarterly	d-p	305	-0.960	-0.000	-0.078	-0.078	5.842***	3.257**	[-0.007 0.047]
	e-p	305	-0.968	0.008	2.878***	5.163***	3.375***	6.347***	**[0.009 0.061]**
Monthly	d-p	913	-0.916	-0.001	-0.841	-0.841	5.568***	2.851	[-0.004 0.011]
	e-p	913	-0.956	0.002	3.060***	5.628***	4.100***	6.466***	**[0.000 0.016]**
Panel C: 1946-2018									
Annual	d-p	73	-0.778	0.072	1.780**	3.345***	1.853	3.483**	**[0.011 0.213]**
	e-p	73	-0.956	0.019	1.048	4.409***	1.048	5.443***	**[0.002 0.214]**
Quarterly	d-p	292	-0.982	0.008	2.178**	5.570***	2.178	3.229**	**[0.003 0.052]**
	e-p	292	-0.997	0.005	1.539*	6.164***	1.652	5.866***	[-0.002 0.045]
Monthly	d-p	876	-0.915	0.003	2.293**	5.310***	2.293*	2.714	**[0.001 0.016]**
	e-p	876	-0.977	0.002	2.143**	6.653***	2.322*	6.066***	[-0.001 0.013]
Panel D: 1946-2002									
Annual	d-p	58	-0.738	0.096	2.235**	3.345***	2.288*	3.630**	**[0.019 0.232]**
	e-p	58	-0.944	0.030	1.397*	4.409***	1.397	3.331**	[-0.011 0.222]
Quarterly	d-p	229	-0.989	0.022	5.570***	5.570***	5.570***	3.666***	**[0.010 0.063]**
	e-p	229	-0.991	0.008	2.263**	6.164***	2.363*	3.501**	**[0.000 0.060]**
Monthly	d-p	685	-0.914	0.007	5.285***	5.304***	5.285***	2.639	**[0.003 0.020]**
	e-p	685	-0.970	0.003	2.834***	6.653***	3.003**	3.881***	[-0.000 0.018]
Panel E: 1952-2018									
Annual	d-p	67	-0.910	0.036	0.801	2.429*	1.646	3.580**	[-0.047 0.173]
	e-p	67	-0.967	-0.003	-0.149	3.278***	0.517	3.158**	[-0.027 0.183]
Quarterly	d-p	268	-1.011	-0.003	-0.901	3.245***	0.882	3.730***	[-0.019 0.021]
	e-p	268	-0.998	-0.001	-0.318	3.994***	0.211	3.537**	[-0.013 0.030]
Monthly	d-p	804	-0.933	-0.001	-0.898	2.717**	0.474	2.726	[-0.006 0.006]
	e-p	804	-0.980	0.000	0.129	4.613***	0.129	3.846***	[-0.004 0.008]
Panel F: 1952-2002									
Annual	d-p	52	-0.882	0.061	1.280*	2.429*	1.327	3.653**	[-0.039 0.197]
	e-p	52	-0.954	0.008	0.398	3.278***	0.398	3.216**	[-0.038 0.199]
Quarterly	d-p	205	-1.026	0.011	3.194***	3.245***	3.194**	3.369**	[-0.009 0.033]
	e-p	205	-0.992	0.002	0.626	3.994***	0.626	3.488**	[-0.011 0.046]
Monthly	d-p	613	-0.936	0.004	2.687***	2.717**	2.723**	2.661	[-0.003 0.010]
	e-p	613	-0.972	0.001	1.015	4.613***	1.015	3.815***	[-0.004 0.012]

Chapter 2

Theory of Low Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference

2.1 Introduction

Economic and financial time series are highly nonstationary [see, e.g., Perron (1989), Stock and Watson (1996b), Ng and Wright (2013), and Giacomini and Rossi (2015)]. We develop theoretical results about the behavior of the sample autocovariance ($\widehat{\Gamma}(k)$, $k \in \mathbb{Z}$) and the periodogram ($I_T(\omega)$, $\omega \in [-\pi, \pi]$) for a short memory nonstationary process.[1] We show that nonstationarity (e.g., time-variation in the mean or autocovariances) induces low frequency contamination, meaning that the sample autocovariance and the periodogram share features that are similar to those of a long memory series. We present explicit expressions for the asymptotic bias of these estimates, showing that it is always positive and increases with the degree of heterogeneity in the data.

The low frequency contamination can be explained as follows. For a short memory series, the autocorrelation function (ACF) is known to display exponential decay and to vanish as the lag length $k \to \infty$, and the periodogram is known to be finite

[1] By short memory nonstationary processes we mean processes that have non-constant moments and whose sum of absolute autocovariances is finite. The latter rules out processes with unbounded second moments (e.g., unit root). For unit root or trending time series, one has to first apply some differencing or de-trending technique.

at the origin. Under general forms of nonstationarity, we show theoretically that $\widehat{\Gamma}(k) = \lim_{T\to\infty} \Gamma_T(k) + d^*$, where $\Gamma_T(k) = T^{-1}\sum_{k+1}^{T} \mathbb{E}(V_t V_{t-k})$, $k \geq 0$ and $d^* > 0$ is independent of k. Assuming positive dependence for simplicity (i.e., $\lim_{T\to\infty} \Gamma_T(k) > 0$), this means that each sample autocovariance overestimates the true dependence in the data. The bias factor $d^* > 0$ depends on the type of nonstationarity. Interestingly, d^* in general does not vanish as $T \to \infty$. This yields a slower than exponential decay. In addition, since short memory implies $\Gamma_T(k) \to 0$ as $k \to \infty$, it follows that d^* generates long memory effects since $\widehat{\Gamma}(k) \approx d^* > 0$ as $k \to \infty$. As for the periodogram, $I_T(\omega)$, we show that under nonstationarity $\mathbb{E}(I_T(\omega)) \to \infty$ as $\omega \to 0$, a feature shared by long memory processes.

This low frequency contamination holds as an asymptotic approximation. We verify analytically the quality of the approximation to finite-sample situations. We distinguish cases where this contamination only occurs as a small-sample problem and cases where it continues to hold asymptotically. The former involves $d^* \approx 0$ asymptotically but a consistent estimate of d^* satisfies $\widehat{d^*} > 0$ in finite-sample. We show analytically that using $\widehat{d^*}$ in place of d^* provides a good approximation in this context. This helps us to explain the long memory effects in situations where asymptotically no low frequency contamination should occur. An example is a t-test on a regression coefficient in a correctly specified model with nonstationary errors that are serially correlated. Other examples include t-tests in the linear model with mild forms of misspecification that do not undermine the conditions for consistency of the least-squares estimator. Further, similar issues arise if one applies some prior de-trending techniques where the fitted model is not correctly specified (e.g., the data follow a nonlinear trend but one removes a linear trend). Yet another example for which our results are relevant is the case of outliers. In all these examples, $d^* \approx 0$ asymptotically but one observes enhanced persistence in finite-sample that can

affect the properties of heteroskedasticity and autocorrelation robust (HAR) inference. Most of the HAR inference in applied work (besides the t- and F-test in regression models) are characterized by nonstationary alternative hypotheses for which $d^* > 0$ even asymptotically. This class of tests is very large. Tests for forecast evaluation [e.g., Casini (2018), Diebold and Mariano (1995), Giacomini and Rossi (2009, 2010), Giacomini and White (2006), Perron and Yamamoto (2019) and West (1996)], tests and inference for structural change models [e.g., Andrews (1993), Bai and Perron (1998), Casini and Perron (2020c, 2020d, 2020b), Elliott and Müller (2007), and Qu and Perron (2007)], tests and inference in time-varying parameters models [e.g., Cai (2007) and Chen and Hong (2012)], tests and inference for regime switching models [e.g., Hamilton (1989) and Qu and Zhuo (2021)] and others are part of this class.

We propose a solution to these problems via nonparametric smoothing over time which we show theoretically to be robust to low frequency contamination. By applying nonparametric smoothing, we prove that the sample local autocovariance and the local periodogram do not exhibit long memory features. Nonparametric smoothing avoids mixing highly heterogeneous data coming from distinct nonstationary regimes as opposed to what the sample autocovariance and the periodogram do. It was introduced recently for robust inference under nonstationarity by Casini (2019). He proposed a new HAC estimator that applies nonparametric smoothing over time in order to account flexibly for nonstationarity. Our results provide a theoretical justification for his double-kernel heteroskedasticity and autocorrelation consistent (DK-HAC) estimators.

Our work is different from the literature on spurious persistence caused by the presence of level shifts or other deterministic trends. Perron (1990) showed two useful results: first, the presence of breaks in mean often induces spurious non-rejection of the unit root hypothesis; second, the presence of a level shift asymptotically biases

the estimate of the AR coefficient towards one. Bhattacharya et al. (1983) demonstrated that certain deterministic trends can induce the spurious presence of long memory. Via simulations, Lamoureux and Lastrapes (1990), and analytically, Hillebrand (2005), showed that when the mean of a GARCH process changes, the sum of the estimated AR parameters of the conditional variance converges in probability to one. Christensen and Varneskov (2017) and McCloskey and Hill (2017) provided methods to estimate the parameters of, respectively, fractional cointegrating vector and a stationary ergodic time series, robust to some forms of mean/trend changes. Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004) and Perron and Qu (2010) showed via simulations and theoretical arguments that changes in mean induce hyperbolically decaying autocorrelations and spectral density estimates that approach infinity at the null frequency. Our results are different from theirs in that we consider a more general problem and we allow for more general forms of nonstationarity using the segmented locally stationary framework of Casini (2019). Importantly, we provide a general solution to these problems and show theoretically its robustness to low frequency contamination. Finally, we discuss in detail the implications of our theory for HAR inference.

HAR inference relies on estimation of the long-run variance (LRV). The latter, from a time domain perspective, is equivalent to the sum of all autocovariances of some relevant process while, from a frequency domain perspective, is equal to 2π times an integrated time-varying spectral density at the zero frequency. From a time domain perspective, estimation involves a weighted sum of the sample autocovariances, while from a frequency domain perspective estimation is based on a weighted sum of the periodogram ordinates near the zero frequency. Therefore, our results on low frequency contamination for the sample autocovariances and the periodogram can have important implications in this context.

There are two main approaches in HAR inference which relies on whether the LRV estimator is consistent or not. The classical approach relies on consistency which results in HAC standard errors [cf. Newey and West (1987, 1994) and Andrews (1991)]. Inference is standard because HAR tests follow asymptotically standard distributions. It was shown early that HAC standard errors can result in oversized tests when there is substantial temporal dependence. This stimulated a second approach based on inconsistent LRV estimators that keep the bandwidth at a fixed fraction of the sample size [cf. Kiefer et al. (2000)]. Inference becomes nonstandard and it is shown that fixed-b achieves high-order refinements [e.g., Sun et al. (2008)] and reduces the oversize problem of HAR tests.[2] However, unlike the classical approach, fixed-b HAR inference is only valid under stationarity.

Recent work by Casini (2019) questioned the performance of HAR inference tests under nonstationarity from a theoretical standpoint. In the past, simulation evidence of serious (e.g., non-monotonic) power or related issues in specific HAR inference contexts were documented by Altissimo and Corradi (2003), Casini (2018), Casini and Perron (2019, 2020c, 2020d), Chan (2020), Chang and Perron (2018), Crainiceanu and Vogelsang (2007), Deng and Perron (2006), Juhl and Xiao (2009), Kim and Perron (2009), Martins and Perron (2016), Perron and Yamamoto (2019) and Vogeslang (1999), among others]. Our theoretical results show that these issues occur because the unaccounted nonstationarity alters the spectrum at low frequencies. Each sample autocovariance is upward biased ($d^* > 0$) and the resulting LRV estimators tend to be inflated. When these estimators are used to normalize test statistics, the latter lose power. Interestingly, d^* is independent of k so that the more lags are included

[2]The fixed-b literature is vast. Various contributions can be found in Dou (2019), Lazarus et al. (2017), Lazarus et al. (2018), Gonçalves and Vogelsang (2011), de Jong and Davidson (2000), Ibragimov and Müller (2010), Jansson (2004), Kiefer and Vogelsang (2002, 2005), Müller (2007, 2014), Phillips (2005), Politis (2011), Preinerstorfer and Pötscher (2016), Pötscher and Preinerstorfer (2018, 2019), Robinson (1998), Sun (2013, 2014a, 2014b), Velasco and Robinson (2001) and Zhang and Shao (2013).

the more severe is the problem. Further, by virtue of weak dependence, we have that $\Gamma_T(k) \to 0$ as $k \to \infty$ but $d^* > 0$ across k. For these reasons, long bandwidths/fixed-b LRV estimators are expected to suffer most because they use many/all lagged autocovariance. Our theoretical results further show that nonparametric smoothing effectively solves the problem; the DK-HAC estimators from Casini (2019) lead to HAR tests with good size and good power even when existing HAR tests have little or no power.

This paper makes an independent contribution relative to other recent working papers by two of the authors. Casini and Perron (2020e) establishes minimax MSE bounds for LRV estimation and proposes a new prewhitening procedure robust to nonstationarity. Casini and Perron (2020a) presents change-point tests for time series with an evolutionary spectra that can also be used to eliminate the bias of the local periodogram when there are discontinuities in the spectrum. The paper is organized as follows. Section 2.2 presents the statistical setting and Section 2.3 establishes the theoretical results. The implications of our results for HAR inference are analyzed analytically and computationally through simulations in Section 2.4. Section 2.5 concludes. The code to implement our method is provided in Matlab, R and Stata languages through a Github repository.

2.2 Statistical Framework for Nonstationarity

Suppose $\{V_{t,T}\}_{t=1}^{T}$ is defined on an abstract probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is the sample space, $\mathscr{F}$ is the σ-algebra and $\mathbb{P}$ is a probability measure. In order to analyze time series models that have a time-varying spectrum it is useful to introduce an infill asymptotic setting whereby we rescale the original discrete time horizon $[1, T]$ by dividing each t by T. Letting $u = t/T$ and $T \to \infty$, we define a new time scale $u \in [0, 1]$ on which as $T \to \infty$ we observe more and more realiza-

tions of $V_{t,T}$ close to time t. As a notion of nonstationarity, we use the concept of segmented locally stationary processes introduced in Casini (2019). This extends the class of locally stationary processes [cf. Dahlhaus (1997)] that have been used widely in both statistics and economics, often simply referred to as time-varying parameter processes [see e.g., Cai (2007) and Chen and Hong (2012)]. Due to imposed smoothness restrictions, these processes exclude many prominent econometric models that account for time variation in the parameters. For example, structural change and regime switching-type models do not belong to this class because parameter changes occur suddenly at a particular point in time rather than smoothly over short periods. Segmented locally stationary processes allow for a finite number of discontinuities in the spectrum over time. We collect the break dates in the set $\mathcal{T} \triangleq \{T_1^0, \ldots, T_m^0\}$. Let $i \triangleq \sqrt{-1}$. A function $G(\cdot, \cdot) : [0, 1] \times \mathbb{R} \to \mathbb{C}$ is said to be left-differentiable at u_0 if $\partial G(u_0, \omega) / \partial_- u \triangleq \lim_{u \to u_0^-} (G(u_0, \omega) - G(u, \omega)) / (u_0 - u)$ exists for any $\omega \in \mathbb{R}$. Section 2.2.1 introduces short-memory segmented locally stationary processes. The extension to long memory processes is presented in Section 2.2.2.

2.2.1 Short Memory Segmented Locally Stationary Processes

Definition 2.2.1 A sequence of stochastic processes $\{V_{t,T}\}_{t=1}^T$ is called segmented locally stationary (SLS) with $m_0 + 1$ regimes, transfer function A^0 and trend $\mu_\cdot$ if there exists a representation

$$V_{t,T} = \mu_j(t/T) + \int_{-\pi}^{\pi} \exp(i\omega t) A_{j,t,T}^0(\omega) \, d\xi(\omega), \qquad \left(t = T_{j-1}^0 + 1, \ldots, T_j^0\right), \tag{2.1}$$

for $j = 1, \ldots, m_0 + 1$, where by convention $T_0^0 = 0$ and $T_{m_0+1}^0 = T$. The following technical conditions are also assumed to hold:

(i) $\xi(\lambda)$ is a stochastic process on $[-\pi, \pi]$ with $\overline{\xi(\omega)} = \xi(-\omega)$ and

$$\operatorname{cum}\{d\xi(\omega_1), \ldots, d\xi(\omega_r)\} = \varphi\left(\sum_{j=1}^{r} \omega_j\right) g_r(\omega_1, \ldots, \omega_{r-1})\, d\omega_1 \ldots d\omega_r,$$

where $\operatorname{cum}\{\cdots\}$ denotes the cumulant spectra of r-th order, $g_1 = 0$, $g_2(\omega) = 1$, $|g_r(\omega_1, \ldots, \omega_{r-1})| \leq M_r$ for all r with M_r being a constant that may depend on r, and $\varphi(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j)$ is the period 2π extension of the Dirac delta function $\delta(\cdot)$.

(ii) There exists a constant K and a piecewise continuous function $A : [0, 1] \times \mathbb{R} \to \mathbb{C}$ such that, for each $j = 1, \ldots, m_0 + 1$, there exists a 2π-periodic function $A_j : (\lambda_{j-1}^0, \lambda_j^0] \times \mathbb{R} \to \mathbb{C}$ with $A_j(u, -\omega) = \overline{A_j(u, \omega)}$, $\lambda_j^0 \triangleq T_j^0/T$ and for all T,

$$A(u, \omega) = A_j(u, \omega) \text{ for } \lambda_{j-1}^0 < u \leq \lambda_j^0, \tag{2.2}$$

$$\sup_{1 \leq j \leq m_0+1} \sup_{T_{j-1}^0 < t \leq T_j^0, \omega} \left|A_{j,t,T}^0(\omega) - A_j(t/T, \omega)\right| \leq KT^{-1}. \tag{2.3}$$

(iii) $\mu_j(t/T)$ is piecewise continuous.

Assumption 2.2.1 (i) $\{V_{t,T}\}$ is a mean-zero SLS process with $m_0 + 1$ regimes; (ii) $A(u, \omega)$ is twice continuously differentiable in u at all $u \neq \lambda_j^0$, $j = 1, \ldots, m_0 + 1$, with uniformly bounded derivatives $(\partial/\partial u) A(u, \cdot)$ and $(\partial^2/\partial u^2) A(u, \cdot)$, and Lipschitz continuous in the second component with index $\vartheta = 1$; (iii) $(\partial^2/\partial u^2) A(u, \cdot)$ is Lipschitz continuous at all $u \neq \lambda_j^0$ $(j = 1, \ldots, m_0 + 1)$; (iv) $A(u, \omega)$ is twice left-differentiable in u at $u = \lambda_j^0$ $(j = 1, \ldots, m_0 + 1)$ with uniformly bounded derivatives $(\partial/\partial_- u) A(u, \cdot)$ and $(\partial^2/\partial_- u^2) A(u, \cdot)$ and has piecewise Lipschitz continuous derivative $(\partial^2/\partial_- u^2) A(u, \cdot)$.

We define the time-varying spectral density as $f_j(u, \omega) \triangleq |A_j(u, \omega)|^2$ for $T_{j-1}^0/T < u = t/T \leq T_j^0/T$. Given $f(u, \omega)$, we can define the local covariance of $V_{t,T}$ at the rescaled time u with $Tu \notin \mathcal{T}$ and lag $k \in \mathbb{Z}$ as $c(u, k) \triangleq \int_{-\pi}^{\pi} e^{i\omega k} f(u, \omega)\, d\omega$. The same definition is also used when $Tu \in \mathcal{T}$ and $k \geq 0$. For $Tu \in \mathcal{T}$ and $k < 0$ it is defined as $c(u, k) \triangleq \int_{-\pi}^{\pi} e^{i\omega k} A(u, \omega) A(u - k/T, -\omega)\, d\omega$.

Next, we impose conditions on the temporal dependence (we omit the second

subscript T when it is clear from the context). Let

$$\begin{aligned}
&\kappa_{V,t}^{(a_1,a_2,a_3,a_4)}(u,\, v,\, w) \\
&\triangleq \kappa^{(a_1,a_2,a_3,a_4)}(t,\, t+u,\, t+v,\, t+w) - \kappa_{\mathscr{N}}^{(a_1,a_2,a_3,a_4)}(t,\, t+u,\, t+v,\, t+w) \\
&\triangleq \mathbb{E}\left(V_t^{(a_1)} - \mathbb{E}V_t^{(a_1)}\right)\left(V_{t+u}^{(a_2)} - \mathbb{E}V_{t+u}^{(a_2)}\right)\left(V_{t+v}^{(a_3)} - \mathbb{E}V_{t+v}^{(a_3)}\right)\left(V_{t+w}^{(a_4)} - \mathbb{E}V_{t+w}^{(a_4)}\right) - \\
&\mathbb{E}\left(V_{\mathscr{N},t}^{(a_1)} - \mathbb{E}V_{\mathscr{N},t}^{(a_1)}\right)\left(V_{\mathscr{N},t+u}^{(a_2)} - \mathbb{E}V_{\mathscr{N},t+u}^{(a_2)}\right)\left(V_{\mathscr{N},t+v}^{(a_3)} - \mathbb{E}V_{\mathscr{N},t+v}^{(a_3)}\right)\left(V_{\mathscr{N},t+w}^{(a_4)} - \mathbb{E}V_{\mathscr{N},t+w}^{(a_4)}\right),
\end{aligned}$$

where $\{V_{\mathscr{N},t}\}$ is a Gaussian sequence with the same mean and covariance structure as $\{V_t\}$, $\kappa_{V,t}^{(a_1,a_2,a_3,a_4)}(u,\, v,\, w)$ is the time-t fourth-order cumulant of $(V_t^{(a_1)},\, V_{t+u}^{(a_2)},\, V_{t+v}^{(a_3)},\, V_{t+w}^{(a_4)})$ while $\kappa_{\mathscr{N}}^{(a_1,a_2,a_3,a_4)}(t,\, t+u,\, t+v,\, t+w)$ is the time-t centered fourth moment of V_t if V_t were Gaussian.

Assumption 2.2.2 (i) $\sum_{k=-\infty}^{\infty}\sup_{u\in[0,\,1]}\|c(u,\, k)\| < \infty$ and $\sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sup_{u\in[0,\,1]}|\kappa_{V,\lfloor Tu\rfloor}^{(a_1,a_2,a_3,a_4)}(k,\, j,\, l)| < \infty$ for all $a_1, a_2, a_3, a_4 \leq p$. (ii) For all $a_1, a_2, a_3, a_4 \leq p$ there exists a function $\widetilde{\kappa}_{a_1,a_2,a_3,a_4} : [0,\, 1]\times\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}\to\mathbb{R}$ such that $\sup_{1\leq j\leq m_0+1}\sup_{\lambda_{j-1}^0<u\leq\lambda_j^0}|\kappa_{V,\lfloor Tu\rfloor}^{(a_1,a_2,a_3,a_4)}(k,\, s,\, l) - \widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\, k,\, s,\, l)| \leq LT^{-1}$ for some constant L; the function $\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\, k,\, s,\, l)$ is twice differentiable in u at all $u \neq \lambda_j^0$ $(j = 1, \ldots,\, m_0+1)$ with uniformly bounded derivatives $(\partial/\partial u)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$ and $(\partial^2/\partial u^2)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$, and twice left-differentiable in u with uniformly bounded derivatives $(\partial/\partial_- u)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$ and $(\partial^2/\partial_- u^2)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$, and piecewise Lipschitz continuous derivative $(\partial^2/\partial_- u^2)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$.

If $\{V_t\}$ is stationary then the cumulant condition of Assumption 2.2.2-(i) reduces to the standard one used in the time series literature [see Andrews (1991)]. Note that α-mixing and moment conditions imply that the cumulant condition of Assumption 2.2.2 holds.

2.2.2 Long Memory Segmented Locally Stationary Processes

One of our goals is to show that the sample autocovariances and the periodogram based on short memory SLS processes have properties similar to those of of long

memory processes. Thus, we need to first define long memory SLS processes and illustrate the properties of these statistics for such processes. Define the backward difference operator $\Delta V_t = \Delta^1 V_t = V_t - V_{t-1}$ and $\Delta^l V_t$ recursively. Long memory features can be expressed as a "pole" in the spectral density at frequency zero. That is, for a stationary process, long memory implies that $f(\omega) \sim \omega^{-2\vartheta}$ as $\omega \to 0$ where $\vartheta \in (0,\, 1/2)$ is the long memory parameter. In what follows, l is some non-negative integer.

Definition 2.2.2 A sequence of stochastic processes $\{V_{t,T}\}$ is called long memory segmented locally stationary with $m_0 + 1$ regimes, transfer function A^0 and trend $\mu_\cdot$ if there exists a representation

$$\Delta^l V_t = \mu_j(t/T) + \int_{-\pi}^{\pi} \exp(i\omega t)\, A^0_{j,t,T}(\omega)\, d\xi(\omega), \qquad \left(t = T^0_{j-1}+1, \ldots, T^0_j\right), \tag{2.4}$$

for $j = 1, \ldots, m_0 + 1$, where by convention $T^0_0 = 0$ and $T^0_{m_0+1} = T$, (i) and (iii) of Definition 2.2.1 hold, and (ii) of Definition 2.2.1 is replaced by

(ii) There exists two constants $L_2 > 0$ and $D < 1/2$ (which depend on j) and a piecewise continuous function $A : [0,\, 1] \times \mathbb{R} \to \mathbb{C}$ such that, for each $j = 1, \ldots, m_0+1$, there exists a 2π-periodic function $A_j : (\lambda^0_{j-1},\, \lambda^0_j] \times \mathbb{R} \to \mathbb{C}$ with $A_j(u, -\omega) = \overline{A_j(u,\, \omega)}$,

$$A(u,\, \omega) = A_j(u,\, \omega) \text{ for } \lambda^0_{j-1} < u \le \lambda^0_j, \tag{2.5}$$

$$\sup_{1\le j\le m_0+1} \sup_{T^0_{j-1}<t\le T^0_j,\, \omega} \left|A^0_{j,t,T}(\omega) - A_j(t/T,\, \omega)\right| \le L_2 T^{-1} |\omega|^{-D}, \tag{2.6}$$

and

$$\sup_{0\le v\le u\le 1,\, u\ne\lambda^0_j\,(j=1,\ldots,\, m_0+1,),\, \omega} |A(u,\, \omega) - A(u',\, \omega)| \le L_2 |u - u'|\, |\omega|^{-D}. \tag{2.7}$$

The spectral density of $\{V_{t,T}\}$ is given by $f_j(u,\, \omega) = |1 - \exp(-i\omega)|^{-2l} |A_j(u,\, \omega)|^{-2}$ for $j = 1, \ldots, m_0 + 1$. We say that the process $\{V_{t,T}\}$ has local memory parameter

$\vartheta(u) \in (-\infty, l+1/2)$ at time $u \in [0, 1]$ if it satisfies (2.4)-(2.7), and its generalized spectral density $f_j(u, \omega)$ $(j = 1, \ldots, m_0+1)$ satisfies the following condition,

$$f_j(u, \omega) = \left|1 - e^{-i\omega}\right|^{-2\vartheta_j(u)} f_j^*(u, \omega), \tag{2.8}$$

with $f_j^*(u, \omega) > 0$ and

$$\left|f_j^*(u, \omega) - f_j^*(u, 0)\right| \leq L_4 f_j^*(u, \omega) |\omega|^{\nu}, \qquad \omega \in [-\pi, \pi], \tag{2.9}$$

where $L_4 > 0$ and $\nu \in (0, 2]$.

Definition 2.2.2 extends Definition 2.2.1 and Assumption 2.2.1 by requiring the bound on the smoothness of $A(\cdot, \omega)$ to depend also on $|\omega|^{-D}$ thereby allowing a singularity at $\omega = 0$. Casini (2019) showed that $f_j(u, \omega) = |A_j(u, \omega)|^2$ for $j = 1, \ldots, m_0+1$. Using similar arguments, we obtain the form $f_j(u, \omega)$ given in (2.8). See Roueff and von Sachs (2011) for a definition of long memory local stationarity. Definition 2.2.2 extends their definition to allow for m_0 discontinuities. We have assumed that breaks in the long memory parameter occur at the same locations as the breaks in the spectrum. This can be relaxed but would provide no added value in this paper.

Example 2.2.1 A time-varying AR fractionally integrated moving average (p, ϑ, q) process with m_0 structural breaks satisfies Definition 2.2.2 with
$\vartheta_j : [0, 1] \to (-\infty, l+1/2)$, $\sigma_j : [0, 1] \to \mathbb{R}_+$, $\phi_j = [\phi_1, \ldots, \phi_p]' : [0, 1] \to \mathbb{R}^q$ and $\theta_j = [\theta_1, \ldots, \theta_q]' : [0, 1] \to \mathbb{R}^p$ are left-Lipschitz functions for each $j = 1, \ldots, m_0+1$ such that $1 - \sum_{k=1}^{p} \phi_{j,k}(u) z^k$ does not vanish for all $u \in [0, 1]$ and $z \in \mathbb{C}$ such that $|z| \leq 1$. Using the latter condition, the local transfer function $A_j(u; \cdot)$ defines for each j a causal autoregressive fractionally integrated moving average (ARFIMA$(p, \vartheta(u) - l, q)$ process whose spectral density satisfies the conditions (2.8) and (2.9) with $\nu = 2$. Using Lemma 3 in Roueff and von Sachs (2011), condition (2.7) holds with with $D > \sup_{1 \leq j \leq m_0+1} \sup_{\lambda_{j-1}^0 < u \leq \lambda_j^0, \omega} \vartheta_j(u) - l$.

Definition 2.2.2 implies that $\rho_V(u, k) \triangleq \text{Corr}(V_{\lfloor Tu \rfloor}, V_{\lfloor Tu \rfloor + k}) \sim Ck^{2d_j(u)-1}$ for $\lambda_{j-1}^0 < u < \lambda_j^0$ and large k where $C > 0$. This means that the rescaled time-u autocor-

relation function (ACF(u)) has a power law decay which implies $\sum_{k=-\infty}^{\infty} |\rho_V(u,\, k)| = \infty$ if $d_j(u) \in (0,\, 1/2)$.

2.3 Theoretical Results on Low Frequency Contamination

In this section we establish theoretical results about the low frequency contamination induced by nonstationarity including results covering the case of misspecification and outliers. We first consider the asymptotic proprieties of two key quantities for inference in time series contexts, i.e., the sample autocovariance and the periodogram. These are defined, respectively, by

$$\widehat{\Gamma}(k) = T^{-1} \sum_{t=|k|+1}^{T} \left(V_t - \overline{V}\right)\left(V_{t-|k|} - \overline{V}\right), \tag{2.10}$$

where $\overline{V}$ is the sample mean and

$$I_T(\omega) = \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \exp(-i\omega t)\, V_t \right|^2, \qquad \omega \in [0,\, \pi],$$

which is evaluated at the Fourier frequencies $\omega_j = (2\pi j)/T \in [0,\, \pi]$. In the context of autocorrelated data, hypotheses testing and construction of confidence intervals require estimation of the so-called long-run variance. Traditional HAC estimators are weighted sums of sample autocovariances while frequency domain estimators are weighted sums of the periodograms. Casini (2019) considered an alternative estimate for the sample autocovariance to be used in the DK-HAC estimators,[3] namely,

$$\widehat{\Gamma}_{\mathrm{DK}}(k) \triangleq \frac{n_T}{T} \sum_{r=1}^{\lfloor T/n_T \rfloor} \widehat{c}_T(r n_T/T,\, k),$$

[3]The DK-HAC estimators are defined in Section 2.4.1.

where $k \in \mathbb{Z}$, $n_T \to \infty$ satisfying the conditions given below, and

$$\begin{aligned}&\widehat{c}_T\left(rn_T/T,\, k\right)\\ =&n_{2,T}^{-1}\sum_{s=0}^{n_{2,T}-1}\left(V_{rn_T+\lfloor|k/2|\rfloor-n_{2,T}/2+s+1}-\overline{V}_{rn_T,T}\right)\left(V_{rn_T-\lfloor|k/2|\rfloor-n_{2,T}+s+1}-\overline{V}_{rn_T,T}\right),\end{aligned} \tag{2.11}$$

where $n_{2,T} \to \infty$ and $\overline{V}_{rn_T,T} = n_{2,T}^{-1}\sum_{s=0}^{n_{2,T}-1} V_{rn_T-n_{2,T}/2+s+1}$. For notational simplicity we assume that n_T and $n_{2,T}$ are even. $\widehat{c}_T\left(rn_T/T,\, k\right)$ is an estimate of the autocovariance at time rn_T and lag k, i.e., $\operatorname{cov}\left(V_{rn_T},\, V_{rn_T-k}\right)$. One could use a smoothed or tapered version. However, to clearly focus on the main intuition, we prefer to omit these estimates to keep the notation simple. The theoretical results remain the same. The estimate $\widehat{\Gamma}_{\mathrm{DK}}\left(k\right)$ is an integrated local sample autocovariance. It extends $\widehat{\Gamma}\left(k\right)$ to better account for nonstationarity. Similarly, the DK-HAC estimator does not relate to the periodogram but to the local periodogram defined by

$$I_{\mathrm{L},T}\left(u,\, \omega\right) \triangleq \left|\frac{1}{\sqrt{n_T}}\sum_{s=0}^{n_T-1} V_{\lfloor Tu\rfloor-n_T/2+s+1,T}\exp\left(-i\omega s\right)\right|^2,$$

where $I_{\mathrm{L},T}\left(u,\, \omega\right)$ is the (untapered) periodogram over a segment of length n_T with midpoint $\lfloor Tu\rfloor$. We also consider the statistical properties of both $\widehat{\Gamma}_{\mathrm{DK}}\left(k\right)$ and $I_{\mathrm{L},T}\left(u,\, \omega\right)$ under nonstationarity. Define $r_j = \left(\lambda_j^0 - \lambda_{j-1}^0\right)$ for $j = 1, \ldots, m_0+1$ with $\lambda_0^0 = 0$ and $\lambda_{m_0+1}^0 = 1$. Note that $\lambda_j^0 = \sum_{s=0}^{j} r_s$.

2.3.1 The Sample Autocovariance under Nonstationarity

We now establish some asymptotic properties of the sample autocovariance under nonstationarity. We consider the case $k \geq 0$ only; the case $k < 0$ is similar and

omitted. Let

$$\overline{\mu}_j = r_j^{-1} \int_{\lambda_{j-1}^0}^{\lambda_j^0} \mu_j(u)\, du, \qquad \text{for } j = 1, \ldots, m_0 + 1,$$

where $\mu_j(\cdot)$ is defined in (2.1). We use $\sum_{j_1 \neq j_2}$ as a shorthand for the double sum $\sum_{\{j_1, j_2 = 1, \ldots, m_0+1,\, j_1 \neq j_2\}}$.

Theorem 2.3.1 *Assume that $\{V_{t,T}\}$ satisfies Definition 2.2.1. Under Assumption 2.2.1-2.2.2,*

$$\widehat{\Gamma}(k) \geq \int_0^1 c(u,\, k)\, du + d^* + o_{\text{a.s.}}(1), \tag{2.12}$$

where $d^* = 2^{-1} \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} (\overline{\mu}_{j_2} - \overline{\mu}_{j_1})^2$. Further, as $k \to \infty$, $\widehat{\Gamma}(k) \geq d^*$ $\mathbb{P}$-a.s. If in addition it holds that $\mu_j(t/T) = \mu_j$ for $j = 1, \ldots, m_0 + 1$, then

$$\widehat{\Gamma}(k) = \int_0^1 c(u,\, k)\, du + d^*_{\text{Sta}} + o_{\text{a.s.}}(1),$$

where $d^*_{\text{Sta}} = 2^{-1} \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} (\mu_{j_2} - \mu_{j_1})^2$ and, as $k \to \infty$, $\widehat{\Gamma}(k) = d^*_{\text{Sta}} + o_{\text{a.s.}}(1)$.

Theorem 2.3.1 reveals interesting features. It is easier for illustrative purposes to discuss the case $\mu_j(t/T) = \mu_j$ for $j = 1, \ldots, m_0 + 1$ for which the mean of V_t in each regime is constant. The theorem states that $\widehat{\Gamma}(k)$ is asymptotically the sum of two terms. The first is the true autocovariance of $\{V_t\}$ at lag k. The second depends on the difference in the mean of $\{V_t\}$ across regimes. This term is always positive and it increases in magnitude with the difference in the mean across regimes. Thus, nonstationarity (here in the form of structural breaks in the mean) induces a positive bias. In the next section, we shall discuss cases in which this bias arises as a finite-sample problem and cases where the bias remains even asymptotically. The result that $\widehat{\Gamma}(k) = d^* + o_{\text{a.s.}}(1)$ as $k \to \infty$ implies that unaccounted nonstationarity generates

long memory effects. The intuition is straightforward. In the previous section we defined a long memory SLS process via the condition $\sum_{k=-\infty}^{\infty} \Gamma(u, k) \to \infty$ for some $u \in (0, 1)$. Similarly, a stationary long memory process implies $\sum_{k=-\infty}^{\infty} \Gamma(k) \to \infty$. The theorem shows that $\widehat{\Gamma}(k)$ exhibits a similar property. Thus, $\widehat{\Gamma}(k)$ decays more slowly than for a corresponding short memory stationary process for small lags and then approaches a strictly positive constant d^* for large lags. A similar result for the case $\mu_j(t/T) = \mu_j$ was discussed under stationarity in Mikosch and Stărică (2004) to explain long memory in the volatility of financial returns. Their result is driven by abrupt breaks in the second moments of a stationary process. Our result is more general since it allows for general forms of nonstationarity and has many empirical implications.

Theorem 2.3.1 suggests that any deviation from stationarity can generate a long memory component d^* or d^*_{Sta} that leads to overestimation of the true autocovariance. That is, either a stationary model with breaks or a locally stationary model or a combination of the two can generate these issues. It follows that also the LRV is overestimated. Since the LRV is used to normalized test statistics, this has important consequences for many HAR inference tests that are characterized by deviations from stationarity under the alternative hypotheses. These include tests for forecast evaluation, tests and inference for structural change models, time-varying parameters models and regime switching models. In the context of the linear regression model, V_t corresponds to the least-squares residuals. Thus, Theorem 2.3.1 is relevant for regression models with mild forms of misspecification that do not undermine the conditions for consistency of the least-squares estimator. In those cases, the misspecification contaminates the residuals and so generates a long memory component. Examples include exclusion of a relevant regressor uncorrelated with the included reggressors, or inclusion of an irrelevant regressor. Also unaccounted nonlinearities and outliers can

contaminate the mean of V_t and therefore contribute to d^*. The difference between d^* and d^*_{Sta} is that the latter is generated only by structural breaks or regime switching.

2.3.2 The Periodogram under Nonstationarity

Classical LRV estimators are weighted averages of periodogram ordinates around the zero frequency. Thus, it is useful to study the behavior of the periodogram as the frequency ω approaches zero. We now establish some properties of the asymptotic bias of the periodogram under nonstationarity. We consider the Fourier frequencies $\omega_l = 2\pi l/T \in (-\pi, \pi)$ for an integer $l \neq 0 \pmod T$ and exclude $\omega_l = 0$ for mathematical convenience.

Assumption 2.3.1 (i) For each $j = 1, \ldots, m_0 + 1$ there exists a $B_j \in \mathbb{R}$ such that

$$\left| \sum_{j=1}^{m_0+1} \sum_{t=\left\lfloor T\lambda^0_{j-1} \right\rfloor + 1}^{\left\lfloor T\lambda^0_j \right\rfloor} \mu_j\left(t/T\right) \exp\left(-i\omega_l t\right) \right|^2 \geq \left| \sum_{j=1}^{m_0+1} B_j \sum_{t=\left\lfloor T\lambda^0_{j-1} \right\rfloor + 1}^{\left\lfloor T\lambda^0_j \right\rfloor} \exp\left(-i\omega_l t\right) \right|^2,$$
$$\omega_l \in (-\pi, \pi),$$

where $B_{j_1} \neq B_{j_2}$ for $j_1 \neq j_2$; (ii) $|\Gamma(u, k)| = C_{u,k} k^{-m}$ for all $u \in [0, 1]$ and all $k \geq C_3 T^{\kappa}$ for some $C_3 < \infty$, $C_{u,k} < \infty$ (which depends on u and k), $0 < \kappa < 1/2$, and $m > 2$.

Part (i) is easily satisfied (e.g., the special case with $\mu_j(t/T) = \mu_j$). Part (ii) is satisfied if $\{V_t\}$ is strong mixing with mixing parameters of size $-2\nu/(\nu - 1/2)$ for some $\nu > 1$ such that $\sup_{t \geq 1} \mathbb{E}|V_t|^{4\nu} < \infty$. This is less stringent than the size condition sufficient for Assumption 2.2.2-(i).

Theorem 2.3.2 *Assume that $\{V_{t,T}\}$ satisfies Definition 2.2.1. Under Assumption 2.2.1-2.2.2 and 2.3.1,*

$$\mathbb{E}\left(I_T(\omega_l)\right) = 2\pi \int_0^1 f(u, \omega_l)\, du \tag{2.13}$$

$$+ \frac{1}{T\omega_l^2} \left| \left[B_1 - B_{m_0+1} - \sum_{j=1}^{m_0} (B_j - B_{j+1}) \exp\left(-2\pi i l \lambda_l^0\right) \right] \right|^2 + o(1).$$

Under Assumption 2.2.1-2.2.2 and 2.3.1-(ii), if $\mu_j(t/T) = \mu_j$ *for each* $j = 1, \ldots, m_0+1$*, then*

$$\mathbb{E}(I_T(\omega_l)) = 2\pi \int_0^1 f(u, \omega_l)\, du + \frac{1}{T\omega_l^2} \left| \left[\mu_j - \mu_{m_0+1} - \sum_{j=1}^{m_0} (\mu_j - \mu_{j+1}) \exp\left(-2\pi i l \lambda_j^0\right) \right] \right|^2 + o(1).$$

In either case, if $T\omega_l^2 \to 0$ *as* $T \to \infty$ *then* $\mathbb{E}(I_T(\omega_l)) \to \infty$ *for many values in* $\{\omega_l\}$ *as* $\omega_l \to 0$.

The theorem suggests that for small frequencies ω_l close to 0, the periodogram attains very large values. This follows because the first term of (2.13) is bounded for all frequencies ω_j. Since $B_1, \ldots, B_{m_0+1}$ are fixed, the order of the second term of (2.13) is $O((T\omega_j^2)^{-1})$. Note that as $\omega_l \to 0$ there are some values l for which the corresponding term involving $|\cdot|^2$ on the right-hand side of (2.13) is equal to zero. In such cases, $\mathbb{E}(I_T(\omega_l)) \geq 2\pi \int_0^1 f(u, \omega_l)\, du > 0$. For other values of $\{l\}$ as $\omega_l \to 0$, the second term of (2.13) diverges to infinity. Thus, considering the behavior of the sequence $\{\mathbb{E}(I_T(\omega_l))\}$ as $\omega_l \to 0$, it generally takes arbitrary unbounded values except for some ω_l for which $\mathbb{E}(I_T(\omega_l))$ is bounded below by $2\pi \int_0^1 f(u, \omega_l)\, du > 0$. This behavior is consistent with long memory as discussed in the previous section. A SLS process with long memory has an unbounded local spectral density $f(u, \omega)$ as $\omega \to 0$ for some $u \in [0, 1]$. Since $f(\cdot, \cdot)$ cannot be negative, it follows that $\int_0^1 f(u, \omega)\, du$ is also unbounded as $\omega \to 0$. Theorem 2.3.2 suggests that nonstationarity consisting of time-varying first moment results in a periodogram sharing features of the periodogram of a long memory series around the zero frequency. Since the periodogram behavior around the zero frequency characterizes a long memory process, nonstationarity can

generate long memory effects.

2.3.3 The Sample Local Autocovariance under Nonstationarity

We now consider the behavior of $\widehat{c}_T(rn_T/T,\, k)$ as defined in (2.11) for fixed k as well as for $k \to \infty$. For notational simplicity we assume that k is even.

Theorem 2.3.3 *Assume that $\{V_{t,T}\}$ satisfies Definition 2.2.1, $n_T,\, n_{2,T} \to \infty$ with $n_T/T \to 0$ and $n_T/n_{2,T} \to 0$. Under Assumption 2.2.1-2.2.2,*

(i) for any $u \in (0,\, 1)$ such that
$T_j^0 \notin [\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1, \ldots,\, \lfloor Tu \rfloor + k/2 + n_{2,T}/2]$
for all $j = 1, \ldots, m_0$, $\widehat{c}_T(u,\, k) = c(u,\, k) + o_{\mathbb{P}}(1)$;

(ii) for any $u \in (0,\, 1)$ such that
$T_j^0 \in [\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1, \ldots,\, \lfloor Tu \rfloor + k/2 + n_{2,T}/2]$
for some $j = 1, \ldots, m_0$ we have two sub-cases: (a) if $(T_j^0 - (\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1))/n_{2,T} \to \gamma$ or $((\lfloor Tu \rfloor + k/2 + n_{2,T}/2 + 1) - T_j^0)/n_{2,T} \to \gamma$ with $\gamma \in (0,\, 1)$, then

$$\widehat{c}_T(u,\, k) \geq \gamma c\left(\lambda_j^0,\, k\right) + (1-\gamma)\, c(u,\, k) + \gamma(1-\gamma)\left(\mu_j\left(\lambda_j^0\right) - \mu_{j+1}(u)\right)^2 + o_{\mathbb{P}}(1).$$

(b) if $(T_j^0 - (\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1))/n_{2,T} \to 0$ or $((\lfloor Tu \rfloor + k/2 + n_{2,T}/2 + 1) - T_j^0)/n_{2,T} \to 0$, then $\widehat{c}_T(u,\, k) = c(u,\, k) + o(1)$.

Further, if there exists an $r = 1, \ldots, \lfloor T/n_T \rfloor$ such that there exists a $j = 1, \ldots, m_0$ with $T_j^0 \in [rn_T + k/2 - n_{2,T}/2 + 1, \ldots, rn_T + k/2 + n_{2,T}/2]$ satisfying (ii-a), then as $k \to \infty$ $\widehat{\Gamma}_{\mathrm{DK}}(k) \geq d_T^*$ $\mathbb{P}$-a.s. where $d_T^* = (n_T/T)\,\gamma(1-\gamma)\,(\mu_j(\lambda_j^0) - \mu_{j+1}(u))^2 > 0$ and $d_T^* \to 0$ as $T \to \infty$.

The theorem shows that the behavior of $\widehat{c}_T(u,\, k)$ depends on whether a change-point in mean is present or not, and if present whether it is close enough to $\lfloor Tu \rfloor$ or not. For a given $u \in (0,\, 1)$ and $k \in \mathbb{Z}$, if the condition of part (i) of the theorem holds, then $\widehat{c}_T(u,\, k)$ is consistent for $\mathrm{cov}(V_{\lfloor Tu \rfloor} V_{\lfloor Tu \rfloor - k}) = c(u,\, k) + O(T^{-1})$ [see Casini (2019)]. If a change-point falls close to either boundary of the window $[\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1, \ldots, \lfloor Tu \rfloor + k/2 + n_{2,T}/2]$, as specified in case (ii-b), then $\widehat{c}_T(u,\, k)$ remains consistent. The only case in which a non-negligible bias arises is

when the change-point falls in the neighborhood around $\lfloor Tu \rfloor$ sufficiently far from either boundary. This represents case (ii-a), for which a biased estimate results. However, the bias vanishes asymptotically. Since $\widehat{\Gamma}_{\mathrm{DK}}(k)$ is an average of $\widehat{c}_T(rn_T, k)$ over blocks $r = 1, \ldots, \lfloor T/n_T \rfloor$, if case (ii-a) holds then $\widehat{\Gamma}_{\mathrm{DK}}(k) \geq d_T^*$ as $k \to \infty$ but $d_T^* \to 0$ as $T \to \infty$. Thus, comparing this result with Theorem 2.3.1, in practice the long memory effects are unlikely to occur when using $\widehat{\Gamma}_{\mathrm{DK}}(k)$ instead of $\widehat{\Gamma}(k)$. Furthermore, one can avoid altogether this issue for $\widehat{\Gamma}_{\mathrm{DK}}(k)$ by appropriately choosing the blocks $r = 1, \ldots, \lfloor T/n_T \rfloor$. A procedure was proposed in Casini (2019) using the methods developed in Casini and Perron (2020a). Another way to see that $\widehat{\Gamma}_{\mathrm{DK}}(k)$ suffers less from these problems is to look at the form of $\widehat{c}_T(rn_T, k)$. Usually one would use a kernel or a taper to assign more weight to observations that are close to $\lfloor Tu \rfloor$. This automatically would reduce the contamination which arises from mixing observations belonging to two different regimes because the shorter regime would be down-weighted by the kernel or taper.

2.3.4 The Local Periodogram under Nonstationarity

We now study the asymptotic properties of $I_{\mathrm{L},T}(u, \omega)$ as $\omega \to 0$ for $u \in [0, 1]$. We consider the Fourier frequencies $\omega_l = 2\pi l/n_T \in (-\pi, \pi)$ for an integer $l \neq 0 \pmod{n_T}$. We need the following high-level conditions.

Assumption 2.3.2 (i) For each $u \in [0, 1]$ with
$T_j^0 \in [\lfloor Tu \rfloor - n_T/2 + 1, \ldots, \lfloor Tu \rfloor + n_T/2]$ there exist $B_j \in \mathbb{R}$, $j = 1, \ldots, m_0$ with $B_{j_1} \neq B_{j_2}$ for $j_1 \neq j_2$ such that

$$\left| \sum_{s=0}^{n_T-1} \mu\left(\left(\lfloor Tu \rfloor - n_T/2 + s + 1\right)/T\right) \exp\left(-i\omega_l s\right) \right|^2 \geq$$
$$\left| B_j \sum_{s=0}^{T_j^0 - (\lfloor Tu \rfloor - n_T/2 + 1)} \exp\left(-i\omega_l s\right) + B_{j+1} \sum_{s=T_j^0 - (\lfloor Tu \rfloor - n_T/2)}^{n_T-1} \exp\left(-i\omega_l s\right) \right|^2 .$$

(ii) $\sup_{u\in[0,\,1]}\left(\partial^2/\partial u^2\right) f\left(u,\,\omega\right)$ is continuous in ω.

Theorem 2.3.4 *Assume that $\{V_{t,T}\}$ satisfies Definition 2.2.1 and that $n_T \to \infty$ with $n_T/T \to 0$. Under Assumption 2.2.1-2.2.2, 2.3.1-(ii) and 2.3.2,*

(i) for any $u \in (0,\,1)$ such that $T_j^0 \notin [\lfloor Tu\rfloor - n_T/2+1,\ldots,\,\lfloor Tu\rfloor + n_T/2]$ for all $j=1,\ldots,\,m_0$, $\mathbb{E}\left(I_{\mathrm{L},T}\left(u,\,\omega_l\right)\right) \geq f\left(u,\,\omega_l\right)$ as $\omega_l \to 0$;

(ii) for any $u \in (0,\,1)$ such that $T_j^0 \in [\lfloor Tu\rfloor - n_T/2+1,\ldots,\,\lfloor Tu\rfloor + n_T/2]$ for some $j=1,\ldots,\,m_0$ we have two sub-cases: (a) if $(T_j^0 - (\lfloor Tu\rfloor - n_T/2+1))/n_T \to \gamma$ or $(T_j^0 - (\lfloor Tu\rfloor + n_T/2+1))/n_T \to \gamma$ with $\gamma \in (0,\,1)$, and $n_T\omega_l^2 \to 0$ as $T\to\infty$, then $\mathbb{E}\left(I_{\mathrm{L},T}\left(u,\,\omega\right)\right) \to \infty$ for many values in the sequence $\{\omega_l\}$ as $\omega_l \to 0$; (b) if $(T_j^0 - (\lfloor Tu\rfloor - n_T/2+1))/n_T \to 0$ or $(T_j^0 - (\lfloor Tu\rfloor + n_T/2+1))/n_T \to 0$, then $\mathbb{E}\left(I_{\mathrm{L},T}\left(u,\,\omega_l\right)\right) \geq f\left(u,\,\omega_l\right)$ as $\omega_l \to 0$.

It is useful to compare Theorem 2.3.4 with Theorem 2.3.2. Unlike the periodogram, the asymptotic behavior of the local periodogram as $\omega_l \to 0$ depends on the vicinity of u to λ_j^0 $(j=1,\ldots,\,m_0)$. Since $I_{\mathrm{L},T}\left(u,\,\omega_l\right)$ uses observations in the window $[\lfloor Tu\rfloor - n_T/2+1,\ldots,\,\lfloor Tu\rfloor + n_T/2]$, if no discontinuity in the mean occurs in this window then $I_{\mathrm{L},T}\left(u,\,\omega_l\right)$ is asymptotically unbiased for the spectral density $f\left(u,\,\omega_l\right)$ and therefore bounded below by it. More complex is its behavior if some T_j^0 falls in the window $[\lfloor Tu\rfloor - n_T/2+1,\ldots,\,\lfloor Tu\rfloor + n_T/2]$. The theorem shows that if T_j^0 is close to the boundary, as indicated in case (ii-b), then $I_{\mathrm{L},T}\left(u,\,\omega_l\right)$ is bounded below by $f\left(u,\,\omega_l\right)$, similarly to case (i). If instead T_j^0 falls sufficiently close to the mid-point $\lfloor Tu\rfloor$, as indicated in case (ii-a), then $\mathbb{E}\left(I_{\mathrm{L},T}\left(u,\,\omega\right)\right) \to \infty$ for many values in the sequence $\{\omega_l\}$ as $\omega_l \to 0$ provided it satisfies $n_T\omega_l^2 \to 0$ as $T\to\infty$. Hence, unless $T\lambda_j^0$ is close to $\lfloor Tu\rfloor$, the local periodogram $I_{\mathrm{L},T}\left(u,\,\omega_l\right)$ behaves very differently from the periodogram $I_T\left(\omega_l\right)$. Accordingly, nonstationarity is unlikely to generate long memory effects if one uses the local periodogram. Further, if one uses preliminary inference procedures for the detection and estimation of the discontinuities in the spectrum and for the estimation of their locations, then one can construct the window efficiently and avoid T_j^0 being too close to $\lfloor Tu\rfloor$. Such procedures have been proposed recently in Casini and Perron (2020a).

2.4 Consequences for HAR Inference

In this section we discuss the implications of the theoretical results from Section 2.3 for inference in the context of potentially autocorrelated data (i.e., HAR inference). We separate the discussion into two parts. We first discuss HAR inference tests for which the issues of low frequency contamination arise as a finite-sample problem. Then we discuss HAR inference tests for which the results presented in the previous section apply even asymptotically. We begin with a review of HAR inference methods and their connection to the estimates considered above.

2.4.1 HAR Inference Methods

There are two main approaches for HAR inference which differ on whether the long-run variance estimator is consistent or not. The classical approach relies on consistency which results in HAC standard errors [cf. Newey and West (1987, 1994) and Andrews (1991)]. Classical HAC standard errors require estimation of the long-run variance defined as $J \triangleq \lim_{T\to\infty} J_T$ where $J_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}(V_s V_t')$. The form of $\{V_t\}$ depends on the specific problem under study. For example, for a t-test on a regression coefficient in the linear model $y_t = x_t\beta_0 + e_t$ $(t = 1, \ldots, T)$ we have $V_t = x_t e_t$. Classical HAC estimators take the following form,

$$\widehat{J}_{\text{Cla},T} \triangleq \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right)\widehat{\Gamma}\left(k\right),$$

where $\widehat{\Gamma}\left(k\right)$ is given in (2.10) with $\widehat{V}_t = x_t\widehat{e}_t$ where $\{\widehat{e}_t\}$ are the least-squares residuals, $K_1\left(\cdot\right)$ is a real-valued kernel and $b_{1,T}$ is bandwidth parameter. One can use the the Bartlett kernel, advocated by Newey and West (1987), or the quadratic spectral kernel as suggested by Andrews (1991), or any other kernel suggested in the literature, see e.g. Ng (1996). Under $b_{1,T} \to 0$ at an appropriate rate, we have $\widehat{J}_{\text{Cla},T} \xrightarrow{\mathbb{P}} J$.

Hence, equipped with $\widehat{J}_{\text{Cla},T}$, HAR inference is standard because HAR tests follow asymptotically standard distributions. This is the simplest approach.

It was shown that classical HAC standard errors can result in oversized tests when there is substantial temporal dependence [e.g., Andrews (1991)]. This stimulated a second approach based on inconsistent long-run variance estimators that keep the bandwidth at some fixed fraction of the sample size [cf. Kiefer et al. (2000)], e.g., using all autocovariances, so that $\widehat{J}_{\text{KVB},T} \triangleq T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(1-|t-s|/T\right)V_tV_s$ which is equivalent to the Newey-West estimator with $b_{1,T}=T^{-1}$, in which case $\widehat{J}_{\text{KVB},T}$ is an inconsistent estimate of J. $\widehat{J}_{\text{KVB},T}$ is essentially a weighted average of the periodogram ordinates with weights that do not spread out as $T\to\infty$. Because of the inconsistency, inference is nonstandard and HAR tests do not have asymptotically standard distributions. The validity of fixed-b HAR inference rests on stationarity. Many authors have considered modifications of $\widehat{J}_{\text{KVB},T}$. However, the one that leads to HAR inference tests that are least oversized is the original $\widehat{J}_{\text{KVB},T}$ [see Casini and Perron (2020e)]. Here for comparison we also report the equally weighted cosine (EWC) estimator of Lazarus et al. (2017). This is an orthogonal series estimators that use large-bandwidths,

$$\widehat{J}_{\text{EWC},T} \triangleq B^{-1}\sum_{j=1}^{B}\Lambda_j^2, \qquad \text{where} \quad \Lambda_j=\sqrt{\frac{2}{T}}\sum_{t=1}^{T}V_t\cos\left(\pi j\left(\frac{t-1/2}{T}\right)\right)$$

with B some fixed integer. Assuming B satisfies some conditions, under fixed-b asymptotics a t-test normalized by $\widehat{J}_{\text{EWC},T}$ follows a t_B distribution where B is the degree of freedom.

Recently, a new HAC estimator was proposed in Casini (2019). Motivated by the signs of low frequency contamination of existing long-run variance estimators, he

proposed a double kernel HAC (DK-HAC) estimator. This is defined as

$$\widehat{J}_{\mathrm{DK},T} \triangleq \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right)\widehat{\Gamma}_{\mathrm{DK}}\left(k\right),$$

where $b_{1,T}$ is a bandwidth sequence and $\widehat{\Gamma}_{\mathrm{DK}}\left(k\right)$ defined in Section 2.3 with $\widehat{c}_T\left(\cdot,\, k\right)$ replaced by

$$\widehat{c}_{\mathrm{DK},T}\left(rn_T/T,\, k\right) = \left(Tb_{2,T}\right)^{-1} \sum_{s=|k|+1}^{T} K_2\left(\frac{\left(rn_T - \left(s - |k|/2\right)\right)/T}{b_{2,T}}\right) V_s V_{s-|k|},$$

with K_2 a real-valued kernel and $b_{2,T}$ a bandwidth sequence. Note that $\widehat{c}_{\mathrm{DK},T}$ and $\widehat{c}_T$ are asymptotically equivalent and the results of Section 2.3 continue to hold for $\widehat{c}_{\mathrm{DK},T}$. More precisely, $\widehat{c}_T$ is a special case of $\widehat{c}_{\mathrm{DK},T}$ with K_2 being a rectangular kernel and $n_{2,T} = Tb_{2,T}$. This approach falls in the first category of standard inference $\widehat{J}_{\mathrm{DK},T} \xrightarrow{\mathbb{P}} J$ and HAR tests normalized by $\widehat{J}_{\mathrm{DK},T}$ follows standard distribution asymptotically. Additionally, Casini and Perron (2020e) proposed prewhitened DK-HAC ($\widehat{J}_{\mathrm{pw,DK},T}$) estimator that improves the size control of HAR tests normalized by $\widehat{J}_{\mathrm{DK},T}$. The estimator $\widehat{J}_{\mathrm{pw,DK},T}$ applies a prewhitening transformation to the data before constructing $\widehat{J}_{\mathrm{DK},T}$ and enjoys the same asymptotic properties as the non-prewhitened DK-HAC estimator $\widehat{J}_{\mathrm{DK},T}$. Due to their ability to more flexibly account for nonstationarity, Casini (2019) and Casini and Perron (2020e) demonstrate that $\widehat{J}_{\mathrm{DK},T}$ and $\widehat{J}_{\mathrm{pw,DK},T}$ have superior power properties relative to the other estimators mentioned above. In terms of size, $\widehat{J}_{\mathrm{pw,DK},T}$ performs better than $\widehat{J}_{\mathrm{Cla},T}$ and $\widehat{J}_{\mathrm{DK},T}$, and is competitive with $\widehat{J}_{\mathrm{KVB},T}$ when the latter works well.[4] We include $\widehat{J}_{\mathrm{DK},T}$ and $\widehat{J}_{\mathrm{pw,DK},T}$ in our simulations below. We report the results only for the DK-HAC estimators that do not use the pre-test for discontinuities in the spectrum [cf. Casini and

[4]There are some empirical cases where $\widehat{J}_{\mathrm{KVB},T}$ does not work well in terms of both size and power. In those cases our method works well. For the cases where $\widehat{J}_{\mathrm{KVB},T}$ leads to tests that have good size, our method is competitive.

Perron (2020a)] because we do not want the results to be affected by the pre-test. Since the pre-test improves the results, what we report here are the worst-case results for the DK-HAC estimators.

2.4.2 Small-Sample Low Frequency Contamination

We now discuss situations in which the low frequency contamination arises as a small-sample problem. These comprises situations where $d^* \approx 0$ asymptotically but a consistent estimate of d^* satisfies $\widehat{d}^* > 0$ in finite-sample. In these situations, no bias due to long memory effects should occur asymptotically but can have an effect with a small sample-size. We begin with a simple model which involves a zero-mean SLS process with changes in persistence. The combination of nonstationarity and serial dependence generates long memory effects because $\widehat{d}^* > 0$.

We specify $\{V_t\}$ as a SLS process given by a two-regimes zero-mean time-varying AR(1), labeled model M1. That is, $V_t = 0.9V_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$ for $t = 1, \ldots, T_1^0$ with $T_1^0 = T\lambda_1^0$, and $V_t = \rho(t/T) V_{t-1} + u_t$, $\rho(t/T) = 0.3(\cos(1.5 - \cos(t/T)))$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 0.5)$ for $t = T_1^0 + 1, \ldots, T$. Note that $\rho(\cdot)$ varies between 0.172 and 0.263. We set $\lambda_1^0 = 0.1$ and $T = 200$. A plot of $\{V_t\}$ is reported in Figure 2·1. Note that $\mathbb{E}(V_t) = 0$ for all t and so $d^* = d^*_{\text{Sta}} = 0$. However, if we replace $\overline{\mu}_1$ and $\overline{\mu}_2$ by $\overline{V}_1$ and $\overline{V}_2$, respectively, where $\overline{V}_1 = (T_1^0)^{-1} \sum_{t=1}^{T_1^0} V_t = 1.27$ and $\overline{V}_2 = (T - T_1^0)^{-1} \sum_{t=T_1^0+1}^{T} V_t = -0.03$, then our estimate $\widehat{d}^*$ of d^* would be different from zero and can generate a finite-sample bias which can give rise to effects akin to long memory.

We first look at the behavior of the sample autocovariance $\widehat{\Gamma}(k)$. We compare it with the theoretical autocovariance $\Gamma_T(k) = T^{-1} \sum_{t=k+1}^{T} \mathbb{E}(V_t V_{t-k})$. The latter is equal to $\Gamma_T(k) \approx \lambda_1^0 0.9^k / (1 - 0.81) + \int_{\lambda_1^0}^{1} c(u, k)\, du$. We can compute $\Gamma_T(k)$ numer-

ically using,

$$c(u,\,k) = \int_{-\pi}^{\pi} e^{ik\omega} f(u,\,\omega)\, d\omega$$
$$= \int_{-\pi}^{\pi} e^{ik\omega} \frac{0.5}{2\pi} \left(1 + \rho(u)^2 - 2\rho(u)\cos(\omega)\right)^{-1} d\omega.$$

Table 2.1 reports $\Gamma_T(k)$, $\widehat{\Gamma}(k)$, $\widehat{\Gamma}(k) - \widehat{d}^*$ and $\widehat{\Gamma}_{\mathrm{DK}}(k)$ for several values of k. It is known that the autocovariance estimates are quite noisy in general. However, it is still possible to discern some patterns. For all k, $\widehat{\Gamma}(k)$ largely overestimates $\Gamma_T(k)$. This is consistent with Theorem 2.3.1 which suggests that this is due to the bias $d^* > 0$. This is also supported by the bias-corrected estimate $\widehat{\Gamma}(k) - \widehat{d}^*$ which is accurate in approximating $\Gamma_T(k)$. This is especially so for small k. In general, Theorem 2.3.1 provides excellent approximations confirming that $\widehat{\Gamma}(k)$ suffers from low frequency contaminations. Theorem 2.3.3 suggests that this issue should not occur for $\widehat{\Gamma}_{\mathrm{DK}}(k)$. In fact, $\widehat{\Gamma}_{\mathrm{DK}}(k)$ is more accurate than $\widehat{\Gamma}(k)$ (except for $k = 0$). For $k \geq 20$ the form of $\Gamma_T(k)$ is different, because $T_1^0 = 20$, and is simply given by the autocovariance of V_t for $t \geq 21$ (i.e., the second regime). Thus, $\Gamma_T(k) \approx 0$ for $k \geq 20$ whereas $\widehat{\Gamma}(k)$ is often small but positive. In contrast, $\widehat{\Gamma}_{\mathrm{DK}}(k) \approx 0$ for $k \geq 20$ thereby confirming that $\widehat{\Gamma}_{\mathrm{DK}}(k)$ does not suffer from low frequency contamination. These results are confirmed in Figure 2·2 which plots the autocorrelation function (ACF) of $V_{1,t} = V_t$ $(t = 1, \ldots, 20)$, $V_{2,t} = V_t$ $(t = 21, \ldots, 200)$ and V_t $(t = 1, \ldots, 200)$. Although the ACF of V_t should be a weighted average of the ACF of $V_{1,t}$ and of $V_{2,t}$, the ACF of V_t in the bottom panel shows much higher persistence than either ACF in the top $(V_{1,t})$ and mid $(V_{2,t})$ panels. This is odd since $V_{1,t}$ is a highly persistent series. Further, it shows that the dependence is essentially always positive. This is also odd. These features are consistent with our theory which suggests that nonstationarity makes V_t appear more persistent and that the bias is positive. Other examples involve V_t given

by the least-squares regression residuals under mild forms of misspecification that do not undermine the conditions for consistency of the least-squares estimator. For example, exclusion of a relevant regressor uncorrelated with the included reggressors, or inclusion of an irrelevant regressor. Another example involves V_t obtained after applying some de-trending techniques where the fitted model is not correctly specified (e.g., the data follow a nonlinear trend but one removes a linear trend). A final example is the case of outliers because outliers influence the mean of V_t and therefore d^*.

What is especially relevant is whether this evidence of long memory feature has any consequence for HAR inference. We obtain the empirical size and power for a t-test on the intercept normalized by several LRV estimators for the model $y_t = \delta + V_t$ with $\delta = 0$ under the null and $\delta > 0$ under the alternative hypothesis. In addition to model M1, we consider other models: M2 involves a locally stationary $V_t = \rho(t/T) V_{t-1} + u_t$, $\rho(t/T) = 0.7(\cos(1.5t/T))$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 0.5)$; M3 is the same as M2 with outliers $V_t \sim$ Uniform $(\underline{c}, 10\underline{c})$ for $t = T/4, T/2, 3T/4$ where $\underline{c} = -1/(\sqrt{2}\text{erfc}^{-1}(3/2))\text{med}(|V - \text{med}(V)|)$ where erfc^{-1} is the inverse complementary error function, $\text{med}(\cdot)$ is the median and $V = (V_t)_{t=1}^T$;[5] M4 involves a locally stationary model with periods of strong persistence where $V_t = \rho(t/T) V_{t-1} + u_t$, $\rho(t/T) = 0.95(\cos(1.5t/T))$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 0.4)$. Note that $\rho(\cdot)$ varies between 0.7 and 0.05 in M2-M3 and between 0.95 and 0.07 in M4.

We consider the DK-HAC estimators with and without prewhitening ($\widehat{J}_{\text{DK},T}$, $\widehat{J}_{\text{DK,pw,SLS},T}$, $\widehat{J}_{\text{DK,pw,SLS},\mu,T}$) of Casini (2019) and Casini and Perron (2020e), respectively; Andrews's (1991) HAC estimator with and without the prewhitening procedure of Andrews and Monahan (1992); Newey and West's (1987) HAC estimator with the usual rule to select the number of lags (i.e., $b_{1,T} = 1/(0.75T^{1/3})$; Newey-West with the

[5] We follow the literature on outlier detection for continuous functions and use the median absolute deviation to generate the outlier. This notion used in this literature does not deem a value smaller than $\underline{c}$ as an outlier.

fixed-b method of Kiefer et al. (2000) (labeled KVB); and the Empirical Weighted Cosine (EWC) of Lazarus et al. (2018). For the DK-HAC estimators we use the data-dependent methods for the bandwidths, kernels and choice of n_T as proposed in Casini (2019) and Casini and Perron (2020e), which are optimal under mean-squared error (MSE).[6] Let $\widehat{V}_t$ denote the least-squares residual based on $\widehat{\delta}$ where the latter is the least-squares estimate of δ. We set $\widehat{b}_{1,T} = 0.6828(\widehat{\phi}\,(2)\,T\widehat{\overline{b}}_{2,T})^{-1/5}$ where

$$\widehat{\phi}\,(2) = \left(18 \left(\frac{n_T}{T} \sum_{j=0}^{\lfloor T/n_{3,T} \rfloor - 1} \frac{(\widehat{\sigma}\,((jn_T+1)/T)\,\widehat{a}_1\,((jn_T+1)/T))^2}{(1-\widehat{a}_1\,((jn_T+1)/T))^4} \right)^2 \right) / \left(\frac{n_T}{T} \sum_{j=0}^{\lfloor T/n_{3,T} \rfloor - 1} \frac{(\widehat{\sigma}\,((jn_T+1)/T))^2}{(1-\widehat{a}_1\,((jn_T+1)/T))^2} \right)^2,$$

with

$$\widehat{a}_1\,(u) = \frac{\sum_{j=t-n_T+1}^{t} \widehat{V}_j \widehat{V}_{j-1}}{\sum_{j=t-n_T+1}^{t} (\widehat{V}_{j-1})^2}, \qquad \widehat{\sigma}\,(u) = \left(\sum_{j=t-n_T+1}^{t} (\widehat{V}_j - \widehat{a}_1\,(u)\,\widehat{V}_{j-1})^2 \right)^{1/2};$$

and

$$\widehat{\overline{b}}_{2,T} = (n_T/T) \sum_{r=1}^{\lfloor T/n_T \rfloor - 1} \widehat{b}_{2,T}\,(rn_T/T)\,, \quad \widehat{b}_{2,T}\,(u) = 1.6786(\widehat{D}_1\,(u))^{-1/5}(\widehat{D}_2\,(u))^{1/5}T^{-1/5}$$

where $\widehat{D}_2\,(u) \triangleq 2\sum_{l=-\lfloor T^{4/25} \rfloor}^{\lfloor T^{4/25} \rfloor} \widehat{c}_{\mathrm{DK},T}\,(u,\,l)^2$ and

$$\widehat{D}_1\,(u) \triangleq$$
$$([S_\omega]^{-1} \sum_{s \in S_\omega} [\frac{3}{\pi}(1 + 0.8(\cos 1.5 + \cos 4\pi u)\exp(-i\omega_s))^{-4}(0.8(-4\pi \sin(4\pi u)))\exp(-i\omega_s)$$
$$- \frac{1}{\pi}\,|1 + 0.8(\cos 1.5 + \cos 4\pi u)\exp(-i\omega_s)|^{-3}\,(0.8(-16\pi^2 \cos(4\pi u)))\exp(-i\omega_s)])^2,$$

with $[S_\omega]$ being the cardinality of S_ω and $\omega_{s+1} > \omega_s$, $\omega_1 = -\pi$, $\omega_{[S_\omega]} = \pi$. We set

[6] See Belotti et al. (2021) for an alternative data-dependent method and for some comparisons.

$n_T = T^{0.6}$, $S_\omega = \{-\pi, -3, -2, -1, 0, 1, 2, 3, \pi\}$. $K_1(\cdot)$ is the QS kernel and $K_2(x) = 6x(1-x)$ for $x \in [0, 1]$.

Table 2.2-2.5 report the results. The t-test normalized by Newey and West's (1987) and Andrews's (1991) prewhitened HAC estimators are excessively oversized.[7] Andrews's (1991) HAC estimator is slightly undersized while KVB's fixed-b and EWC are severely undersized.[8] These outcomes arise from nonstationarity and is consistent with our theoretical analysis of $\widehat{\Gamma}(k)$. Since $\widehat{\Gamma}(k) > 0$ for many large lags k, the KVB's fixed-b and EWC's estimators that include many lags (i.e., long bandwidths) are inflated and reduce the magnitude of the test statistic even under the null hypotheses. For the t-test on the intercept, $\widehat{J}_{\mathrm{DK},T}$ can be oversized when there is strong dependence, as shown in Table 2.2. However, the prewhitened DK-HAC estimators $\widehat{J}_{\mathrm{DK,pw,SLS},T}$ and $\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$ show very accurate rejection rates. Overall, inspection of the size properties suggests that the DK-HAC estimators do not suffer from low frequency contamination which, in contrast, affects the LRV estimators that rely on the full sample estimates $\widehat{\Gamma}(k)$ or equivalently on $I_T(\omega)$ (e.g., the EWC). The same issue also affects the power properties of the tests. The KVB's fixed-b and EWC's estimators suffer from relatively large power losses. The power of Newey and West's and Andrews's (1991) prewhitened HAC is not comparable because they are signifi-

[7]This is not in contradiction with our theoretical results. The prewhitening of Andrews and Monahan (1992) is unstable when there is nonstationarity as shown by Casini and Perron (2020e). The reason is that there is a bias both in the whitening and the recoloring stages. The biases have opposite signs so that here the underestimation of the LRV dominates. Newey-West uses a fixed rule for determining the number of lags. The number of included lags is small. This estimator is known to be largely oversized when the data are stationary with high dependence. Our results say that the included sample autocovariances may be inflated if there is nonstationarity. However, given that the fixed rule selects a small number of lags then nonstationarity results in a smaller oversize problem.

[8]In general, Andrews's (1991) HAC estimator leads to tests that are oversized when the data are stationary with strong dependence. Here nonstationarity reduces the oversize problem. It follows a similar argument as for the Newey-West estimator even though Andrews's (1991) HAC estimator uses a data-dependent method for the selection of the number of lags. Thus, it selects more lags than the ones suggested by the fixed rule. Consequently, more sample autocovariances are overestimated and this helps to reduce its oversize problem.

cantly oversized. The DK-HAC estimators have the best power, the second best being Andrews's (1991) HAC estimator.

Turning to M2, Table 2.3 shows even larger size distortions and power losses for KVB's fixed-b and EWC's estimators. All the DK-HAC estimators display accurate size control and good power. Newey and West's and Andrews's (1991) prewhitened HAC estimators are again excessively oversized. Andrews's (1991) HAC estimator sacrifices some power relative to the DK-HAC even though the margin is not high. For model M3-M4, Table 2.4-2.5 show that Andrews's (1991) HAC estimator also suffers from strong size distortions and power losses, thus sharing the same issues as KVB's fixed-b and EWC's estimators. In model M4, even Newey and West's and Andrews's (1991) prewhitened HAC estimators are undersized and have relatively low power. The DK-HAC estimators perform best both in terms of size and power. Table 2.2-2.5 suggest that the low frequency contamination can equally arise from different forms of nonstationarity. Overall, the evidence for quite substantial underejection and power losses in model M1-M4 for the existing LRV estimators is consistent with our theoretical results. These represent situations where the contamination occurs as a small-sample problem. In the next section, we show that when the contamination holds asymptotically then the size distortions and power problems can be even more severe.

2.4.3 General Low Frequency Contamination

We now discuss statistical environments and HAR inference tests for which the low frequency contamination results of Section 2.3 hold even asymptotically. This means that $d^* > 0$ for all T and as $T \to \infty$. This comprises the class of HAR tests that admit a nonstationary alternative hypotheses. This class is very large and include most HAR-based tests. Examples include tests for forecast evaluation [e.g., Casini (2018), Diebold and Mariano (1995), Giacomini and Rossi (2009, 2010), Giacomini

and White (2006), Perron and Yamamoto (2019) and West (1996)], tests for structural changes [e.g., Andrews (1993), Bai and Perron (1998), Casini and Perron (2020c, 2020d, 2020b), Elliott and Müller (2007), and Qu and Perron (2007)], tests for time-varying parameters [e.g., Cai (2007) and Chen and Hong (2012)], tests for regime switching [e.g., Hamilton (1989) and Qu and Zhuo (2021)] and many others.[9] Here we consider the Diebold-Mariano test for the sake of illustration and remark that similar issues apply to the other HAR tests mentioned above.

The Diebold-Mariano test statistic is defined as $t_{\mathrm{DM}} \triangleq T_n^{1/2}\overline{d}_L/\sqrt{\widehat{J}_{d_L,T}}$, where $\overline{d}_L$ is the average of the loss differentials between two competing forecast models, $\widehat{J}_{d_L,T}$ is an estimate of the LRV of the the loss differential series and T_n is the number of observations in the out-of-sample. We use the quadratic loss. We consider an out-of-sample forecasting exercise with a fixed forecasting scheme where, given a sample of T observations, $0.5T$ observations are used for the in-sample and the remaining half is used for prediction [see Perron and Yamamoto (2019) for recommendations on using a fixed scheme in the presence of breaks]. The DGP under the null hypotheses is given by $y_t = 1 + \beta_0 x_{t-1}^{(0)} + e_t$ where $x_{t-1}^{(0)} \sim$ i.i.d. $\mathscr{N}(1, 1)$, $e_t = 0.3e_{t-1} + u_t$ with $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$, and we set $\beta_0 = 1$ and $T = 400$. The two competing models both involve an intercept but differ on the predictor used in place of $x_t^{(0)}$. The first forecast model uses $x_t^{(1)}$ while the second uses $x_t^{(2)}$ where $x_t^{(1)}$ and $x_t^{(2)}$ are independent i.i.d. $\mathscr{N}(1, 1)$ sequences, both independent from $x_t^{(0)}$.

[9] In some cases the low frequency contamination could be reduced if one uses a test that accounts properly for the form of nonstationary under the alternative hypotheses. For example, consider testing for breaks. Suppose that there are two breaks and one first uses a test for one break versus no break and then a test of two breaks versus one break if the first test rejects the null hypotheses of no break, and so on. The test in the first step would suffer from low frequency contamination since under the alternative of one break the estimate of the LRV is contaminated by the presence of the second break. Thus, one would conclude that there is no break. In contrast, if one uses a test that allows for an unknown number of breaks given some upper bound [i.e., the UDmax test of Bai and Perron (1998)], the contamination can be reduced because this test would account for the correct form of nonstationarity. However, if instead of two breaks the true model involves other forms of nonstationarity (e.g., slowly-varying parameters, smooth breaks, etc.) then also the UDmax test would suffer from low frequency contamination.

Each forecast model generates a sequence of $\tau\,(=1)$-step ahead out-of-sample losses $L_t^{(j)}$ $(j=1,\,2)$ for $t=T/2+1,\ldots,\,T-\tau$. Then $d_t \triangleq L_t^{(2)}-L_t^{(1)}$ denotes the loss differential at time t. The Diebold-Mariano test rejects the null hypotheses of equal predictive ability when $\overline{d}_L$ is sufficiently far from zero. Under the alternative hypothesis, the two competing forecast models are as follows: the first uses $x_t^{(1)}=x_t^{(0)}+u_{X_1,t}$ where $u_{X_1,t}\sim$ i.i.d.$\mathscr{N}\,(0,\,1)$ while the second uses $x_t^{(2)}=x_t^{(0)}+0.2z_t+2u_{X_2,t}$ for $t\in[1,\ldots,\,3T/4-1,\,3T/4+21,\ldots T]$ and $x_t^{(2)}=\delta\,(t/T)+0.2z_t+2u_{X_2,t}$ for $t=3T/4,\ldots,\,3T/4+20$ with $u_{X_2,t}\sim$ i.i.d.$\mathscr{N}\,(0,\,1)$, where z_t has the same distribution as $x_t^{(0)}$.

We consider four specifications for $\delta\,(\cdot)$. In the first $x_t^{(2)}$ is subject to an abrupt break in the mean $\delta\,(t/T)=\delta>0$, in the second $x_t^{(2)}$ is locally stationary with time-varying mean $\delta\,(t/T)=\delta\,(\sin\,(t/T-3/4))$, in the third specification $x_t^{(2)}=x_t^{(0)}+0.2z_t+2u_{X_2,t}$ for $t\in[1,\ldots,\,T/2-30,\,T/2+21,\ldots T]$ and $x_t^{(2)}=\delta\,(t/T)+0.2z_t+2u_{X_2,t}$ for $t=T/2-30,\ldots,\,T/2+20$ with $\delta\,(t/T)=\delta(\sin(t/T-1/2\,-30/T))$, in the fourth $x_t^{(2)}$ is the same as in the second with in addition two outliers $x_t^{(2)}\sim$ Uniform $(|\underline{c}|\,,\,5\,|\underline{c}|)$ for $t=6T/10,\,8T/10$ where $\underline{c}=-1/\left(\sqrt{2}\mathrm{erfc}^{-1}\,(3/2)\right)\mathrm{med}(|x^{(2)}-\mathrm{med}\,\left(x^{(2)}\right)|)$ where $x^{(2)}=\left(x_t^{(2)}\right)_{t=1}^{T}$. That is, in the second model $x_t^{(2)}$ is locally stationary only in the out-of-sample, in the third it is locally stationary in both the in-sample and out-of sample and in the fourth model $x_t^{(2)}$ has two outliers in the out-of-sample. The location of the outliers is irrelevant for the results; they can also occur in the in-sample.

Table 2.6 reports the size and the power for all models. We begin with the case $\delta\,(t/T)=\delta>0$ (top panel). The size of the test using the DK-HAC estimators is accurate while the test using other LRV estimators are oversized with the exception of the KVB's fixed-b method for which the rejection rate is equal to zero. The HAR tests using existing LRV estimators have lower power relative to that obtained with the DK-HAC estimators for small values of δ. When δ increases the tests standardized

by the HAC estimators of Andrews (1991) and Newey and West (1987), and by the KVB's fixed-b and EWC LRV estimators display non-monotonic power gradually converging to zero. In contrast, when using the DK-HAC estimators the test has monotonic power that reaches and maintains unit power. The results for the other models are even stronger. In general, except for the DK-HAC estimators, all other tests display serious power problems. Thus, either form of nonstationarity or outliers leads to similar implications, consistent with our theoretical results.

In order to further assess the theoretical results from Section 2.3, Figure 2·3-2·5 report the plots of d_t, its sample autocovariances and its periodogram, for $\delta =$ 1, 2, 5, respectively. We only consider the case $\delta_t = \delta > 0$. The other cases lead to the same conclusions. For $\delta = 1$, Figure 2·3 (mid panel) shows that $\widehat{\Gamma}(k)$ decays slowly. As δ increases, Figure 2·4-2·5 (mid panels), $\widehat{\Gamma}(k)$ decays even more slowly at a rate far from the typical exponential decay of short memory processes. It shows a power-like decay that is statistically significant at the 10% significance level up to lag 80. This suggests evidence of long memory. However, the data are short memory with small temporal dependence. What is generating the spurious long memory effect is the nonstationarity present under the alternative hypotheses. This is visible in the top panels which present plots of d_t for the first specification. The shift in the mean of d_t for $t = 3T/4, \ldots, 3T/4+20$ is responsible for the long memory effect. This corresponds to the second term of (2.12) in Theorem 2.3.1. The negative autocovariances follow from the behavior of d_t for observations $t \neq 3T/4, \ldots, 3T/4+$ 20. That is, a positive d_t at time t corresponds to $x_{t-1}^{(2)}$ predicting y_t worse than what $x_{t-1}^{(1)}$ does and this is likely to be followed by $x_t^{(2)}$ predicting y_{t+1} better than what $x_t^{(1)}$ does. Thus, the overall behavior of the sample autocovariance is as predicted by Theorem 2.3.1. For small lags, $\widehat{\Gamma}(k)$ shows a power-like decay and it is positive. As k increases to medium lags, the autocovariances turn negative. However, d^* in (2.12)

makes these autocovariances closer to zero since $d^* > 0$. For large k, $\widehat{\Gamma}(k)$ approaches zero. Next, we move to the bottom panels which plot the periodogram of $\{d_t\}$. It is unbounded at frequencies close to $\omega = 0$ as predicted by Theorem 2.3.2 and as would occur if long memory was present. It also explains why the Diebold-Mariano test normalized by Newey-West's, Andrews', KVB's fixed-b and EWC's LRV estimators have serious power problems. These LRV estimators are inflated and consequently the tests lose power. The figures show that as we raise δ the more severe these issues and the power losses so that the power eventually reaches zero. This is consistent with our theory since d^* is increasing in δ (cf. $d^* \approx 0.1 \cdot 0.9\delta^2$).

We now verify the results about the local sample autocovariance $\widehat{c}_T(u, k)$ and the corresponding local periodogram from Theorem 2.3.3-2.3.4. We set $n_{2,T} = T^{0.6} = 36$ following the MSE criterion of Casini (2019). We consider (i) $u = 236/T$, (ii-a) $u = T_1^0/T = 3/4$ and (ii-b) $u = 264/T$. Note that cases (i)-(ii-b) correspond to parts (i)-(ii-b) in Theorem 2.3.3-2.3.4. We consider $\delta = 1$, 2 and 5. According to Thereon 2.3.3-2.3.4, we should expect long memory features only for case (ii-a). Figure 2·6-2·11 confirm this. The results pertaining to case (ii-a) are plotted in the middle panels. Figures 2·6, 2·8 and 2·10 show that the local autocovariance displays slow decay similar to the pattern discussed above for the sample autocovariance $\widehat{\Gamma}(k)$ and that this problem becomes more severe as δ increases. Such long memory features also appear for $I_{\mathrm{L}}(3/4, \omega)$. The middle panels in Figure 2·7, 2·9 and 2·11 show that the local periodogram at $u = 3/4$ and at a frequency close to $\omega = 0$ are extremely large. The latter result is consistent with Theorem 2.3.4-(ii-a) which suggests that $I_{\mathrm{L},T}(3/4, \omega) \to \infty$ as $\omega \to 0$. For case (i) and (ii-b) both figures show that the local autocovariance and the local periodogram do not display long memory features. Indeed, they have forms similar to those of a short memory process, a result consistent with Theorem 2.3.3-2.3.4 also for cases (i) and (ii-b).

It is interesting to explain why HAR inference based on the DK-HAC estimators does not suffer from the low frequency contamination even if case (ii-a) occurs. Note that the DK-HAC estimator computes an average of the local spectral density over time blocks. If it happens that one of these blocks contains a discontinuity in the spectrum, then as in case (ii-a) some bias would arise for the local spectral density estimate corresponding to that block. However, by virtue of the time-averaging over blocks that bias becomes negligible. Hence, nonparametric smoothing over time asymptotically cancels the bias, so that inference based on the DK-HAC estimators is robust to nonstationarity.

2.4.4 Discussion

In summary, the sample autocovariance and the periodogram are sensitive to nonstationarity in that they may display characteristics typical of long memory even if the data are short memory. We refer to this phenomenon as low frequency contamination induced by unaccounted nonstationarity. In some situations these issues only imply a small-sample problem [cf. Section 2.4.2]. In others, they have an effect even asymptotically. In either case, the theory in Section 2.3 provides useful guidance about the properties of the sample autocovariance and the periodogram in finite-samples. It also provides accurate approximations for misspecified models or models with outliers as discussed above. Since LRV estimates are direct inputs for inference methods in the context of potentially autocorrelated data, our theory offers new insights for the properties of HAR inference tests. The use of HAC standard errors has become the standard practice, and recent theoretical developments in HAR inference advocated the use of fixed-b or long bandwidth LRV estimators on the basis that they offer better size control when there is strong dependence in the data. Our results suggest that care is needed before applying such methods because they are highly sensitive to effects akin to long memory arising from nonstationarity. The concern is

then that the use of long bandwidths leads to overestimation of the true LRV due to low frequency contamination. Consequently, HAR tests can lose power dramatically, a problem that occurs also for the classical HAC estimators, though to a lesser extent since they use a smaller number of sample autocovariances.

Our theory also suggests a solution to this problem. This entails the use of nonparametric smoothing over time which avoids combining observations that belong to different regimes. This accounts for nonstationarity and prevents spurious long memory effects. An exception where some bias may arise for $\widehat{c}_T(u, k)$ and $I_T(u, \omega)$ is when a discontinuity in the spectrum falls close to $\lfloor Tu \rfloor$. However, this problem is simple to address because one can use a pre-test for discontinuities in the spectrum and exclude those $u \in (0, 1)$ that are close to a discontinuity in the spectrum. Casini and Perron (2020a) proposed such a test and used it for the DK-HAC estimators which were shown to be robust to such cases.

2.5 Conclusions

Economic time series are highly nonstationary and models might be misspecified. If nonstationary is not accounted for properly, parameter estimates and, in particular, asymptotic variance estimates can be largely biased. We establish results on the low frequency contamination induced by nonstationarity and misspecification for the sample autocovariance and the periodogram. These estimates can exhibit features akin to long memory when the data are nonstationary short memory. We distinguish cases where this contamination only implies a small-sample problem and cases where the problem remains asymptotically. We propose a solution to this problem based on nonparametric smoothing which is shown, using theoretical arguments, to be robust. Since the autocovariances and the periodogram are basic elements for heteroskedasticity and autocorrelation robust (HAR) inference, our results provide

insights on the important debate between consistent versus inconsistent LRV estimation. Indeed, the properties of long bandwidths/fixed-b methods are only known under stationarity. Our results show that existing LRV estimators tend to be inflated when the data are nonstationarity. This results in HAR tests that can be undersized and exhibit dramatic power losses or even no power. Long bandwidths/fixed-b HAR tests suffer more from low frequency contamination relative to HAR tests based on HAC estimators, whereas the DK-HAC estimators do not suffer from this problem.

2.6 Figures

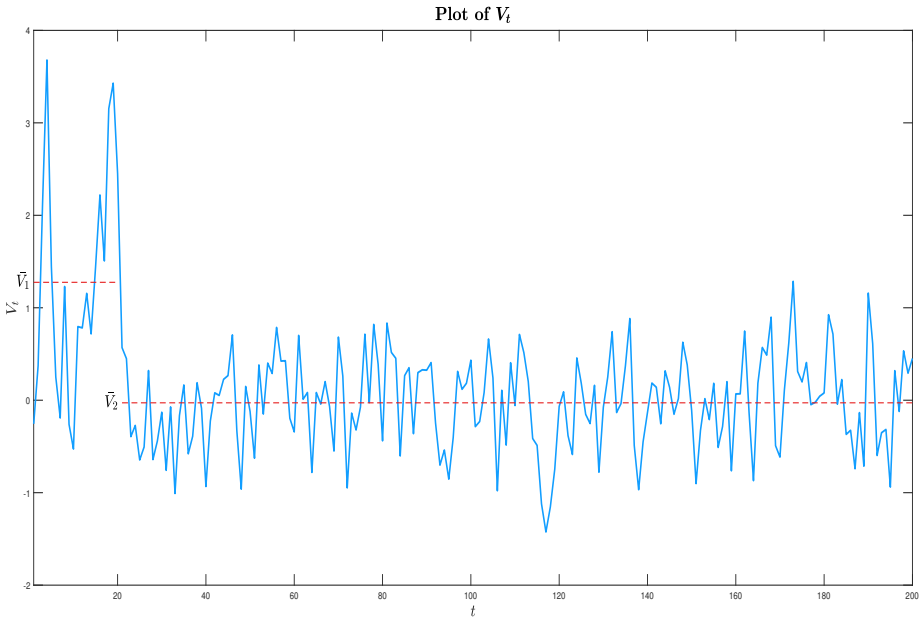


Figure 2·1: Plot of V_t for model M1. The sample size is $T = 200$. $T_1^0 = 20$. Also reported in red dashed lines are the sample averages in the two regimes with $\overline{V}_1 = 1.27$ and $\overline{V}_2 = -0.03$.

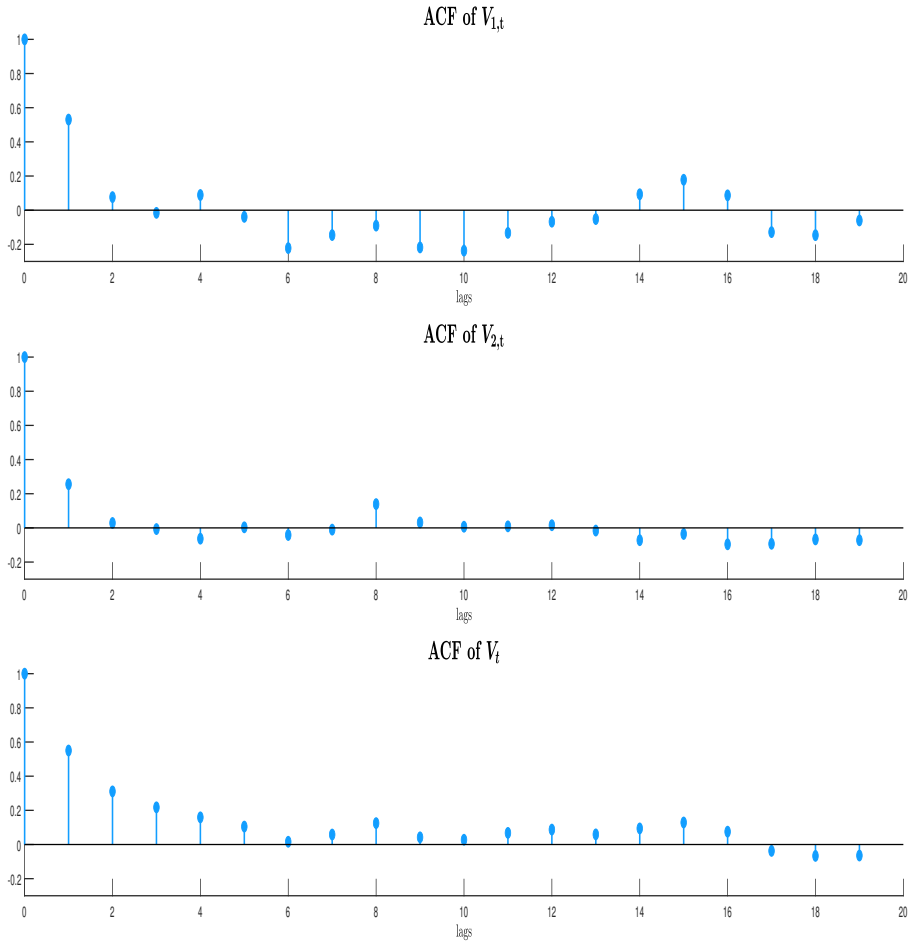


Figure 2·2: ACF of $V_{1,t}$ (top panel), ACF of $V_{2,t}$ (mid panel) and ACF of V_t (bottom panel) for model M1.

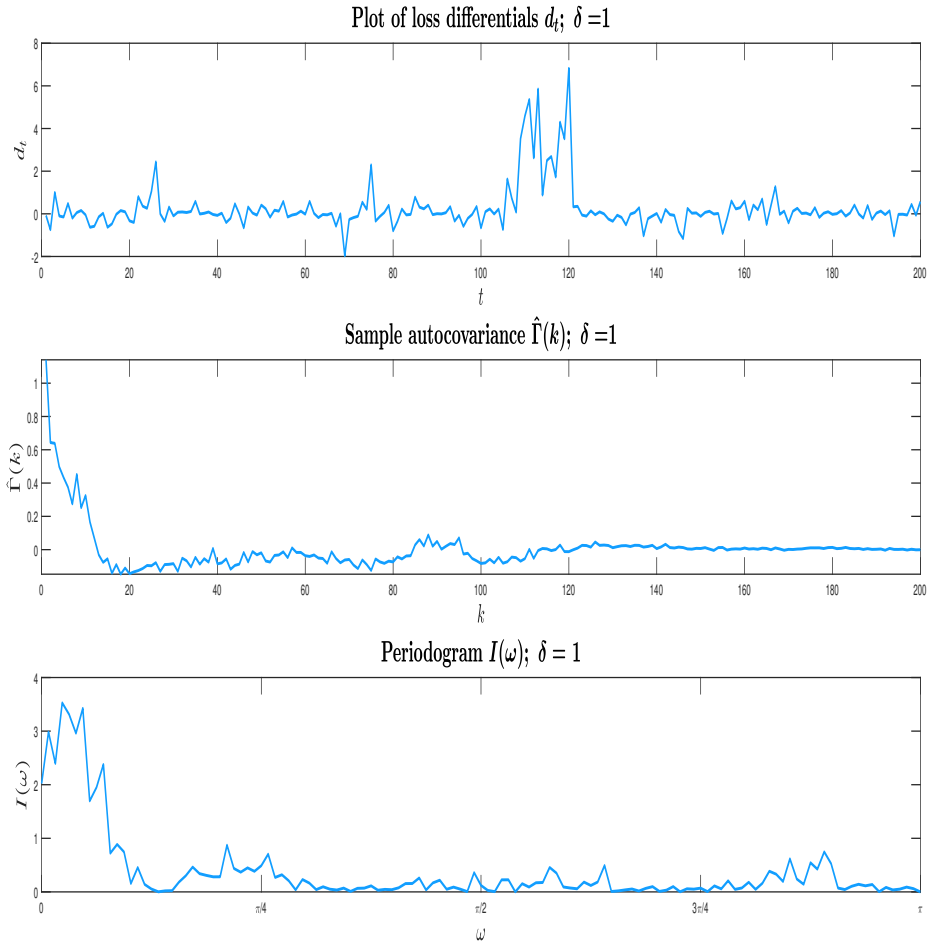


Figure 2·3: a) top panel: plot of $\{d_t\}$; b) mid-panel: plot of the sample autocovariances $\widehat{\Gamma}(k)$ of $\{d_t\}$; c) bottom panel: plot the periodogram $I(\omega)$ of $\{d_t\}$. In all panels $\delta = 1$.

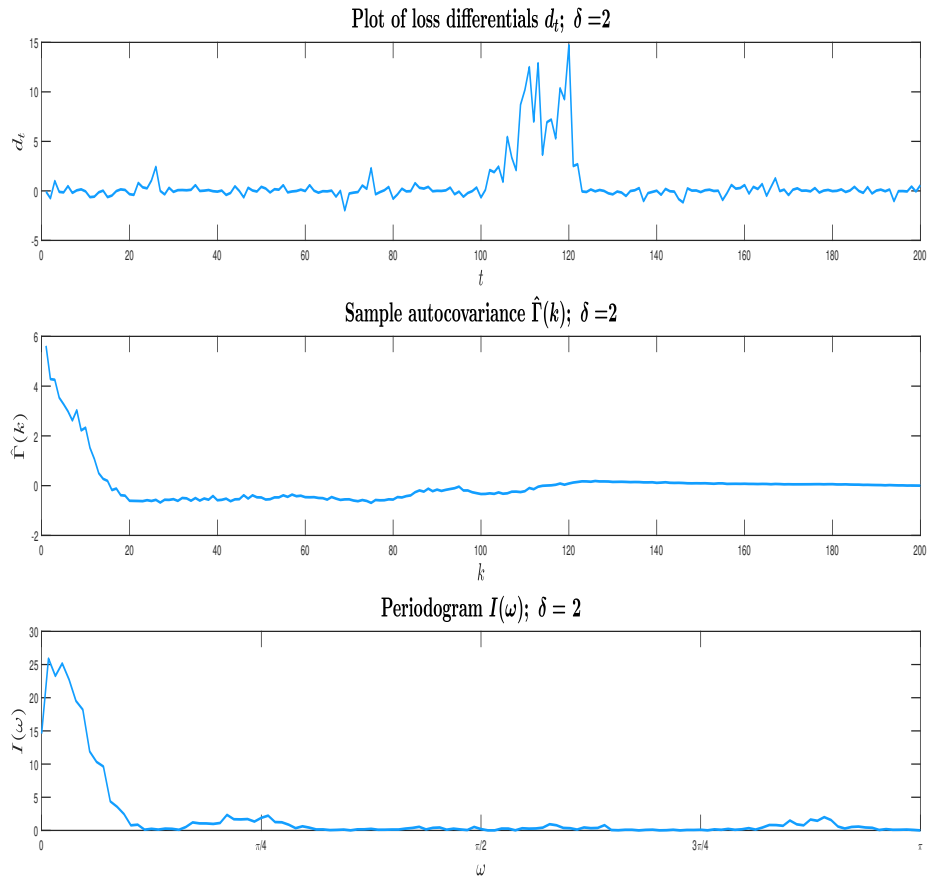


Figure 2·4: a) top panel: plot of $\{d_t\}$; b) mid-panel: plot of the sample autocovariances $\widehat{\Gamma}(k)$ of $\{d_t\}$; c) bottom panel: plot the periodogram $I(\omega)$ of $\{d_t\}$. In all panels $\delta = 2$.

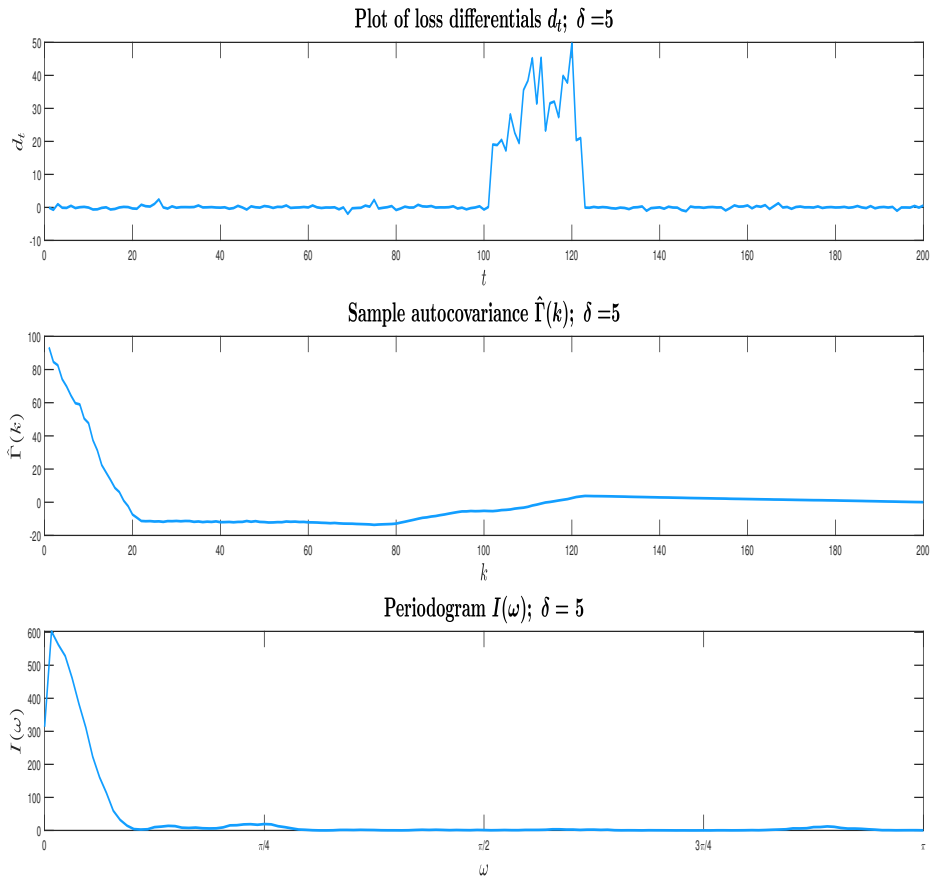


Figure 2·5: a) top panel: plot of $\{d_t\}$; b) mid-panel: plot of the sample autocovariances $\widehat{\Gamma}(k)$ of $\{d_t\}$; c) bottom panel: plot the periodogram $I(\omega)$ of $\{d_t\}$. In all panels $\delta = 5$.

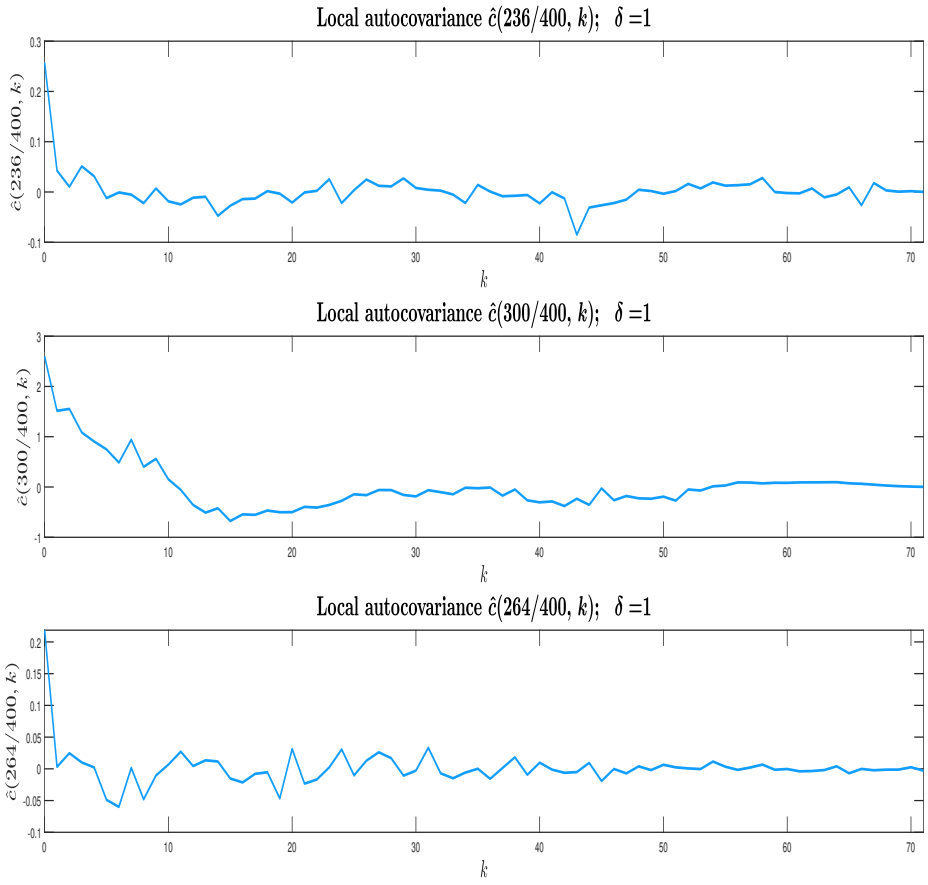


Figure 2·6: The figure plots $\widehat{c}_T(u, k)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 1$.

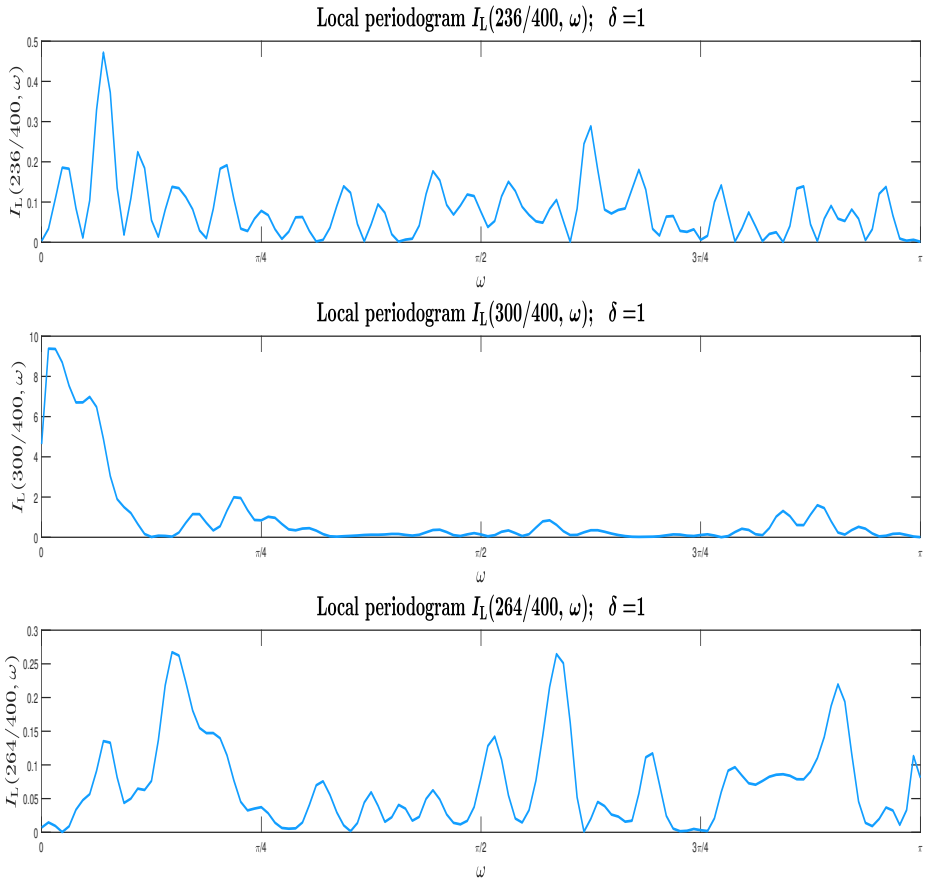


Figure 2·7: The figure plots $I_{\mathrm{L}}(u, \omega)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 1$.

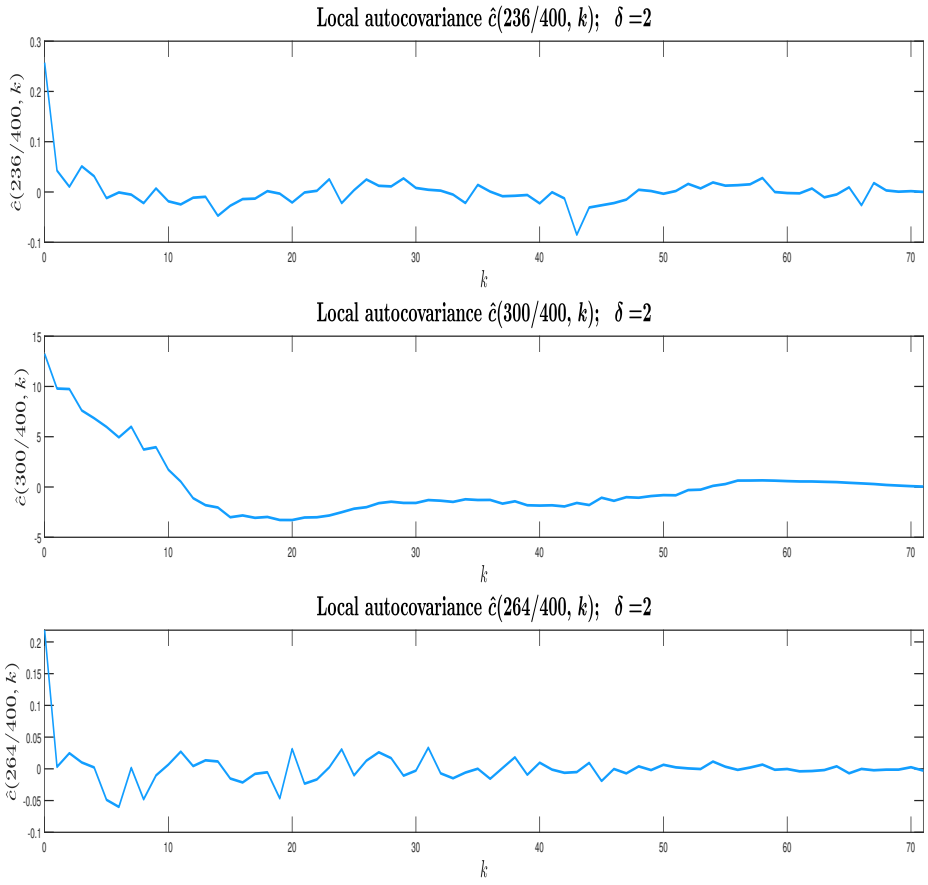


Figure 2·8: The figure plots $\widehat{c}_T(u, k)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 2$.

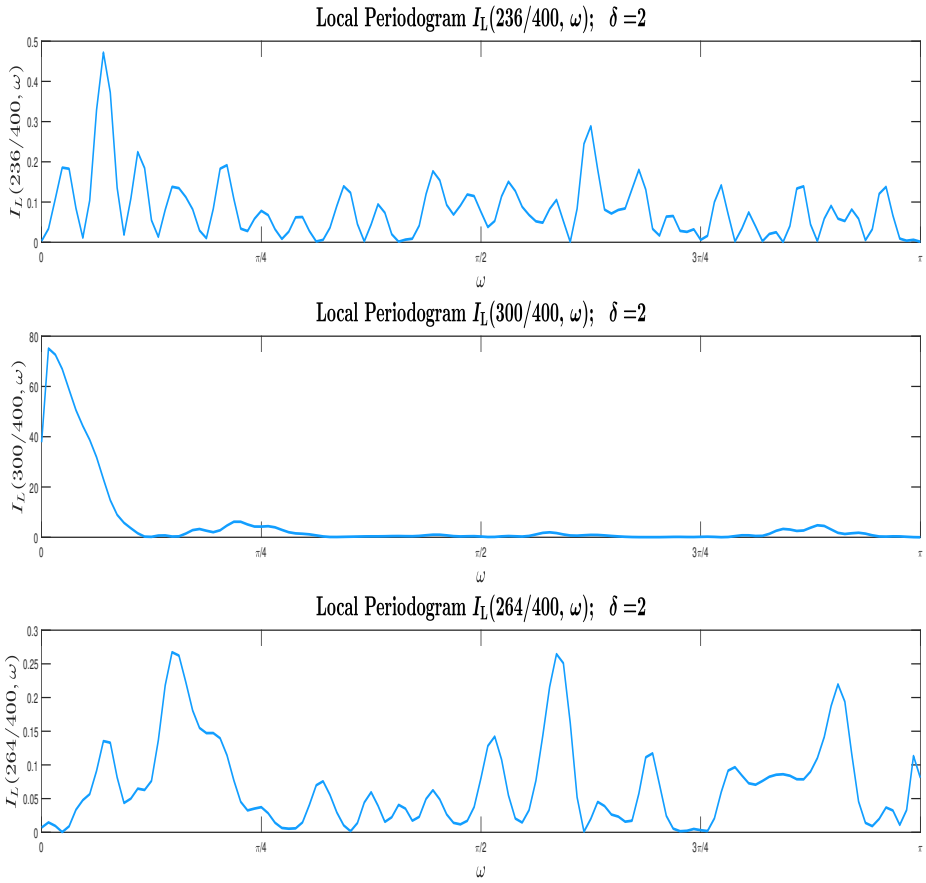


Figure 2·9: The figure plots $I_{\mathrm{L}}(u, \omega)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 2$.

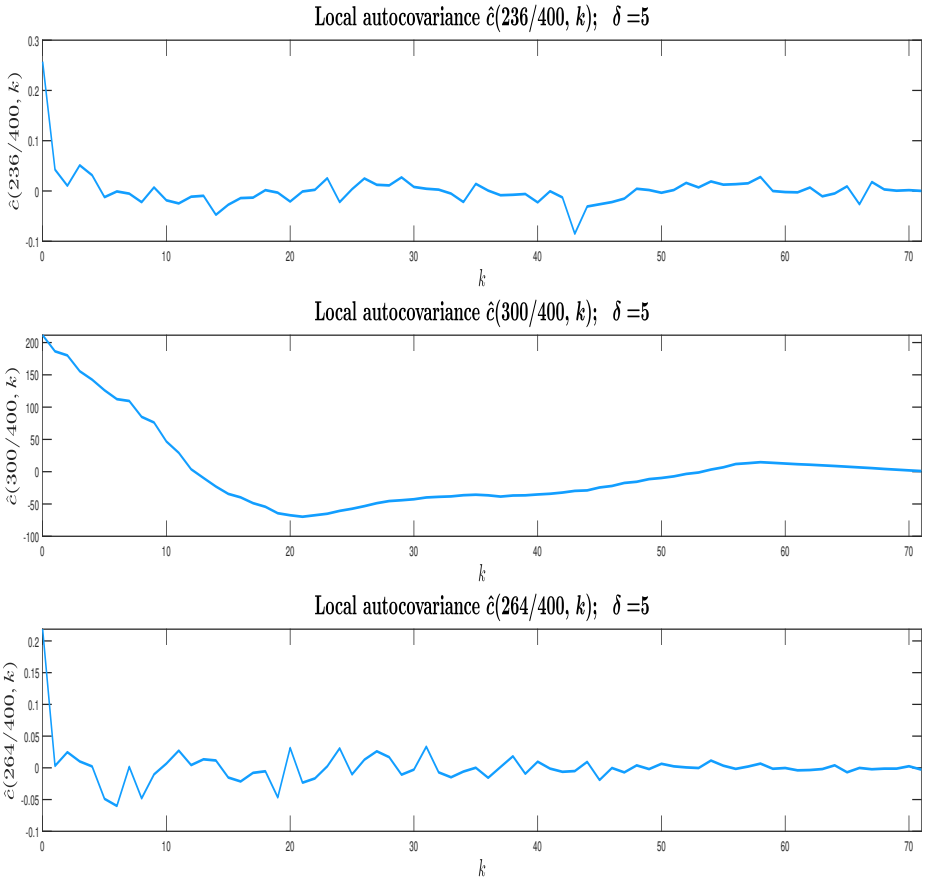


Figure 2·10: The figure plots $\widehat{c}_T(u, k)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 5$.

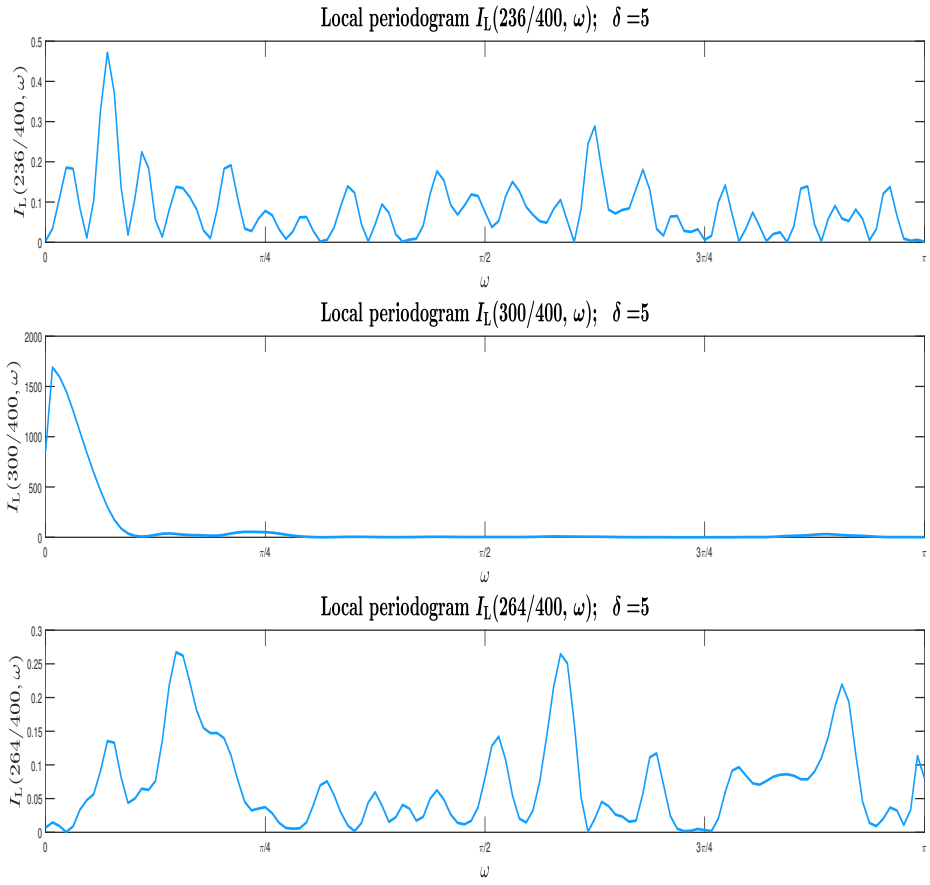


Figure 2·11: The figure plots $I_{\mathrm{L}}(u, \omega)$ for $u = 236/400$, λ_1^0, $264/400$ in the top, mid and bottom panel, respectively. In all panels $\delta = 5$.

2.7 Tables

Table 2.1: Comparison between the theoretical autocovariance and the sample estimates

	$\Gamma_T(k)$	$\widehat{\Gamma}(k)$	$\widehat{\Gamma}(k) - \widehat{d}^*$	$\widehat{\Gamma}_{\mathrm{DK}}(k)$
$k = 0$	0.37	0.54	0.37	0.63
$k = 1$	0.13	0.29	0.13	0.25
$k = 2$	0.01	0.16	0.01	0.03
$k = 5$	0.00	0.05	-0.03	0.00
$k = 10$	-0.03	0.02	-0.07	0.00

The table reports the values of a) the theoretical autocovariance $\Gamma_T(k)$; b) the sample autocovariance $\widehat{\Gamma}(k)$; c) the corrected sample autocovariance $\widehat{\Gamma}(k) - \widehat{d}$; d) the double kernel sample autocovariance $\widehat{\Gamma}_{\mathrm{DK}}(k)$, for $k = 0$, 1, 2, 5 and 10, with the process $\{V_t\}$ as in model M1 with $T = 200$.

Table 2.2: Empirical small-sample size and power of t-test for model M1

$\alpha = 0.05$, $T = 200$	$\delta = 0$ (size)	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.15$	$\delta = 0.25$	$\delta = 1$	$\delta = 1.5$
$\widehat{J}_{\mathrm{DK},T}$	0.068	0.189	0.286	0.460	0.661	0.992	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.045	0.085	0.199	0.332	0.612	0.976	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.046	0.090	0.202	0.333	0.613	0.977	1.000
Andrews (1991)	0.039	0.095	0.185	0.383	0.623	0.968	0.999
Andrews (1991), prewhite	0.115	0.168	0.304	0.447	0.650	0.988	0.999
Newey-West (1987)	0.209	0.272	0.398	0.516	0.689	0.997	1.000
Newey-West (1987), fixed-b (KVB)	0.004	0.018	0.063	0.139	0.301	0.870	0.969
EWC	0.011	0.038	0.137	0.273	0.539	0.978	0.999

Table 2.3: Empirical small-sample size and power of the of t-test for model M2

$\alpha = 0.05$, $T = 200$	$\delta = 0$ (size)	$\delta = 0.15$	$\delta = 0.2$	$\delta = 0.25$	$\delta = 0.3$	$\delta = 0.5$
$\widehat{J}_{\mathrm{DK},T}$	0.059	0.415	0.815	0.974	0.974	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.058	0.262	0.632	0.899	0.899	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.053	0.246	0.616	0.894	0.894	1.000
Andrews (1991)	0.064	0.228	0.564	0.892	0.830	1.000
Andrews (1991), prewhite	0.252	0.564	0.904	0.992	0.991	1.000
Newey-West (1987)	0.133	0.388	0.821	0.981	0.971	1.000
Newey-West (1987), fixed-b (KVB)	0.000	0.077	0.018	0.356	0.356	0.971
EWC	0.004	0.045	0.255	0.632	0.637	1.000

Table 2.4: Empirical small-sample size and power of the of t-test for model M3

$\alpha = 0.05$, $T = 200$	$\delta = 0$ (size)	$\delta = 0.1$	$\delta = 0.15$	$\delta = 0.2$	$\delta = 0.25$
$\widehat{J}_{\mathrm{DK},T}$	0.086	0.552	0.930	0.992	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.065	0.436	0.887	0.971	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.063	0.415	0.875	0.981	1.000
Andrews (1991)	0.017	0.257	0.696	0.950	0996
Andrews (1991), prewhite	0.036	0.456	0.864	0.952	0.998
Newey-West (1987)	0.031	0.344	0.795	0.976	0.994
Newey-West (1987), fixed-b (KVB)	0.000	0.084	0.245	0.442	0.627
EWC	0.000	0.051	0.299	0.699	0.937

Table 2.5: Empirical small-sample size and power of the of t-test for model M4

$\alpha = 0.05$, $T = 200$	$\delta = 0$ (size)	$\delta = 0.1$	$\delta = 0.15$	$\delta = 0.2$	$\delta = 0.25$	$\delta = 0.5$
$\widehat{J}_{\mathrm{DK},T}$	0.067	0.558	0.748	0.870	0.945	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.065	0.301	0.495	0.618	0.736	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.037	0.351	0.508	0.656	0.766	1.000
Andrews (1991)	0.016	0.253	0.448	0.564	0.675	0.999
Andrews (1991), prewhite	0.456	0.804	0.916	0.904	0.992	1.000
Newey-West (1987)	0.346	0.954	0.981	0.821	0.981	1.000
Newey-West (1987), fixed-b (KVB)	0.000	0.000	0.000	0.005	0.015	0.333
EWC	0.024	0.240	0.486	0.596	0.681	0.999

Table 2.6: Empirical small-sample size and power of the DM (1995) test

			(1) $\delta > 0$			
$\alpha = 0.05$, $T = 200$	size	$\delta = 0.2$	$\delta = 0.5$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_{\mathrm{DK},T}$	0.033	0.312	0.551	0.997	1.000	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$	0.042	0.322	0.563	0.999	1.000	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$	0.046	0.348	0.573	0.998	1.000	1.000
Andrews (1991)	0.085	0.254	0.305	0.114	0.000	0.000
Andrews (1991), prewhite	0.085	0.246	0.293	0.401	0.045	0.000
Newey-West (1987)	0.083	0.246	0.299	0.612	0.817	0.782
Newey-West (1987), fixed-b (KVB)	0.002	0.212	0.185	0.000	0.000	0.000
EWC	0.083	0.252	0.268	0.045	0.000	0.000
			(2) $\delta\,(t/T)$ locally stationary			
$\alpha = 0.05$, $T = 200$		$\delta = 0.2$	$\delta = 0.5$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_{\mathrm{DK},T}$		0.278	0.297	0.592	0.889	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$		0.301	0.363	0.634	0969	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$		0.327	0.368	0.642	0.969	1.000
Andrews (1991)		0.255	0.259	0.255	0.110	0.005
Andrews (1991), prewhite		0.249	0.243	0.268	0.188	0.031
Newey-West (1987)		0.281	0.282	0.313	0.268	0.078
Newey-West (1987), fixed-b (KVB)		0.203	0.202	0.178	0.025	0.000
EWC		0.244	0.252	0.219	0.045	0.000
			(3) $\delta\,(t/T)$ locally stationary			
$\alpha = 0.05$, $T = 200$		$\delta = 0.2$	$\delta = 1$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_{\mathrm{DK},T}$		0.540	0.862	0.992	1.000	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$		0.396	0.664	0.988	1.000	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$		0.412	0.724	0.987	1.000	1.000
Andrews (1991)		0.328	0.234	0.235	0.241	0.777
Andrews (1991), prewhite		0.342	0.315	0.512	0.296	0.882
Newey-West (1987)		0.381	0.384	0.720	0.972	0.999
Newey-West (1987), fixed-b (KVB)		0.100	0.032	0.000	0.002	0.040
EWC		0.312	0.152	0.142	0.296	0.852
			(4) case (2) with outliers			
$\alpha = 0.05$, $T = 400$		$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 5$	$\delta = 10$
$\widehat{J}_{\mathrm{DK},T}$		0.694	0.733	0.822	0.981	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},T}$		0.724	0.777	0.846	0.982	1.000
$\widehat{J}_{\mathrm{DK,pw,SLS},\mu,T}$		0.727	0.771	0.847	0.981	1.000
Andrews (1991)		0.192	0.242	0.245	0.203	0.022
Andrews (1991), prewhite		0.182	0.233	0.243	0.288	0.114
Newey-West (1987)		0.222	0.271	0.245	0.345	0.225
Newey-West (1987), fixed-b (KVB)		0.203	0.222	0.212	0.075	0.000
EWC		0.186	0.221	0.174	0.062	0.000

Chapter 3

A New Particle Filter for Nonlinear Dynamic Stochastic General Equilibrium Models

3.1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models are widely used by researchers working in the field of macroeconomics. These models are also increasingly used by central banks around the world, to evaluate the current state of the economy, to analyze the effects of monetary or fiscal policies, and to forecast prominent macroeconomic variables such as the aggregate output and inflation.

The parameter estimation and inference for DSGE models typically requires computing the likelihood function. If the model is driven by Gaussian structural shocks and solved through (log) linearization, evaluating the likelihood can be efficiently implemented using the Kalman filter. In other cases, when the shocks are not Gaussian or the model is nonlinear, alternative filters are required.

One such choice is the particle filter. Since its introduction in Gordon et al. (1993), the particle filter has become a popular method for studying non-linear or non-Gaussian state-space models. Surveys are provided, for example, by Doucet and Johansen (2011) and Creal (2012). Pitt and Shephard (1999) and Kim et al. (1998) are among the first to apply particle filters in finance and economics. When it comes to DSGE models, Fernández-Villaverde and Rubio-Ramírez (2007) is the

first to demonstrate how to use particle filter to evaluate the likelihood function of a nonlinear model. Much progress has been made since; see the review in Herbst and Schorfheide (2015). A challenging issue with particle filtering, however, is that the filter in its standard implementation will breakdown when there is no error term in the measurement equation. To address this, researchers commonly assume that there is a nontrivial amount of measurement errors, say 20% standard deviation of the data. Unfortunately, this practice is questionable because the measurement errors are not introduced to improve the model or its match with the data, but instead as a convenience such that the filter is applicable. It potentially leads to miss-specification, which can further cause bias in parameter estimation and inference.

Motivated by this limitation, in this paper we develop a new particle filter for nonlinear DSGE models without adding measurement errors. The measurement equation in a DSGE model is usually a selection equation that selects a subset of variables from the state vector for the estimation. We can map the entire state vector into two subvectors: a subvector whose components are observed and a subvector whose components are latent. Then, by only sampling and propagating particles of the latent variables, we avoid the need to introduce measurement errors. To implement this filter, a key step is to nonparametrically evaluate the data's density conditional on the latent subvector. We propose a global approximation approach based on series expansion such as the Gram-Charlier A series for this purpose.

As a by-product, the new filter also allows us to study singular DSGE model using the composite likelihood. Economic theory often produce stochastically singular DSGE models when the dimension of the observable is greater than the dimension of shocks. Qu (2018) proposed to use the composite likelihood to bridge such models with data. By combining the composite likelihood with our new filter, we can provide a unified treatment of both singular and nonlinear DSGE models.

3.2 The Model

Throughout this paper we are concerned with the filtering problem of nonlinear Dynamic Stochastic General Equilibrium (DSGE) models whose solution has the following state space representation:

$$Y_t = \Psi(\theta)X_t + u_t, \tag{3.1}$$

$$X_t = \Phi(\theta, X_{t-1}, \epsilon_t). \tag{3.2}$$

Equation (3.1) is the measurement equation characterizing the relationship between data and state. Equation (3.2) is the transition equation describing the dynamics of the system. Specifically, the n_y-by-1 vector Y_t contains observables available as data and the n_x-by-1 vector X_t includes the states, whose dynamics is driven by the structural shocks contained in the n_ϵ-by-1 vector ϵ_t. For convenience, we assume the shocks are orthogonal with unit variance. The n_u-by-1 vector u_t comprises the measurement errors, just to relate our method to the conventional ones. In particular, u_t can have zero variance, meaning measurement errors do not exist as we commonly see in DSGE models. The vector θ consists of the structural parameters, serving as input for the function $\Psi(\cdot)$ and transition function $\Phi(\cdot, X, \epsilon)$. Here, we allow $\Phi(\theta, \cdot, \cdot)$ to be nonlinear. For a concise notation, we will henceforth suppress θ unless further clarity is needed. When Φ is linear, say

$$X_t = \Phi_1 X_{t-1} + \Phi_\epsilon \epsilon_t,$$

we are back to the familiar linear DSGE models, e.g. small scale models such as Lubik and Schorfheide (2004) and medium scale models such as Smets and Wouters (2007). When it comes to the filtering problems of linear models, Kalman filter is the standard choice. For general cases involving nonlinear dynamics, the particle filter

has become a popular approach.

If $n_y > n_\epsilon + n_u$, the system is stochastically singular. It is well known that the conventional time and frequency domain likelihoods are not well defined in this case because the conditional covariance and the spectral density of Y_t are singular. A few approaches have been proposed, including "measurement error" approach (e.g. Altug (1989), Sargent (1989), Ireland (2004)), "more structural shocks" approach (e.g., Schorfheide (2000), Rabanal and Rubio-Ramírez (2005), Smets and Wouters (2007)), and "fewer observables" approach (e.g., Guerron-Quintana (2010), Canova (2014)). Instead of augmenting measurement errors, introducing additional structural shocks, or excluding observables, Qu (2018) suggests a framework based on the composite likelihood. This approach can be conveniently combined with our new particle filter to handle singular nonlinear DSGE models. For now, we assume $n_y \leq n_\epsilon + n_u$.

We will use the same letter, say, X, to denote a random variable and its realization interchangeably. That is, small letter x will NOT be used to denote the realization of X. Instead, it will reserved for the random variable X's standardized counterpart

$$x = \hat{\mathbb{V}}^{-\frac{1}{2}}(X)(X - \hat{\mathbb{E}}(X)),$$

Moreover, Y^t is a shorthand for the filtration $\sigma(Y_1, Y_2, \ldots, Y_t)$, i.e., the information provided by the history of Y_t.

3.3 Traditional Particle Filter and the No-measurement-error Problem

Before introducing the new particle filter, we first review the traditional particle filter and the measurement error issue. The traditional particle filter is constructed using the properties of a Markovian process. Suppose the initial distribution of X is given by $p(X_0|Y^{-1}) = p(X_0)$. For any $t \geq 1$, given $p(X_{t-1}|Y^{t-1})$, the predictive distribution

is

$$p(X_t|Y^{t-1}) = \int p(X_t|X_{t-1}, Y^{t-1})p(X_{t-1}|Y^{t-1})dX_{t-1}.$$

The filtered distribution follows from this distribution and the Bayes Rule:

$$p(X_t|Y^t) = \frac{p(Y_t|X_t, Y^{t-1})p(X_t|Y^{t-1})}{p(Y_t|Y^{t-1})}.$$

The traditional particle filter uses a swarm of particles $\{X_t^j, W_t^j\}_{j=1}^N$ represent a state space model and approximate the distributions of interest. For instance, we approximate the filtered distribution using

$$\hat{p}(X_t|Y_t) = \frac{1}{N} W_t^j \delta(X_t = X_t^j),$$

where $\delta(\cdot)$ is the Dirac measure. Through out this paper, we let $W_t^j = 1$ as we resample in each period. The predictive distribution can be obtained in a similar fashion. Formally, the traditional particle filter proceeds as follows.

Algorithm 3.3.1 (Traditional Particle Filter) The algorithm is initialized by drawing N particles at time zero:

$$X_0^j \sim p(X_0|Y^{-1}) = p(X_0), \quad j = 1, \ldots, N.$$

For any $t \geq 1$, the particles are propagated and updated through three steps recursively:

1. Propagation: Given period-$(t-1)$ particles, i.e.,

$$X_{t-1}^j \sim p(X_{t-1}|Y^{t-1}), \quad j = 1, \ldots, N,$$

propagating according to proposal distribution of choice yields the predictive particles

$$\tilde{X}_t^j \sim q(X_t|X_{t-1}^j, Y^t), \quad j = 1, \ldots, N.$$

The notation of the proposal distribution indicates potential use of information contained in Y_t.

2. Weighting: The predictive particles $\tilde{X}_t^j$ are evaluated using unnormalized weights

$$w_t^j = p(Y_t|\tilde{X}_t^j, Y_{t-1}^j)\frac{p(\tilde{X}_t|X_{t-1}^j, Y^{t-1})}{q(\tilde{X}_t^j|X_{t-1}^j, Y^t)}.$$

   The second factor is the importance weight taking into account the difference between the proposal distribution and transition distribution.

3. Resampling: The recursion is closed by updating filtered state particles X_t^j. Since we have pairs $(\tilde{X}_t^j, w_t^j)$, resampling X_t^j according to w_t^j yields the result.

Remark 3.3.1 When $q(X_t|X_{t-1}, Y^t) = p(X_t|X_{t-1}, Y^{t-1})$, i.e., when the proposal distribution is chosen to be the transition distribution, the particle filter is called bootstrap particle filter. We will use bootstrap particle filter as the default method to compare with our new filter.

Now, we examine how the measurement issue affects computing the likelihood. Suppose we obtain an approximation for the filtered distribution $p(Y_t|X_t, Y^{t-1})$ using the particle filter. Then, the period- t likelihood follows after integrating out X_t:

$$p(Y_t|Y^{t-1}) = \int p(Y_t|X_t, Y^{t-1})p(X_t|Y^{t-1})dX_t.$$

Therefore, we can approximate the period-t likelihood by averaging the unnormalized weights. In particular, for the bootstrap particle filter,

$$\hat{p}(Y_t|Y^{t-1}) = \frac{1}{N}\sum_{j=1}^{N} p(Y_t|\tilde{X}_t^j, Y^{t-1}).$$

Inside the summation, the conditional density $p(Y_t|X_t, Y^{t-1})$ is given by the measurement equation (3.1). Under normality,

$$Y_t|X_t \sim \mathcal{N}(\Psi X_t, \Sigma_u)$$

which gives

$$p(Y_t|X_t) = (2\pi)^{-\frac{n_y}{2}}|\Sigma_u|^{-\frac{1}{2}}(Y_t - \Psi X_t)'\Sigma_u^{-1}(Y_t - \Psi X_t).$$

Clearly, this is well-defined only if Σ_u is nonsingular. Otherwise, suppose the measurement error $u_t = 0$, or $\Sigma_u = 0$, then

$$p(Y_t|X_t^j, Y^{t-1}) = \begin{cases} \infty & \Psi X_t^j = Y_t \\ 0 & \text{otherwise} \end{cases},$$

As a result, the traditional particle filter breaks down. More generally, the filter becomes computationally inefficient or even impractical when the measurement errors have small variances because most particles will receive weights close to zero. To remedy this problem, practitioners often assume there is a sizeable measurement error for each observable. A common practice is to make the magnitude of its standard deviation 10 percent or 20 percent of observable's standard deviation. Here, we argue that a new particle filter can be developed without taking this questionable direction.

3.4 The New Particle Filter

Recall that the measurement equation (3.1) links the states to the observables. When measurement error is absent, the n_y-by-n_x selection matrix Ψ induces a n_y-dimensional subspace containing observables. In this case, we define a companion matrix Π such that

$$\Psi\Pi' = 0,$$

where Π induces a $(n_x - n_y)$-dimensional subspace orthogonal to the row space of Ψ. As it turns out, this observation is the key step toward designing our filter. We thus formally define the following notion:

Definition 3.4.1 (Unobservable) Let X be the state vector and Ψ the selection matrix. A companion matrix Π can be constructed such that $\Psi\Pi' = 0$. Then, $Z = \Pi X$ is defined as unobservable vector whose components are called unobservables.

Define a transformation matrix Λ such that

$$\Lambda = \begin{bmatrix} \Psi \\ \Pi \end{bmatrix}.$$

Note that Λ is invertible by definition. Because Ψ is a selection matrix, with appropriate ordering, Λ will often have a diagonal form, i.e.,

$$\Lambda = \begin{bmatrix} \Psi_y & 0 \\ 0 & I_z \end{bmatrix},$$

where Ψ_y is a n_y-by-n_y matrix with nonzero column vectors and I_z is a n_z-by-n_z identity matrix. Further, we stack the observables Y_t and unobservables $Z_t = \Pi X_t$ as a transformed state vector

$$S_t = \begin{bmatrix} Y_t \\ Z_t \end{bmatrix} = \Lambda X_t.$$

Note that S_t contains the information as the original state variable X_t because Λ is invertible. We can now rewrite the transition equation as

$$S_t = \Lambda X_t = \Lambda \Phi(\Lambda^{-1}\Lambda X_{t-1}, \epsilon_t) = \Lambda \Phi(\Lambda^{-1} S_{t-1}, \epsilon_t) = \Phi_s(S_{t-1}, \epsilon_t).$$

Consequently, the measurement equation can be rewritten as

$$Y_t = \begin{bmatrix} I_y & 0 \end{bmatrix} S_t,$$

where I_y is the n_y-by-n_y identity matrix. This is essentially saying only a subset of the transformed states directly contribute to the observations. As we will see, this representation ensures that when resampling particles, the observable particles can be exactly matched to observations in each recursion.

Turning to the likelihood, we propose using the density of data Y_t conditioning

only on unobservable Z_t, not X_t:

$$p(Y_t|Y^{t-1}) = \int p(Y_t|Z_t, Y^{t-1})p(Z_t|Y^{t-1})dZ_t.$$

That is, after we generate $Z_t^j \sim p(Z_t|Y^{t-1})$ and estimate $p(Y_t|Z_t^j, Y^{t-1})$, the period-t likelihood can be approximated as

$$\hat{p}(Y_t|Y^{t-1}) \approx \frac{1}{N}\sum_{i=1}^{N} \hat{p}(Y_t|Z_t^j, Y^{t-1}).$$

Since unobservable Z_t is orthogonal to observable Y_t, $p(Y_t|Z_t, Y^{t-1})$ is always well-defined and not degenerated. Now we formally lay out the algorithm for the New (Bootstrap) Particle Filter.

Algorithm 3.4.1 (New Particle Filter in Generic Form) The algorithm is initialized by drawing N particles at period-0,

$$X_0^j \sim p(X_0), \quad j = 1, \ldots, N.$$

For any $t \geq 1$ period, the propagation and updating of particles takes three steps:

1. Propagation: Given period-$(t-1)$ particles

   $$X_{t-1}^j \sim p(X_{t-1}|Y^{t-1}), \quad j = 1, \ldots, N,$$

   we propagate them according to the transition equation to obtain

   $$\tilde{X}_t^j \sim p(X_t|X_{t-1}^j, Y^{t-1}), \quad j = 1, \ldots, N.$$

2. Decomposition: The predictive state particle $\tilde{X}_t^j$ can be decomposed as observable particle and unobservable particle using the decomposition matrix Λ:

   $$\Lambda\tilde{X}_t^j = \begin{bmatrix} \Phi\tilde{X}_t^j \\ \Pi\tilde{X}_t^j \end{bmatrix} = \begin{bmatrix} Y_t^j \\ Z_t^j \end{bmatrix} \equiv S_t^j.$$

3. Weighting: The unobservable particles Z_t are evaluated by unnormalized weight

$$w_t^j = \hat{p}(Y_t|Z_t^j, Y_{t-1}),$$

where the conditional density is nonparametrically approximated using the particle pairs (Y_t^j, Z_t^j), described later.

4. Updating: Resample unobservable particles Z_t^j according to the weights w_t^j while fix the unobservable particles $Y_t^j = Y_t$. Then state particles X_t^j can be updated through inversion:

$$X_t^j = \Lambda^{-1} \begin{bmatrix} Y_t \\ Z_t^j \end{bmatrix}$$

Crucially, in the weighting step we approximately compute the observable's density conditional on the unobservables, i.e., $p(Y_t|Z_t, Y^{t-1})$ for each unobservable particle Z_t^j that we generate. We propose a global approximation approach based on the Gram-Charlier A series expansion. For a clear presentation of the idea underlying this approximation, we start by considering the approximation of an unconditional density.

3.4.1 Gram-Charlier Unconditional Density Approximation

Univariate Case

We examine a continuous random variable X. Let $\varphi(t)$ be the characteristic function of its distribution whose density function is $p(x)$, and κ_r its cumulants. By the definition of a characteristic function, we have

$$\varphi(t) = \int_{-\infty}^{\infty} p(x) \exp(itx) dx.$$

Alternatively, $\varphi(t)$ can be rewritten entirely in terms of the cumulants,

$$\varphi(t) = \exp\left(\sum_{r=1}^{\infty} \kappa_r \frac{(it)^r}{r!}\right).$$

Then, we have the cumulant generating function

$$\log \varphi(t) = \sum_{r=1}^{\infty} \kappa_r \frac{(it)^r}{r!}.$$

We consider expanding $p(x)$ around a known distribution with probability density function $q(x)$, characteristic function $\phi(t)$, and cumulants γ_r:

$$\phi(t) = \exp\left(\sum_{r=1}^{\infty} \gamma_r \frac{(it)^r}{r!}\right)$$

The density $q(x)$ here, as generally, is chosen to be that of the normal distribution, but other choices are possible as well. Then,

$$\varphi(t) = \exp\left(\sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(it)^r}{r!}\right) \phi(t).$$

Note that the exponential component can be expressed in terms of Bell polynomials, i.e.,

$$\exp\left(\sum_{r=1}^{\infty} x_r \frac{t^r}{r!}\right) = \sum_{r=0}^{\infty} B_r(x_1, \ldots, x_r) \frac{t^r}{r!},$$

leading to

$$\varphi(t) = \sum_{r=0}^{\infty} B_r(\kappa_1 - \gamma_1, \ldots, \kappa_r - \gamma_r) \frac{(it)^r}{r!} \phi(t).$$

But $(it)^r q(t)$ is the Fourier transform of $(-1)^r d^r q(-x)/dx$ when $q(x)$ is symmetric and $\lim_{x\to\infty} d^r q(x)/dx = 0$ for $r = 0, 1, \ldots$, as

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{d^r q(-x)}{dx^r} \exp(-itx) dx \\
&= -\frac{d^{r-1} q(-x)}{dx^{r-1}} \exp(-itx)\Big|_{-\infty}^{\infty} + (-it) \int_{-\infty}^{\infty} \frac{d^{r-1} q(-x)}{dx^{r-1}} \exp(-itx) dx \\
&= \cdots = (-it)^r \int_{-\infty}^{\infty} q(-x) \exp(-itx) dx = (-1)^r (it)^r \phi(t).
\end{aligned}$$

This implies

$$\begin{aligned} p(x) &= \sum_{r=0}^{\infty} \frac{1}{r!} B_r(\kappa_1 - \gamma_1, \ldots, \kappa_r - \gamma_r)(-1)^r \frac{d^r q(x)}{dx} \\ &= q(x) \sum_{r=0}^{\infty} \frac{1}{r!} B_r(\kappa_1 - \gamma_1, \ldots, \kappa_r - \gamma_r) \left\{ (-1)^r \frac{d^r q(x)}{dx} q^{-1}(x) \right\} \\ &= q(x) \sum_{r=0}^{\infty} \frac{1}{r!} B_r(\kappa_1 - \gamma_1, \ldots, \kappa_r - \gamma_r) H_r(x). \end{aligned}$$

In the last line of the display, when $q(x)$ is the normal density, $H_r(x)$ is the probablist's (not the physicist's) Hermite polynomials, defined by

$$H_r(x) = (-1)^r \frac{d^r q(x)}{dx} q^{-1}(x).$$

In practice, before applying the above expansion, we usually standardize a variable X such that its mean is zero and variance equals to one. That is, we consider

$$x = \hat{\mathbb{V}}(X)^{-\frac{1}{2}}(X - \hat{\mathbb{E}}(X)).$$

Once we achieve an approximation for the density of standardized variable x, we can recover that of X by a change-of-variable technique

$$p_X(X) = \hat{\mathbb{V}}(X)^{-\frac{1}{2}} p_x(x), \quad \text{where } x = \hat{\mathbb{V}}(X)^{-\frac{1}{2}}(X - \hat{\mathbb{E}}(X)).$$

Here, x is used to denote a standardized variable, not realization of the random variable X. Further, we use $p(\cdot)$ to denote the standardized variable's density $p_x(\cdot)$ unless extra clarification is needed.

This standardization simplifies the expressions substantially. First, by construction, x has mean zero and variance one. Second, the cumulants of a standard normal

variable are zero for $r \geq 2$, therefore

$$p(x) = q(x) \sum_{r=0}^{\infty} \frac{1}{r!}(0, 0, \kappa_3, \kappa_4, \ldots, \kappa_r) H_r(x).$$

Applying a fourth order approximation, we get

$$\hat{p}(x) = q(x) \left(1 + \frac{\kappa_3}{3!} H_3(x) + \frac{\kappa_4}{4!} H_4(x)\right).$$

Multivariate Case

For multivariate case, follow Sobel (1963) to consider a k-dimensional random vector $x = (x_1, \ldots, x_n)$ that has zero mean and covariance matrix

$$\mathbb{V}(x) = (1 - \rho) I_n + \rho 1_n 1_n',$$

where $-(k-1)^{-1} < \rho < 1$. Let $s = r_1 + \cdots + r_n$. The multivariate Hermite polynomial is defined by

$$H_{r_1,\ldots,r_n}(x; \rho) = (-1)^s \frac{\partial^s q(x)}{\partial x_1^{r_1} \partial x_2^{r_2} \cdots \partial x_n^{r_n}} q^{-1}(x).$$

Expanding $p(x; \rho)$ around $q(x; \rho)$ gives

$$\begin{aligned} p(x) &= \sum_{r_1=0}^{\infty} \cdots \sum_{r_n=0}^{\infty} (-1)^s \frac{c_{r_1,\ldots,r_n}}{r_1! \cdots r_n!} \frac{\partial^s q(x)}{\partial x_1^{r_1} \cdots \partial x_n^{r_n}}, \\ &= q(x) \sum_{r_1=0}^{\infty} \cdots \sum_{r_n=0}^{\infty} \frac{c_{r_1,\ldots,r_n}}{r_1! \cdots r_n!} H_{r_1,\ldots,r_n}(x; \rho). \end{aligned}$$

By orthogonality of hermite polynomials

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_{i_1,\ldots,i_n}(x; p) H_{r_1,\ldots,r_n}(x; p) q(x; \rho) dx_1 \cdots dx_n = \prod_{\alpha=1}^{k} (r_\alpha) \delta_{ij},$$

where δ_{ij} is the Kronecker delta. This implies that the coefficients satisfy

$$c_{r_1,\ldots,r_n} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_{r_1,\ldots,r_p}(x;\rho)p(x)dx_1 \cdots dx_n. \tag{3.3}$$

On the other hand, for hermite polynomials up to the fourth order we have

$$H_{0,\ldots,0} = 1,$$

$$H_{1,0,\ldots,0} = x_1,$$

$$H_{2,0,\ldots,0} = x_1^2 - 1,$$

$$H_{1,1,0,\ldots,0} = x_1x_2 - \rho,$$

$$H_{3,0,\ldots,0} = x_1^3 - 3x_1,$$

$$H_{2,1,0,\ldots,0} = x_1^2x_2 - x_2 - 2x_1\rho,$$

$$H_{1,1,1,0,\ldots,0} = x_1x_2x_3 - \rho(x_1 + x_2 + x_3)$$

$$H_{4,0,\ldots,0} = x_1^4 - 6x_1^2 + 3,$$

$$H_{3,1,0,\ldots,0} = (x_1^3 - 3x_1)x_2 - 3\rho(x_1^2 - 1),$$

$$H_{2,2,0,\ldots,0} = (x_1^2 - 1)(x_2^2 - 1) - 4\rho(x_1x_2 - \rho) - 2\rho^2,$$

$$H_{2,1,1,0,\ldots,0} = (x_1^2 - 1)(x_2x_3 - \rho) - 2\rho(x_1x_2 - \rho) - 2\rho(x_1x_3 - \rho) - 2\rho^2,$$

$$H_{1,1,1,1,0,\ldots,0} = x_1x_2x_3x_4 - \rho(x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4 - 6\rho) - 3\rho^2.$$

Let $\mu_{r_1,\ldots,r_p} = \mathbb{E}\prod_{\alpha=1}^{k} X_\alpha^{r_\alpha}$. Combining the Hermite polynomial results with (3.3) yields

$$c_{0,\ldots,0} = 1,$$

$$c_{1,0,\ldots,0} = 0,$$

$$c_{2,0,\ldots,0} = 0,$$

$$c_{3,0,\ldots,0} = \mu_{3,0,\ldots,0},$$

$$
\begin{aligned}
c_{2,1,\ldots,0} &= \mu_{2,1,\ldots,0}, \\
c_{1,1,1,0,\ldots,0} &= \mu_{1,1,1,0,\ldots,0}, \\
c_{4,0,\ldots,0} &= \mu_{4,0,\ldots,0} - 3, \\
c_{3,1,0,\ldots,0} &= \mu_{3,1,0,\ldots,0} - 3\rho, \\
c_{2,2,0,\ldots,0} &= \mu_{2,2,0,\ldots,0} - 2\rho^2 - 1, \\
c_{2,1,1,0,\ldots,0} &= \mu_{2,1,1,0,\ldots,0} - \rho - 2\rho^2, \\
c_{1,1,1,1,0,\ldots,0} &= \mu_{1,1,1,1,0,\ldots,0} - 3\rho^2.
\end{aligned}
$$

Note that in our setting x has already been standardized such that $\rho = 0$. Hence expressions above can be further simplified. Once a fourth order Gram-Charlier approximation for standardized variable has been achieved, we apply change of variable to recover the targeting density approximation.

Algorithm 3.4.2 (Gram-Charlier Unconditional Density Approximation in Practice) In practice, unconditional density can be approximated as follows:

1. Compute the sample mean and covariance, and the third and fourth sample cumulants of data.

2. Standardize the variable using the sample mean and covariance.

3. Apply, say, the fourth order Gram-Charlier expression. For Univariate case,

$$\hat{p}(x) = q(x)\left(1 + \frac{\kappa_3}{3!}H_3(x) + \frac{\kappa_4}{4!}H_4(x)\right).$$

4. Apply the change of variable technique to compute the density for the original variable X

$$\hat{p}_X(X) = \hat{\mathbb{V}}(X)^{-\frac{1}{2}}\hat{p}(x), \quad \text{where } x = \hat{\mathbb{V}}(X)^{-\frac{1}{2}}(X - \hat{\mathbb{E}}(X)).$$

3.4.2 Gram-Charlier Conditional Density Approximation

The above approximation is for unconditional density but we need to approximate the density of the observable Y conditional on the unobservable Z. Recall when approximating unconditional density with Gram-Charlier expansions, we standardize the variable of interest using its mean and variance. When it comes to conditional density approximation, what we need for standardization, naturally, is conditional mean and covariance. For a general model, the conditional mean is a nonlinear function of the conditioning variable. Here, we use Hermite polynomials to approximate this function. Specifically, we use powers of Y as left hand side variable while hermite polynomials up to certain order s of standardized Z as right hand side variables. Again, with slight abuse of notation, the small letter

$$z = \hat{\mathbb{V}}(Z)^{-\frac{1}{2}}(Z - \hat{\mathbb{E}}(Z)),$$

denotes standardized Z. The reason why Z has to be standardized is that hermite polynomials are derived from standard normal density. Also note that conditioning on Z is the same as conditioning on z because the standardizing transformation is monotone.

Let the column regressor vector $H(z, s)$ contain all Hermite polynomials $H_{r_1,\ldots,r_{n_z}}(z)$ such that $s = r_1 + \cdots + r_{n_z}$ where n_z is the dimension of z. For univariate case $n_z = 1$, an example of $s = 4$ is

$$H(z, 4) = \begin{bmatrix} 1 \\ z \\ z^2 - 1 \\ z^3 - 3z \\ z^4 - 6z^2 + 3 \end{bmatrix}.$$

Since the dimension of Y, n_y, is usually bigger than 1, we will use multivariate

regression model:

$$Y^j = AH(z^j, s) + v^j, \quad v^j \overset{i.i.d.}{\sim} (0, \Sigma), \quad j = 1, \ldots, N.$$

Recall that with identical regressors across equations the GLS reduces OLS equation by equation, implying

$$\hat{A} = (\mathcal{Y}\mathcal{H}')(\mathcal{H}\mathcal{H}')^{-1}, \quad \text{where } \mathcal{Y} = \begin{bmatrix} Y^1 & \cdots & Y^N \end{bmatrix} \text{ and } \mathcal{H} = \begin{bmatrix} H(z^1, s) & \cdots & H(z^N, s) \end{bmatrix}.$$

It follows that the observable's first moment estimate conditional on a typical observable particle z is

$$\hat{\mathbb{E}}(Y|z) = \hat{A}H(z, s).$$

Approximating the conditional second moment is slightly more complicated as this involves cross products of observables when Y is not univariate. To facilitate the estimation process, we can vectorize unique cross products as left hand side variable but maintaining the right hand side variable. Once elements of the conditional second moment matrix are estimated, repositioning gives the observable's second moment estimate conditional on a typical observable particle z, $\hat{\mathbb{E}}(YY'|z)$. It follows that the observable's covariance estimate conditional on for a typical observable particle z is

$$\hat{\mathbb{V}}(Y|z) = \hat{\mathbb{E}}(YY'|z) - \hat{\mathbb{E}}(Y|z)\mathbb{E}(Y|z)'.$$

With conditional mean and variance estimates at hand, we standardize each observable particle through

$$y = \hat{\mathbb{V}}(Y|z)^{-\frac{1}{2}}(Y - \hat{\mathbb{E}}(Y|z)).$$

Now the standardized observable y is our new targets, and we again have N standardized observable particles y^j as new "data". To get the conditional moments estimate of standardized observable y, we can again apply Hermite polynomial regression.

That is, we estimate desired regressions of powers of y using Hermite polynomials of z up to certain order.

With conditional moments of y, we can recover its conditional cumulants, from which Gram-Charlier Approximation $\hat{p}(y|z)$ for density of standardized observable y conditional on standardized unobservable z is obtained. Again, implementing change of variable recovers

$$\hat{p}_Y(Y|Z) = \hat{p}_Y(Y|z) = \hat{\mathbb{V}}(Y|z)^{-\frac{1}{2}}\hat{p}(y|z), \quad \text{where } y = \hat{\mathbb{V}}(Y|z)^{-\frac{1}{2}}(Y - \hat{\mathbb{E}}(Y|z)).$$

We summarize the algorithm as follows.

Algorithm 3.4.3 (Gram-Charlier Unconditional Density Approximation in Practice) In practice, unconditional density can be approximated as follows:

1. Compute the mean and covariance of conditioning variable Z.
2. Standardize the conditioning variable Z using the sample mean and covariance, reaching
$$z = \hat{\mathbb{V}}(Z)^{\frac{1}{2}}(Z - \hat{\mathbb{E}}(Z)).$$
3. Apply Hermite polynomial regression to compute the conditional mean and variance of Y.
4. Standardize the observable Y using the sample conditional mean and covariance, reaching
$$y = \hat{\mathbb{V}}(Y|z)^{\frac{1}{2}}(Y - \hat{\mathbb{E}}(Y|z)).$$
5. Apply, say, the fourth order Gram-Charlier expression. For univariate case,
$$p(y|z) \approx q(y|z)\left(1 + \frac{\kappa_3(z)}{3!}H_3(y) + \frac{\kappa_4(z)}{4!}H_4(y)\right).$$
6. Apply the change of variable technique to compute the conditional density for

the original variable Y

$$\hat{p}_{Y|Z}(Y|Z) = \hat{\mathbb{V}}(Y|z)^{-\frac{1}{2}}\hat{p}(y|z), \quad \text{where } y = \hat{\mathbb{V}}(Y|z)^{-\frac{1}{2}}(Y - \hat{\mathbb{E}}(Y|z)).$$

With the new generic particle filter algorithm and the Gram-Charlier conditional density approximation algorithm, we have a new particle filter. However, the new particle filter assumes there are no redundant states. In practice, we can use the following operation to remove redundant states from the state vector.

3.4.3 Redundant State Elimination

Redundant states often arise in a linear or partially linear system. Due to the need of standardization, the new filter requires the system to be represented using a minimal number of states. That is, X_t has to be independent in the sense that there exists no vector α such that $\alpha' X_t = 0$. To achieve this, the following procedure can be applied in advance:

We start by computing the unconditional covariance matrix of the state X_t. Let the symmetric n-by-n matrix

$$\Omega = \mathbb{V}(X_t),$$

which has rank r. Apply the eigenvalue decomposition for Ω such that

$$\Omega = U\Lambda U' = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1' \\ U_2' \end{bmatrix} = U_1 \Lambda_1 U_1',$$

where the submatrix U_1 contains eigenvectors corresponding to the nonzero eigenvalues. This implies that the transformed state vector $U'X_t$ has a covariance

$$\mathbb{V}(U'X_t) = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix},$$

which essentially means the next $n - r$ states are redundant as the orthonormal

eigenmatrix U always has a inverse U'. Rewrite the state transition equation to keep only the first r non-redundant state vectors

$$U_1'X_t = U_1'\Phi U \begin{bmatrix} U_1'X_{t-1} \\ 0 \end{bmatrix} + U_1'\Sigma_\epsilon \epsilon_t$$

where we have a zero because $U_2'X_{t-1} = 0$ since it has zero variance. Define

$$X_t^* = U_1'X_t, \quad \Sigma_\epsilon^* = U_1'\Sigma_\epsilon.$$

Also, let

$$\Phi^* = U_1'\Phi U_1,$$

the first r columns of $U_1'\Phi U$ that will multiply X_t^*. Then, the new transition equation is

$$X_t^* = \Phi^* X_t^* + \Sigma_\epsilon^* e_t.$$

We now turn to the measurement equation. We have

$$Y_t = \Psi X_t = \Psi U \begin{bmatrix} U_1'X_t \\ 0 \end{bmatrix}.$$

Define

$$\Psi^* = \Psi U_1$$

the first r columns of ΨU that will multiply $U_1'X_t$. Then,

$$Y_t = \Psi^* X_t^*.$$

The above reduction steps leads to a new representation

$$Y_t \;\; = \;\; \Psi^* X_t^*$$

$$X_t^* = \Phi^* X_t^* + \Sigma_\epsilon^* \epsilon_t.$$

In this presentation, X_t^* is r-by-1 and $n - r$ redundant states are dropped.

Algorithm 3.4.4 (Redundant State Elimination)

1. Compute the unconditional covariance matrix Ω for the state vector X_t. Apply eigenvalue decomposition such that $U \Lambda U' = \Omega$.

2. Collect the orthonormal sub-eigenmatrix U_1 corresponding to the nonzero eigenvalues.

3. Define

$$X_t^* = U_1'^* = U_1' \Phi U_1, \quad Q^* = U_1'^* = \Psi U_1,$$

   Then a minimal state representation is obtained.

3.5 Particle MCMC

Previously we mainly discussed how to approximate likelihood function using a new particle filter. The approximate likelihood can further be embeded into a posterior sampler such as Markov Chain Monte Carlo (MCMC), through which estimation and inference can be conducted. In one sentence: Particle MCMC uses likelihood approximation $\hat{\mathcal{L}}$ in place of the likelihood. The theoretical justification is provided by Andrieu et al. (2010). To see why this works, we introduce the random variable $\mathcal{L}$ as an auxiliary variable. Consider the joint distribution

$$p(\theta, \mathcal{L}|y^T) = p(\theta|y^T) p(\mathcal{L}|\theta, y^T) = \frac{p(y^T|\theta) p(\theta) p(\mathcal{L}|\theta, y^T)}{p(y^T)}.$$

Replacing the intractable likelihood $p(y^T|\theta)$ with its estimator $\hat{\mathcal{L}}$, we have

$$\pi(\theta, \mathcal{L}|y^T) = \frac{\mathcal{L} p(\theta) p(\mathcal{L}|\theta, y^T)}{p(y^T)}.$$

Suppose $\hat{\mathcal{L}}$ is a non-negative and unbiased estimate of $p(y^T|\theta)$,

$$\int \pi(\theta, \mathcal{L}|y^T)d\mathcal{L} = \frac{p(\theta)}{p(y^T)} \int \mathcal{L}p(\mathcal{L}|\theta, y^T)d\mathcal{L} = p(\theta|y^T).$$

For the Metropolis-Hastings ratio,

$$\begin{aligned}
&\frac{\pi(\theta', \mathcal{L}'^T)}{\pi(\theta_m, \mathcal{L}_m|y^T)} \frac{q(\theta|\theta')p(\mathcal{L}_m|\theta_m, y^T)}{q(\theta'|\theta_m)p(\mathcal{L}'|\theta'^T)} \\
=&\frac{\mathcal{L}'p(\theta')p(\mathcal{L}'|\theta'^T)}{\mathcal{L}_m p(\theta_m)p(\mathcal{L}_m|\theta_m, y^T)} \frac{q(\theta_m|\theta')p(\mathcal{L}_m|\theta_m, y^T)}{q(\theta'|\theta_m)p(\mathcal{L}'|\theta'^T)} \\
=&\frac{\mathcal{L}'p(\theta')}{\mathcal{L}_m p(\theta_m)} \frac{q(\theta_m|\theta')}{q(\theta'|\theta_m)}.
\end{aligned}$$

This is idea is known as pseudo-marginal approach.

Algorithm 3.5.1 (Particle Metropolis-Hastings)

1. Initialization: Set θ_0 and run the new particle filter to obtain $\hat{\mathcal{L}}_0$.
2. For $m = 1$ to M,
   (a) Sample $\theta' \sim q(\theta|\theta_{m-1})$ and compute $\hat{\mathcal{L}}$.
   (b) Compute the acceptance probability

$$\alpha_m = \min\left(1, \frac{\hat{\mathcal{L}}'p(\theta')}{\hat{\mathcal{L}}_{m-1}p(\theta_{m-1})} \frac{q(\theta_{m-1}|\theta')}{q(\theta'|\theta_{m-1})}\right) \tag{3.4}$$

   and accept $\{\theta', \hat{\mathcal{L}}'\}$ with probability α_m.

3.6 Nonlinear Composite Likelihood Method

Building on the composite likelihood concept of Lindsay (1988), Qu (2018) develops a likelihood-based framework for analyzing linear singular DSGE models that does not require adding measurement errors, introducing new structural shocks, or excluding observables for the purpose of convenience. The composite likelihood framework

extends naturally the existing state space representation by introducing a selection matrix P_s:

$$Y_{s,t} = P_s Y_t = P_s \Psi X_t + P_s \Psi u_t. \tag{3.5}$$

Consequently, the selection matrix P_s induces a submodel s consisting of the new measurement equation (3.5) and the original transition (3.2). In particular, P_s can be defined such that any submodel s is nonsingular.

Let $Y_{1,t}, \ldots, Y_{S,t}$ be some subvectors of Y_t, each satisfying (3.5) through its selection matrix P_s. Suppose any submodel s is nonsingular and denote their corresponding log-likelihood functions by $\ell_s(\theta)(s = 1, \ldots, S)$. Then, the composite log-likelihood is defined as

$$\ell(\theta) = \sum_{s=1}^{S} \ell_s(\theta).$$

By approximating the component log-likelihood functions ℓ_s using the new particle filter, we now have a nonlinear composite likelihood method. Estimation and inference can be conducted accordingly.

3.7 A Stochastic Volatility Model

For illustrative purposes, we implement the new particle filter algorithm for a stochastic volatility (SV) model. Contrary to the autoregressive conditional heteroscedasticity (ARCH) framework (Engle (1982), Bollerslev et al. (1994)) which models the conditional volatility as a deterministic function of previous observations and past volatilities, a SV model specifies the volatility to follow some latent stochastic process. After discretization, the transition equation ($t \geq 1$) can be written as

$$r_t = \rho_1 r_{t-1} + \exp(0.5 h_{t-1}) u_t, \tag{3.6}$$

$$h_t = \omega + \rho_2(h_{t-1} - \omega) + \sigma v_t, \tag{3.7}$$

where r_t is the observed asset return and

$$u_t, v_t \sim i.i.d.(0, 1).$$

The system is initialized by

$$h_0 \sim \omega + \frac{\sigma}{\sqrt{1-\rho_2^2}} v_0,$$
$$r_0 \sim \frac{\exp(0.5h_0)}{\sqrt{1-\rho_1^2}} u_0.$$

Since u_t plays the role of the measurement error, the traditional particle filter works well. This allows us to use the traditional particle filter as a benchmark to evaluate the new particle filter.

3.7.1 New Particle Filter for a Stochastic Volatility Model

To apply the new particle filter, the state vector are defined as

$$X_t = \begin{bmatrix} r_t \\ h_t \end{bmatrix},$$

while the measurement equation is

$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} r_t \\ h_t \end{bmatrix} = r_t.$$

The latent state is

$$Z_t = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} r_t \\ h_t \end{bmatrix} = h_t.$$

Below we explain how to implement the new particle filter algorithm for this concrete example. Since the unobservable is univariate, no redundant state elimination is needed. In a typical period t, we propagate old particles (r_{t-1}^j, h_{t-1}^j) by entering

the shocks (u_t^j, v_t^j) into the transition equation (3.6) and (3.7).

Now we estimate the conditional density $\hat{p}(r_t|h_t)$ by treating generated particles (r_t^j, h_t^j) as data. We start by transforming the original unobservable particles h_t^j, obtaining

$$z_t^j = \frac{h_t^j - \hat{\mu}(h_t)}{\hat{\sigma}(h_t)}, \quad j = 1, \ldots, N,$$

where

$$\hat{\mu}(h_t) = \frac{1}{N}\sum_{j=1}^{N} h_t^j \quad \text{and } \hat{\sigma}(h_t) = \sqrt{\hat{\mu}(h_t^2) - \hat{\mu}(h_t)^2}.$$

With standardized unobservable particles z_t^j, we can form Hermite polynomials as regressors. Specifically, we construct Hermite polynomials up to fourth order as regressors:

$$H(z_t, 4) = \begin{bmatrix} H_0(z_t) \\ H_1(z_t) \\ H_2(z_t) \\ H_3(z_t) \\ H_4(z_t) \end{bmatrix} = \begin{bmatrix} 1 \\ z_t \\ z_t^2 - 1 \\ z_t^3 - 3z_t \\ z_t^4 - 6z_t^2 + 3 \end{bmatrix}.$$

Next, we need to estimate the conditional mean and standard deviation for the standardization of observable particles. This can be done through Hermite polynomials regressions:

$$r_t^j = H(z_t^j, 4)'\beta_t^1 + v_t^{1,j}, \quad j = 1, \ldots, N,$$

and

$$(r_t^j)^2 = H(z_t^j, 4)'\beta_t^2 + v_t^{2,j}, \quad j = 1, \ldots, N.$$

With regression estimates, we have

$$\hat{\mu}(r_t|z_t^j) = H(z_t^j, 4)'\hat{\beta}_t^1,$$
$$\hat{\sigma}(r_t|z_t^j) = \sqrt{\hat{\mu}(h_t^2|z_t^j) - \hat{\mu}(h_t|z_t^j)^2} = \sqrt{H(z_t^j, 4)'\hat{\beta}_t^2 - (H(z_t^j, 4)'\hat{\beta}_t^1)^2}.$$

Therefore, observable particles r_t^j are standardized as

$$y_t^j = \frac{r_t^j - \hat{\mu}(r_t|z_t^j)}{\hat{\sigma}(r_t|z_t^j)}, \quad j = 1, \ldots, N.$$

Remark 3.7.1 Due to approximation error and estimation uncertainty, the value of $\hat{\sigma}(r_t|z_t^j)$ for certain outlying particles z_t^j can be very small or even negative. The corresponding standardized observable can thus take extremely large values or have the wrong sign. Although such cases are fairly rare, they can have large effects on the estimation. A rule of thumb is to set

$$\hat{\sigma}(r_t|h_t) = 0.1\hat{\sigma}(r_t) \quad \text{whenever } \hat{\sigma}(r_t|h_t) < 0.1\hat{\sigma}(r_t).$$

Now that we have standardized observable and unobservable particles (y_t^j, z_t^j), we are ready for a fourth order Gram-Charlier conditional density approximation. To compute the conditional cumulants, we first need to approximate standardized observable's moments up to fourth order, conditional on standardized unobservable. That is, we need to run

$$\begin{aligned}
y_t^j &= H(z_t^j, 4)'\beta_t^1 + e_t^j, \\
(y_t^j)^2 &= H(z_t^j, 4)'\beta_t^2 + e_t^{2,j}, \\
(y_t^j)^3 &= H(z_t^j, 4)'\beta_t^3 + e_t^{3,j}, \\
(y_t^j)^4 &= H(z_t^j, 4)'\beta_t^4 + e_t^{4,j}, \\
&\text{for } j = 1, \ldots, N.
\end{aligned}$$

Let

$$m_1(z_t) = \hat{\mu}(y_t|z_t) = H(z_t, 4)'\hat{\beta}_t^1,$$
$$m_2(z_t) = \hat{\mu}(y_t^2|z_t) = H(z_t, 4)'\hat{\beta}_t^2,$$
$$m_3(z_t) = \hat{\mu}(y_t^3|z_t) = H(z_t, 4)'\hat{\beta}_t^3,$$
$$m_4(z_t) = \hat{\mu}(y_t^4|z_t) = H(z_t, 4)'\hat{\beta}_t^4.$$

Then, the conditional cumulants can be recovered through

$$k_3(z_t) = m_3(z_t) - 3m_2(z_t)m_1(z_t) + 2m_1(z_t)^3,$$
$$k_4(z_t) = m_4(z_t) - 4m_3(z_t)m_1(z_t) - 3m_2(z_t)^2 + 12m_2(z_t)m_1(z_t) - 6m_1^4(z_t).$$

Therefore, a fourth order Gram-Charlier approximation for the standardized observable's density conditional on unobservable is

$$\hat{p}(y_t|z_t) = \hat{q}(y_t)\left(1 + \frac{k_3(z_t)}{3!}H_3(y_t) + \frac{k_4(z_t)}{4!}H_4(y_t)\right)$$

By the changing-of-variable technique, the original conditional density is

$$\hat{p}_r(r_t|h_t) = \frac{1}{\hat{\sigma}(r_t|z_t)}\hat{p}(y_t|z_t), \quad \text{where } y_t = \frac{r_t - \hat{\mu}(r_t|z_t)}{\hat{\sigma}(r_t|z_t)}.$$

With a valid conditional density approximation at hand, the weighting stage of the new particle filter assigns to the unobservable particles z_t^j the weights

$$w_t^j = \hat{p}_r(r_t|h_t^j),$$

where the value of r_t is fixed to the value of the observation at period t. The likelihood in period t can be easily approximated, as

$$p(r_t|r^{t-1}) = \int p(r_t|h_t, r^{t-1})p(h_t|r^{t-1})dh_t \approx \frac{1}{N}\sum_{j=1}^{N} w_t^j.$$

Next the unobservable particles are resampled with probability proportional to their weights so that particles with larger weights are replicated more often. Finally, apply the inverted transformation matrix, here just the 2-dimensional identity matrix, on the particle pairs (r_t, h_t^j). Then the state particles are updated.

3.7.2 A Comparison with the Traditional Particle Filter

Now we evaluate the performance of the new particle filter by comparing it with the traditional particle filter. It is noteworthy to emphasize that the implementation of the traditional particle filter would require a different formulation: the state vector would be defined as

$$X_t^o = \begin{bmatrix} r_{t-1} \\ h_{t-1} \end{bmatrix},$$

while the measurement equation

$$Y_t^o = r_t = \rho_1 r_{t-1} + \exp(0.5 h_{t-1}) u_t.$$

Since

$$r_t \sim \mathcal{N}(\rho_1 r_{t-1}, \exp(h_{t-1})),$$

a particle (r_{t-1}, h_{t-1}) is evaluated by

$$w_t^o = \frac{1}{\sqrt{2\pi \exp(h_{t-1})}} \exp\left(-\frac{(r_t - \rho_1 r_{t-1})^2}{2 \exp(h_{t-1})} \right).$$

As a result, traditional particle filter works fairly well for SV model and can serve as a good benchmark.

To conduct the simulation exercise, the parameter values are set according to estimates in Table 7 of Kim, Shephard, and Chib (1998),

$$\rho_1 = 0, \quad \rho_2 = 0.9780, \quad \omega = -0.8650, \quad \sigma = 0.1580.$$

The sample size $T = 1,000$ and the particle size $N = 200,000$. Given a typical set of parameter values, the approximated log-likelihoods of each period produced by the new particle filter and the traditional particle filter are compared in Figure 3·1. Even though the two filters use entirely different state formulations and particle weighting schemes, the plots suggest they yield very similar per period log-likelihood approximations, except for the initial period. Figure 3·2 displays the full log-likelihood comparison given by the two filters. This shows the new particle filter indeed performs fairly well for the stochastic volatility model.

3.8 A New Keysesian Monetary DSGE Model

We consider a prototypical New Keynesian Monetary DSGE model studied in Clarida et al. (2000), Lubik and Schorfheide (2004), and Qu (2018). Since this small scale DSGE model is linear and Gaussian, Kalman filter can produce its exact likelihood. This will allow us to evaluate the performance of the new particle filter in more details.

The model can be concisely summarized by three equilibrium conditions, i.e., IS schedule, Phillips curve, and monetary policy:

$$y_t = \mathbb{E}_t y_{t+1} - \tau(r_t - \mathbb{E}_t \pi_{t+1}) + g_t, \tag{3.8}$$

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa(y_t - z_t), \tag{3.9}$$

$$r_t = \rho_r r_{t-1} + (1 - \rho_r)(\psi_1 \pi_t + \psi_2[y_t - z_t]) + \epsilon_{r,t}, \tag{3.10}$$

where π_t and r_t denote (log) deviations of inflation, and nominal interest rate from their corresponding steady states, and in the case of output y_t, from a trend path. The IS schedule (3.8) results from households' intertemporal decisions regarding consumptions and savings. The parameter β is the discount factor and τ the intertemporal substitution elasticity. The exogenous process g_t captures the aggregate demand shock. The expectational Phillips curve (3.9) characterizes inflation dynamics due

to monopolistically competitive firms' pricing rigidity. The slope κ describes how responsive the pricing adjustment is with respect to the output gap $y_t - z_t$, where z_t is the natural rate, i.e., the output level provided that the economy is frictionless. The monetary policy rule (3.10) specifies the central bank sets a nominal interest rate by responding to the output gap and inflation. The parameters ψ_1 and ψ_2 determine the tradeoff. The interest rate dynamics is also subject to its lagged value, as well as the monetary shock $\epsilon_{r,t}$.

The system is driven by three exogenous forces. Besides the serially uncorrelated monetary policy shock $\epsilon_{r,t}$, the demand shock g_t and natural rate z_t evolve according to AR(1) with autoregressive coefficients ρ_g and ρ_g:

$$g_t = \rho_g g_{t-1} + \epsilon_{g,t}, \quad z_t = \rho_z z_{t-1} + \epsilon_{z,t}.$$

The three innovations satisfy

$$\begin{bmatrix} \epsilon_{r,t} \\ \epsilon_{g,t} \\ \epsilon_{z,t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_r^2 & 0 & 0 \\ 0 & \sigma_g^2 & \rho_{gz}\sigma_g\sigma_z \\ 0 & \rho_{gz}\sigma_g\sigma_z & \sigma_z^2 \end{bmatrix} \right),$$

where σ_r, σ_g, and σ_g demote their standard deviations. In particular, $\epsilon_{g,t}$ and $\epsilon_{z,t}$ are allowed to have nonzero correlation ρ_{gz}.

In a typical setting, the observables are (log) levels of output, inflation, and interest rate:

$$\begin{aligned} \bar{Y}_t &= Y^* + CY_t \\ &= \begin{bmatrix} 0 \\ \pi^* \\ \gamma^* + \pi^* \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \\ r_t \end{bmatrix}. \end{aligned}$$

Here the output is pre-filtered and π^* and r^* are annualized steady state of inflation and real interest rates. Now the discount factor in (3.9) can be replaced by $\beta = (1 + r^*/100)^{-1/4}$. For convenience, the structural parameters are installed into the

vector $\theta = (\tau, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z, \rho_{gz}, \pi^*, r^*)'$.

In order to solve the model using Sims (2002) method, we augment the original system with two expectation equations

$$y_t = \mathbb{E}_{t-1} y_t + \eta_{y,t}, \quad \pi_t = \mathbb{E}_{t-1}\pi_t + \eta_{\pi,t}.$$

Define the state vector $X_t = (y_t, \pi_t, r_t, g_t, z_t, \mathbb{E}_t y_{t+1}, \mathbb{E}_t \pi_{t+1})'$. Then, in the canonical form $\Gamma_0 X_t = \Gamma_1 X_{t-1} + \Psi \epsilon_t + \Pi \eta_t$, we have

$$\begin{bmatrix} 1 & 0 & \tau & -1 & 0 & -1 & -\tau \\ -\kappa & 1 & 0 & 0 & \kappa & 0 & -\beta \\ -(1-\rho_r)\psi_2 & -(1-\rho_r)\psi_1 & 1 & 0 & (1-\rho_r)\psi_2 & 0 & (1-\rho_r)\psi_2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \\ r_t \\ g_t \\ z_t \\ \mathbb{E}_t y_{t+1} \\ \mathbb{E}_t \pi_{t+1} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_r & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \\ g_{t-1} \\ z_{t-1} \\ \mathbb{E}_{t-1} y_t \\ \mathbb{E}_{t-1} \pi_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{r,t} \\ \epsilon_{g,t} \\ \epsilon_{z,t} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{y,t} \\ \eta_{\pi,t} \end{bmatrix}. \quad (3.11)$$

Since the number of structural shocks is only three while the number of states are seven, the redundant state elimination algorithm need to be run in advance to ensure the state space representation has reached the minimal form. As a result, the selection matrix Ψ will not necessary have a zero block and the companion matrix Π must be solved through $\Psi\Pi' = 0$. Fortunately, many computation sorfwares provide such built-in algorithm. Assume this is done and we have unobservable

$$Z_t = \Pi X_t.$$

As we will see, the linear and Gaussian structure of the model allows us to evaluate

the performance of the new particle filter through the following three key items that can be exactly computed:

1. The particle weights $w^j = p(Y_t|Z_t^j, Y^{t-1})$, given unobservable particles Z_t^j, can be computed through the Kalman particle filter that is to be discussed.

2. The filtering density $p(Z_t|Y^t)$ can be generated by smoothing the unobservable particles resampled according to the particle weights given by the Kalman particle filter.

3. The log-likelihood $\log p(Y^t|\theta)$ can be produced by the Kalman filter.

3.8.1 Kalman Particle Filter

Once prior filtering distribution $p(Z_{t-1}|Y^{t-1})$ and particles Z_t^j are given, we can compute the exact particle weights $w^j = p(Y_t|Z_t^j, Y^{t-1})$ by exploiting the joint normality, similar to the derivation of Kalman filter. Since the filtering distribution $p(Z_t|Y^t)$ can be updated as a result of resampling, we name the procedure the Kalman particle filter for convenience. For notation, let

$$\bar{X}_{t|s} \equiv \mathbb{E}(X_t|Y^s), \quad \bar{Y}_{t|s} \equiv \mathbb{E}(Y_t|Y^s),$$
$$\Omega_{t|s} \equiv \mathbb{V}(X_t|Y^s), \quad \Sigma_{t|s} \equiv \mathbb{V}(Y_t|Y^s), \quad \Gamma_{t|s} \equiv \mathbb{E}(X_t Y_t'^s) - \bar{X}_{t|s}\bar{Y}'_{t|s}.$$

Recall the Kalman filter is initialized by setting

$$X_0 \sim \mathcal{N}(\bar{X}_{0|0}, \Omega_{0|0}).$$

Starting from this initial condition, the filtering proceeds recursively. In particular, given

$$X_{t-1}|Y^{t-1} \sim \mathcal{N}(\bar{X}_{t-1|t-1}, \Omega_{t-1|t-1}), \quad \epsilon_t \sim \mathcal{N}(0, \Omega_\epsilon),$$

the optimal prediction of X_t and its mean squared error (MSE) are

$$\bar{X}_{t|t-1} = \Phi \bar{X}_{t-1|t-1},$$
$$\Omega_{t|t-1} = \Phi \Omega_{t-1|t-1} \Phi' + \Omega_\epsilon,$$

the optimal prediction of Y_t and its MSE are

$$\bar{Y}_{t|t-1} = \Psi \bar{X}_{t|t-1},$$
$$\Sigma_{t|t-1} = \Psi \Omega_{t|t-1} \Psi' + \Sigma_u,$$

and the covariance between $X_{t|t-1}$ and $Y_{t|t-1}$satisfies

$$\Gamma_{t|t-1} = \Omega_{t|t-1} \Psi',$$

which leads to the relationship

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} \Big| Y^{t-1} \sim \mathcal{N}\left(\begin{bmatrix} \bar{X}_{t|t-1} \\ \bar{Y}_{t|t-1} \end{bmatrix}, \begin{bmatrix} \Omega_{t|t-1} & \Gamma_{t|t-1} \\ \Gamma'_{t|t-1} & \Sigma_{t|t-1} \end{bmatrix} \right). \tag{3.12}$$

The forecast error and Kalman gain are

$$v_t = Y_t - \bar{Y}_{t|t-1},$$
$$K_t = \Gamma_{t|t-1} \Sigma^{-1}_{t|t-1}.$$

Therefore, the filtering distribution can be updated as

$$X_t | Y^{t-1} \sim \mathcal{N}(\bar{X}_{t|t}, \Omega_{t|t}),$$

where

$$\bar{X}_{t|t} = \bar{X}_{t|t-1} + K_t v_t,$$
$$\Omega_{t|t} = \Omega_{t|t-1} - K_t \Gamma'_{t|t-1}.$$

The derivation of the Kalman particle filter follows the same logic since the joint Gaussian distribution (3.12) also implies the conditional density $p(Y_t|X_t, Y^{t-1})$. Define the following conditional moments:

$$\bar{Y}_{t|X,s} \equiv \mathbb{E}(Y_t|X_t, Y^s), \quad \Omega_{t|s} \equiv \mathbb{V}(X_t|Y^s),$$
$$\Sigma_{t|X,s} \equiv \mathbb{V}(Y_t|X_t, Y^s), \quad \Gamma_{t|s} \equiv \mathbb{E}(X_t Y_t'^s) - \bar{X}_{t|s}\bar{Y}'_{t|s}.$$

The state forecast error and Kalman gain are

$$e_t = X_t - \bar{X}_{t|t-1},$$
$$P_t = \Gamma'_{t|t-1}\Omega^{-1}_{t|t-1}.$$

Then,

$$Y_t|X_t, Y^{t-1} \sim \mathcal{N}(\bar{Y}_{t|X,t-1}, \Sigma_{t|X,t-1}),$$

where

$$\bar{Y}_{t|X,t-1} = \bar{Y}_{t|t-1} + P_t e_t,$$
$$\Sigma_{t|X,t-1} = \Sigma_{t|t-1} - P_t\Gamma_{t|t-1}.$$

Recall $Z_t = \Pi X_t$ and define the particle Kalman gain

$$Q_t = (\Pi\Gamma_{t|t-1})'(\Pi\Omega_{t|t-1}\Pi'^{-1}.$$

Then, we arrive at

$$Y_t|Z_t, Y^{t-1} \sim \mathcal{N}(\bar{Y}_{t|Z,t-1}, \Sigma_{t|Z,t-1}),$$

where

$$\bar{Y}_{t|Z,t-1} = \bar{Y}_{t|t-1} + Q_t\Pi e_t, \tag{3.13}$$
$$F_{t|Z,t-1} = F_{t|t-1} - Q_t\Pi\Gamma_{t|t-1}. \tag{3.14}$$

3.8.2 A New Particle Filter for the New Keysesian Monetary Model

Now we illustrate how to implement the new particle filter for the New Keysesian Monetary Model. We still exploit Hermite polynomial regressions and Gram-Charlier density approximation to estimate the conditional density $p(Y_t|Z_t, Y^{t-1})$. The linear Gaussian structure of the model allows us to refine the algorithm. To facilitate the discussion and show how the new particle filter nests simpler method, we will present the refinement, although we can certainly avoid making use of such a special structure.

Note that this time we have an observable vector

$$Y_t = \begin{bmatrix} y_t \\ \pi_t \\ r_t \end{bmatrix} = \Psi X_t.$$

The linear structure of (3.13) implies only Hermite polynomials up to two are needed, so the regressors are set as

$$H(Z_t, 2) = \begin{bmatrix} 1 \\ Z_t \end{bmatrix}.$$

After observable and unobservable particles are generated, we arrange them through

$$\mathcal{Y}_t = \begin{bmatrix} Y_t^1 & Y_t^2 & \cdots & Y_t^N \end{bmatrix}, \quad \mathcal{H}_t = \begin{bmatrix} H(Z_t^1, 2) & H(Z_t^2, 2) & \cdots & H(Z_t^N, 2) \end{bmatrix}.$$

So the parameter matrix is

$$\hat{A}_t = (\mathcal{Y}_t \mathcal{H}_t')(\mathcal{H}_t \mathcal{H}_t'^{-1}.$$

We then have the conditional expectation

$$\hat{\mathbb{E}}(\mathcal{Y}_t|\mathcal{Z}_t) = \begin{bmatrix} \hat{\mathbb{E}}(y_t|Z_t^1) & \hat{\mathbb{E}}(y_t|Z_t^2) & \cdots & \hat{\mathbb{E}}(y_t|Z_t^N) \\ \hat{\mathbb{E}}(\pi_t|Z_t^1) & \hat{\mathbb{E}}(\pi_t|Z_t^2) & \cdots & \hat{\mathbb{E}}(\pi_t|Z_t^N) \\ \hat{\mathbb{E}}(r_t|Z_t^1) & \hat{\mathbb{E}}(r_t|Z_t^2) & \cdots & \hat{\mathbb{E}}(r_t|Z_t^N) \end{bmatrix} = \hat{A}_t \mathcal{H}_t.$$

Moreover, (3.14) implies homoskedacticity holds for the multivariate regression. Hence the conditional covariance matrix for standardization is simply

$$\mathbb{V}(Y_t|Z_t) = \frac{1}{N}(\mathcal{Y}_t - \hat{A}_t\mathcal{H}_t)(\mathcal{Y}_t - \hat{A}_t\mathcal{H}_t)'$$

Finally, since the normal distribution only have the first two nontrivial cumulants, the particle weights can be obtained through

$$p(Y_t|Z_t) = \hat{\mathbb{V}}(Y_t|Z_t)^{-\frac{1}{2}} \frac{1}{(2\pi)^{\frac{n_y}{2}}} \exp\left(-\frac{1}{2} y_t y_t'\right), \quad \text{where } y_t = \hat{\mathbb{V}}(Y_t|Z_t)^{-\frac{1}{2}}(Y_t - \hat{\mathbb{E}}(Y_t|Z_t)).$$

3.8.3 Evaluating the New Particle Filter's Performance

Now we evaluate the performance of the new particle filter. To rule out any indeterminacy, the parameter values are calibrated as in Table 3.1, according to the posterior means reported in Lubik and Schorfheide (2004) when using the post-1982 sample from 1982:IV to 1997:IV that excludes the Volcker-disinflation period:

Table 3.1: Calibration for Lubik and Schorfheide (2004) Model

Parameter	Value	Parameter	Value
ψ_1	2.19	ψ_2	0.30
ρ_r	0.84	π^*	3.43
r^*	3.01	κ	0.58
τ^{-1}	1.86	ρ_g	0.83
ρ_z	0.85	ρ_{gz}	0.36
σ_r	0.18	σ_g	0.18
σ_z	0.64		

Solving (3.8) in the form of vector autoregression $X_t = \Phi X_{t-1} + \Sigma_\epsilon \epsilon_t$ yields

$$\begin{bmatrix} y_t \\ \pi_t \\ r_t \\ g_t \\ z_t \\ \mathbb{E}_t y_{t+1} \\ \mathbb{E}_t \pi_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -0.90 & 1.62 & 0.59 & 0 & 0 \\ 0 & 0 & -0.96 & 1.69 & -0.28 & 0 & 0 \\ 0 & 0 & 0.46 & 0.67 & -0.11 & 0 & 0 \\ 0 & 0 & 0 & 0.83 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.85 & 0 & 0 \\ 0 & 0 & -0.44 & 0.76 & -0.13 & 0 & 0 \\ 0 & 0 & -0.41 & 0.74 & 0.60 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \\ g_{t-1} \\ z_{t-1} \\ \mathbb{E}_{t-1} y_t \\ \mathbb{E}_{t-1} \pi_t \end{bmatrix}$$
$$+ \begin{bmatrix} -0.19 & 0.51 & 0.41 \\ -0.21 & 0.29 & -0.20 \\ 0.10 & 0.12 & -0.08 \\ 0 & 0.18 & 0 \\ 0 & 0.23 & 0.60 \\ -0.09 & 0.13 & -0.09 \\ -0.09 & 0.32 & 0.42 \end{bmatrix} \begin{bmatrix} \epsilon_{r,t} \\ \epsilon_{g,t} \\ \epsilon_{z,t} \end{bmatrix}.$$

Note that $\epsilon_t = (\epsilon_{r,t}, \epsilon_{g,t}, \epsilon_{z,t})'$ are uncorrelated shocks with unit variance.

Depending on the number of observables, we have three cases:

1. One observable and two unobservables.

2. Two observables and one unobservable.

3. Three observables and no unobservable.

In the simulation study, the number of particles N is set to be 200,000 and the sample size T is set to be 200.

We use the second case to demonstrate how well the new particle filter perform in the weighting stage and the updating stage. Say we observe y_t and π_t. In a typical period, we plot the conditional density $p(Y_t|Z_t^j, Y^{t-1})$ and the filtering density $p(Z_t|Y^t)$ in Figure 3·3. The results clearly demonstrate that the particle filter does a fairly good job in obtaining the observable's density conditional on the unobservable as particle weights.

Finally, for all three cases, we compare the log-likelihood produced by the new particle filter with the Kalman filter and a traditional particle filter. The Kalman

filter should obviously serve as a good benchmark. On the other hand, a traditional particle filter requires the addition of measurement errors. Following the procedure of An and Schorfheide (2007), Herbst and Schorfheide (2015), Herbst and Schorfheide (2019) and many others, we choose the standard deviations of the measurement errors to be 20% of the sample standard deviation of the time series. The results are plotted in Figure 3·4-3·6. The graphs suggest for the small scale DSGE model, the new particle filter is much more accurate than the traditional particle filter.

Remark 3.8.1 The new particle filter relies crucially on the conditional density $p(Y_t|Z_t, Y^{t-1})$. One concern is that when the number of observables equals to that of states, there will be no room for unobservable particles. As it turns out, in this case the conditional density trivially reduces to $p(Y_t|Y^{t-1})$, whose approximation only requires the cumulants of Y_t conditional on Y^{t-1}. The regression method used to estimate moments now only uses constant as the regressor. Moreover, all particles receive the same weights and thus resampling steps can be skipped.

3.9 Conclusion

This paper has developed a new particle filter for analyzing nonlinear DSGE models. It can avoid arbitrary measurement error specification that is commonly needed by the traditional particle filter. The main feature of the new particle filter is that only unobservable particles are updated and corresponding conditional density, or particle weights, are estimated nonparametrically using series expansions. To estimate the conditional momments used in the series expansion step, flexible regression method is proposed. As a plus, it allows us to extend the composite likelihood framework to a nonlinear setting, thus providng a unified econometric framewok for analyzing both singular and nonsingular DSGE models.

3.10 Figures

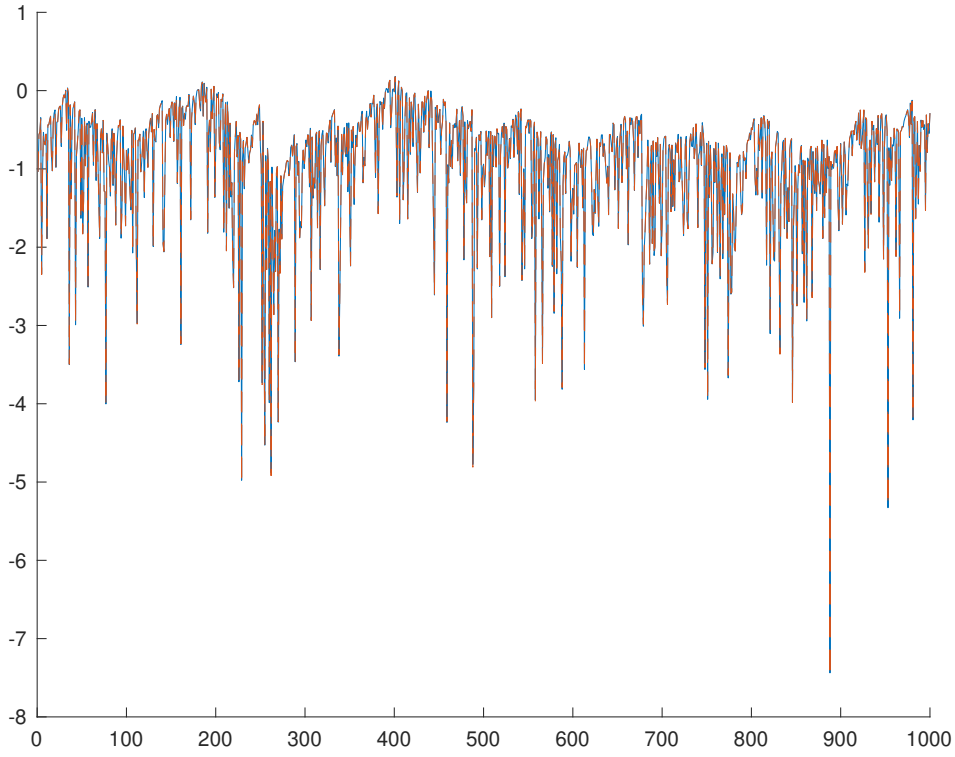


Figure 3·1: Log-likelihood of each period ($T = 1,000, N = 200,000$): New Particle Filter (Blue Solid) vs Traditional Particle Filter (Red Dashed)

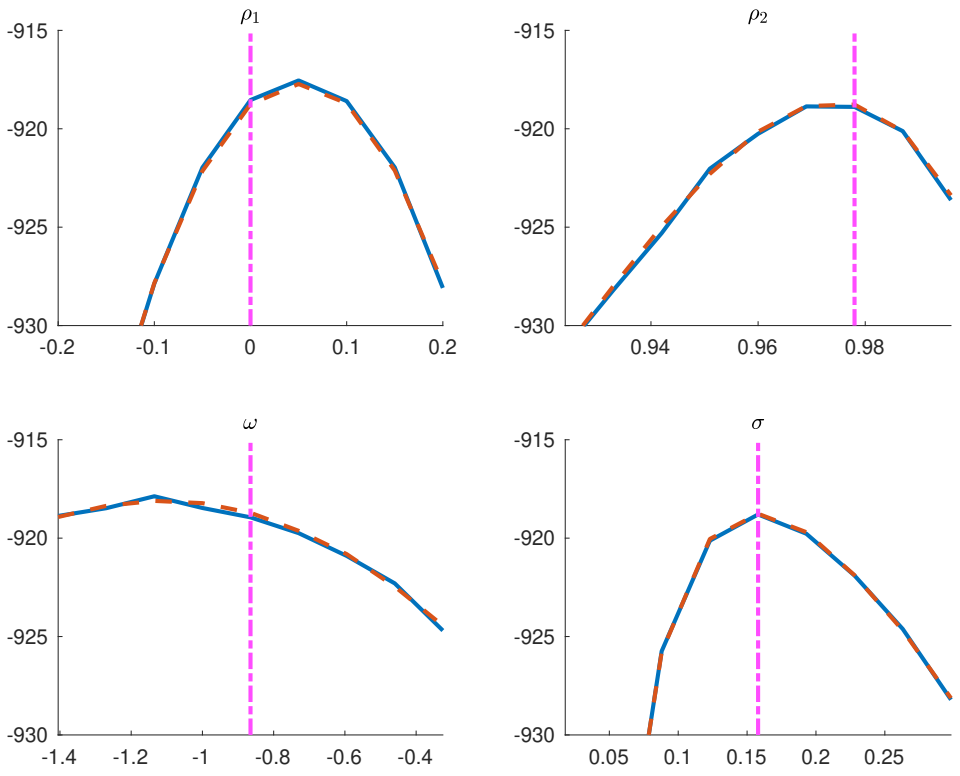


Figure 3·2: Log-likelihood ($T = 1,000, N = 200,000$): New Particle Filter (Blue Solid) vs Traditional Particle Filter (Red Dashed). Vertical dashed-dotted lines signify true values.

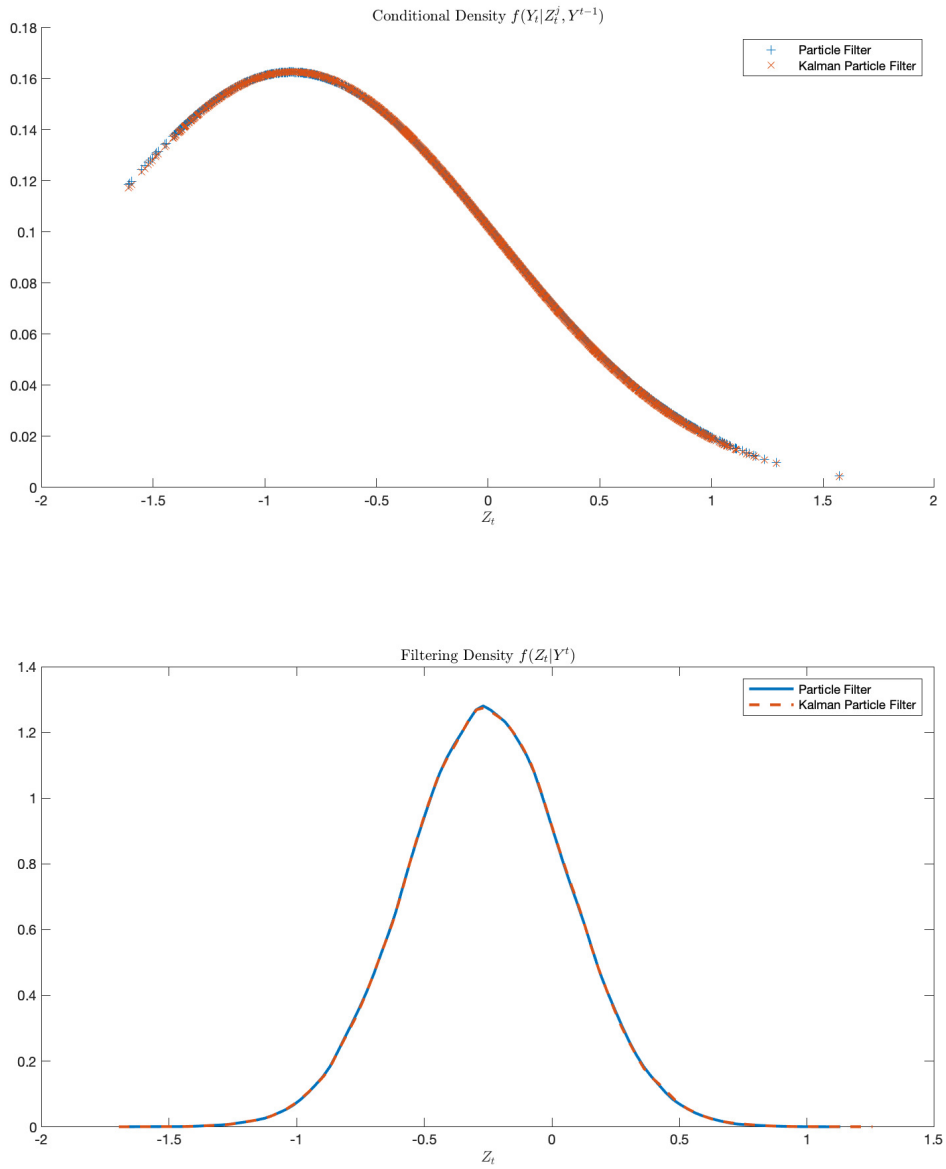


Figure 3·3: Particle Filter vs. Kalman Particle Filter. Top Panel: Conditional Densities of Particle Filter (Blue Plus) and Kalman Particle Filter (Red Cross). Bottom Panel: Filtering Densities of Particle Filter (Blue Solid) and Kalman Particle Filter (Red Dashed)

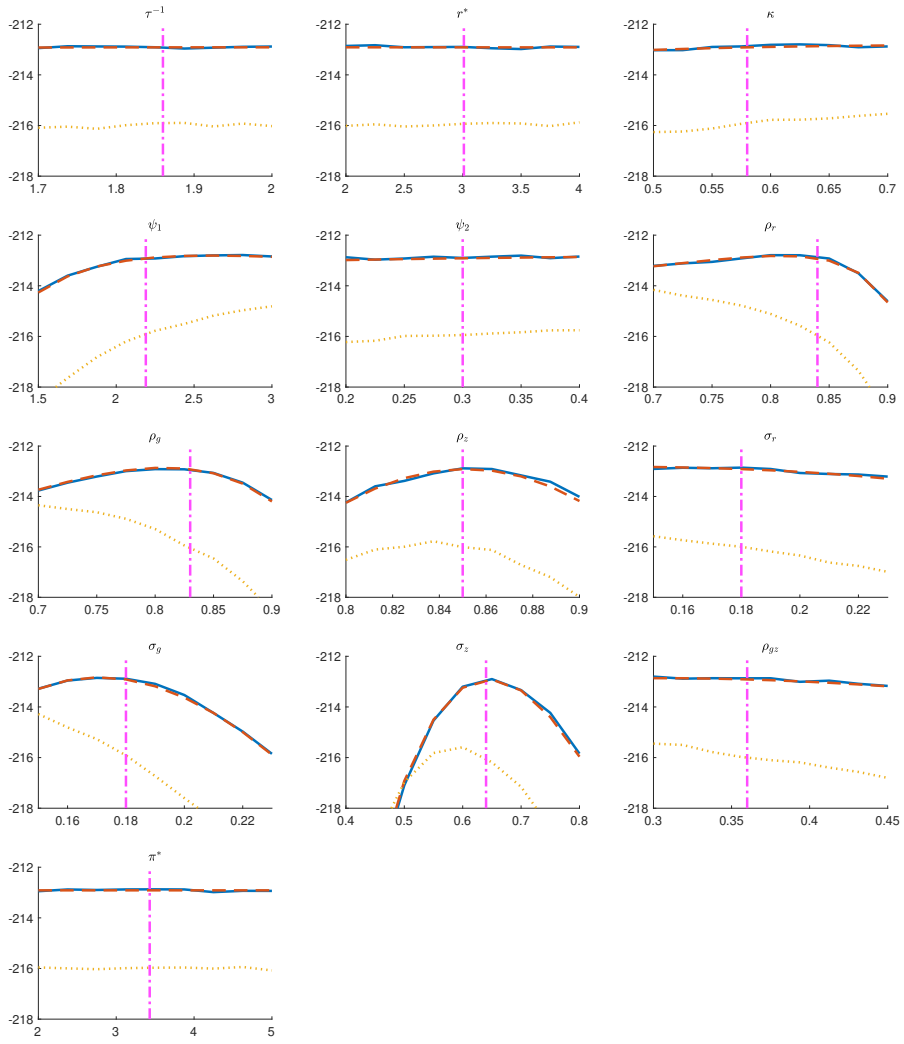


Figure 3·4: New Particle Filter (Blue Solid) vs. Kalman Filter (Red Dashed) and Traditional Particle Filter (Yellow Dotted): log-liklihood profile for each parameter. One observable and two unobservables. $T = 200$ data and $N = 200,000$ particles are used. Vertical dashed-dotted lines signify true values.

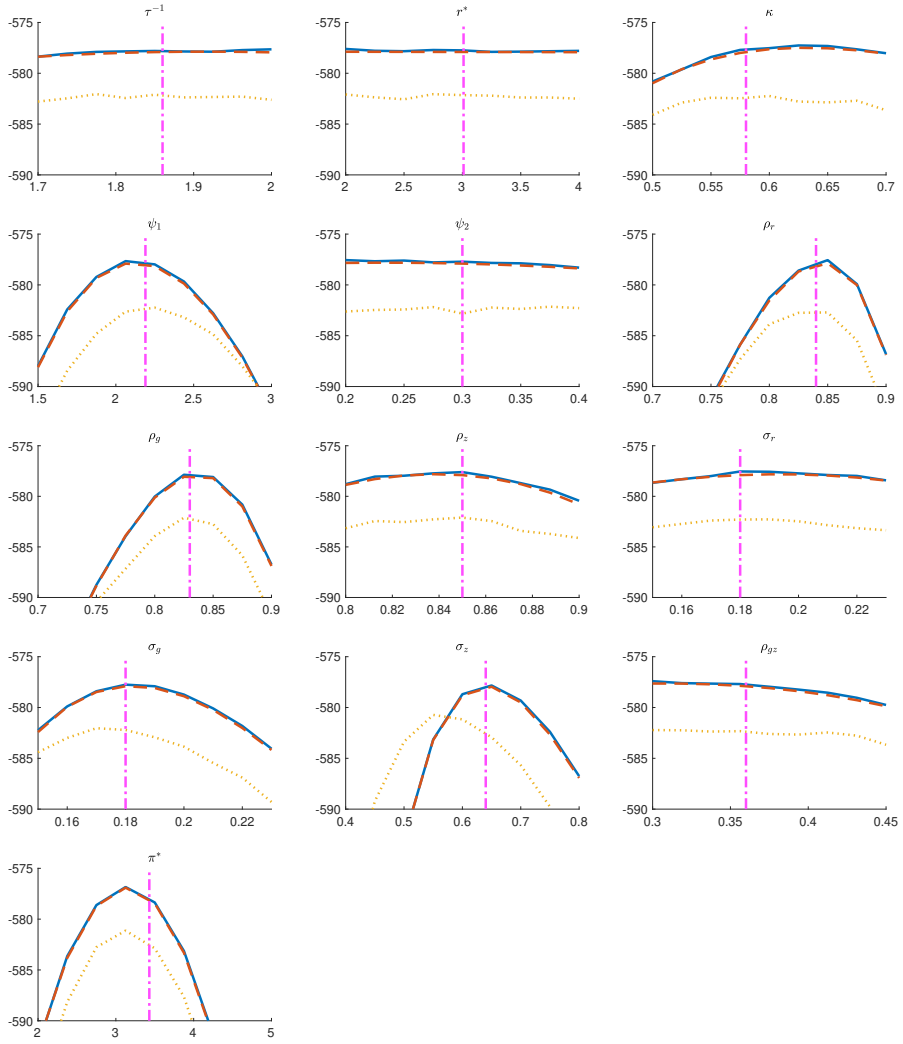


Figure 3·5: New Particle Filter (Blue Solid) vs. Kalman Filter (Red Dashed) and Traditional Particle Filter (Yellow Dotted): log-liklihood profile for each parameter. Two observables and one unobservable. $T = 200$ data and $N = 200,000$ particles are used. Vertical dashed-dotted lines signify true values.

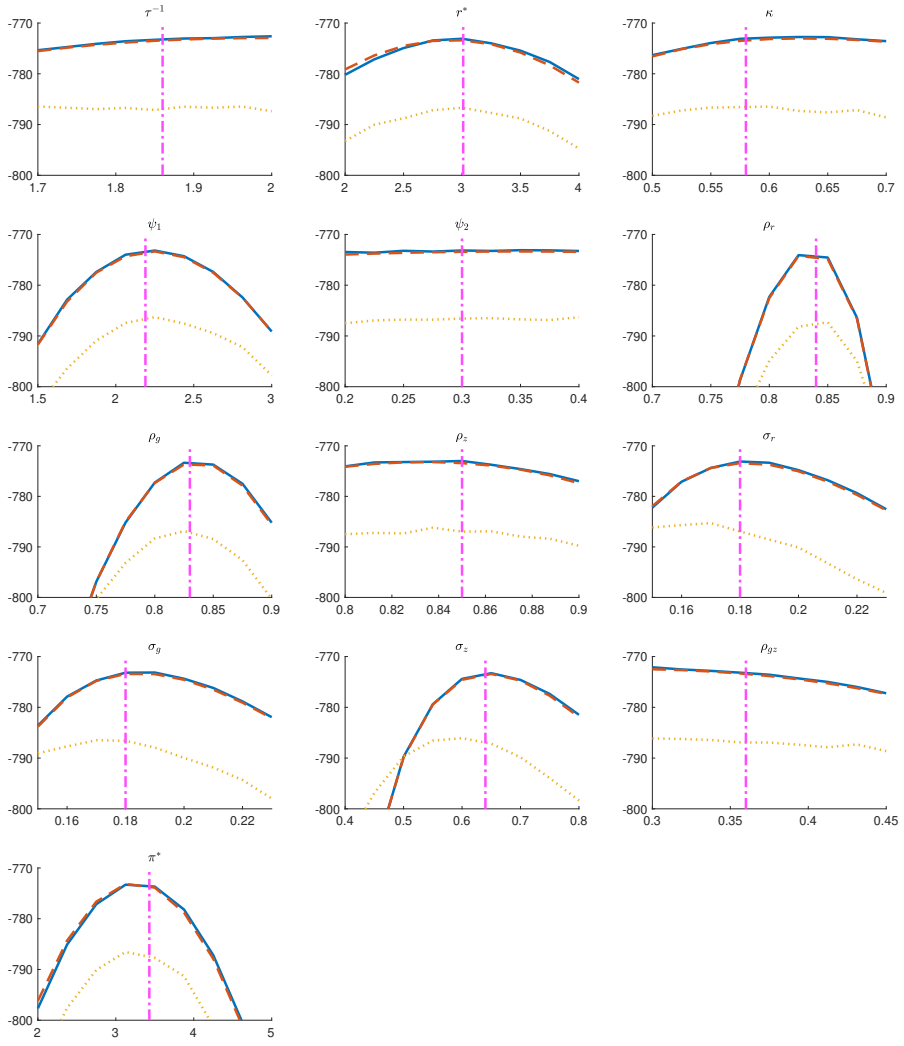


Figure 3·6: New Particle Filter (Blue Solid) vs. Kalman Filter (Red Dashed) and Traditional Particle Filter (Yellow Dotted): log-liklihood profile for each parameter. Three observables and zero unobservable. $T = 200$ data and $N = 200,000$ particles are used. Vertical dashed-dotted lines signify true values.

Appendix A

Proof of Chapter 1

A.1 Proof of Theorem 1.4.1

Since $(u_t, v_t)'$ is bivariate normal and $E[e_t|v_t] = 0$ by construction, e_t and v_t must be independent for all t. The intertemporal independence of $(u_t, v_t)'$ implies e_t must be independent of $v_1, \ldots, v_T$ and thus $x_0, \ldots, x_{T-1}$. The result follows. □

A.2 Proof of Theorem 1.4.2

To derive the asymptotic distribution for $\hat{\beta}$ and $\hat{\delta}$, I apply the Frisch-Waugh-Lovell theorem to simplify computation. Denote the matrix and its demeaned counterpart

$$X = \begin{bmatrix} \vdots & \vdots \\ x_{t-1} & \tilde{v}_t \\ \vdots & \vdots \end{bmatrix}, \quad \mathcal{X} = \begin{bmatrix} \vdots & \vdots \\ x_{t-1}^{\mu} & \tilde{v}_t^{\mu} \\ \vdots & \vdots \end{bmatrix},$$

which implies

$$\mathcal{X}'X = \begin{bmatrix} \sum x_{t-1}^{\mu} x_{t-1} & \sum x_{t-1}^{\mu} \tilde{v}_t \\ \sum x_{t-1}^{\mu} \tilde{v}_t & \sum \tilde{v}_t^{\mu} \tilde{v}_t \end{bmatrix}, \quad \mathcal{X}'\epsilon = \begin{bmatrix} \sum x_{t-1}^{\mu} \epsilon_t \\ \sum \tilde{v}_t^{\mu} \epsilon_t \end{bmatrix}.$$

Moreover, denote the demeaned Ornstein-Uhlebeck process and scaling matrix, (the reason of whose elements' different degrees will become obvious)

$$J_c^{\mu}(r) = J_c(r) - \int J_c(r)dr, \quad \Gamma_T = \begin{bmatrix} T & 0 \\ 0 & T^{1/2} \end{bmatrix}.$$

Since the proxy variable

$$\tilde{v}_t = \frac{c - \tilde{c}}{T} x_{t-1} + v_t,$$

the resulting error term can be represented as

$$\epsilon_t = e_t + \frac{\tilde{c} - c}{T} \delta x_{t-1}.$$

For entries in $\Gamma_T^{-1} \mathcal{X} X \Gamma_T^{-1}$,

$$\begin{aligned}
&T^{-2} \sum x_{t-1}^{\mu} x_{t-1} \\
=&T^{-2} \sum x_{t-1}^2 - \left(T^{-3/2} \sum x_{t-1} \right)^2 \\
\Rightarrow&\sigma_v^2 \int J_c(r)^2 dr - \sigma_v^2 \left(\int J_c(r) dr \right)^2 \\
=&\sigma_v^2 \int J_c^{\mu}(r)^2 dr,
\end{aligned}$$

$$\begin{aligned}
&T^{-3/2} \sum x_{t-1}^{\mu} \tilde{v}_t \\
=&(c - \tilde{c}) T^{-5/2} \sum x_{t-1}^2 + T^{-3/2} \sum x_{t-1} v_t \\
&+ (\tilde{c} - c) T^{-1/2} \left(T^{-3/2} \sum x_{t-1} \right)^2 - T^{-5/2} \sum x_{t-1} \sum v_t \\
\rightarrow&0,
\end{aligned}$$

$$\begin{aligned}
&T^{-1} \sum \tilde{v}_t^{\mu} \tilde{v}_t \\
=&(c - \tilde{c})^2 T^{-3} \sum x_{t-1}^2 + 2(c - \tilde{c}) T^{-2} \sum x_{t-1} v_t + T^{-1} \sum v_t^2 \\
&- (c - \tilde{c})^2 T^{-1} \left(T^{-3/2} \sum x_{t-1} \right)^2 + 2(\tilde{c} - c) T^{-3} \sum x_{t-1} \sum v_t - T^{-1} \left(T^{-1/2} \sum v_t \right)^2 \\
\rightarrow&\sigma_v^2.
\end{aligned}$$

Note the off diagonal entry vanish asymptotically due to different rate of convergence among the regressors. For entries in $\Gamma_T^{-1}\mathcal{X}\epsilon$,

$$\begin{aligned}
&T^{-1}\sum x_{t-1}^{\mu}\epsilon_t \\
=&T^{-1}\sum x_{t-1}e_t - T^{-2}\sum x_{t-1}\sum e_t \\
&+(\tilde{c}-c)\delta T^{-2}\sum x_{t-1}^2 + (c-\tilde{c})\delta\left(T^{-3/2}\sum x_{t-1}\right)^2 \\
\Rightarrow&\sigma_v\sigma_e\int J_c(r)dG(r) - \sigma_v\sigma_e\int J_c(r)drZ + (\tilde{c}-c)\delta\sigma_v^2\int W(r)^2dr \\
&+(c-\tilde{c})\delta\sigma_v^2\left(\int W(r)dr\right)^2 \\
=&\sigma_v\sigma_e\int W_c^{\mu}(r)dG(r) + (\tilde{c}-c)\delta\sigma_v^2\int J_c^{\mu}(r)^2dr,
\end{aligned}$$

$$\begin{aligned}
&T^{-1/2}(\tilde{c}-c)\sum x_t^{\mu}\epsilon_t \\
=&(c-\tilde{c})\delta T^{-3/2}\sum x_{t-1}e_t - (c-\tilde{c})^2\delta^2T^{-5/2}\sum x_{t-1}^2 + T^{-1/2}\sum v_te_t \\
&+(\tilde{c}-c)\delta T^{-3/2}\sum x_{t-1}v_t + (\tilde{c}-c)\delta T^{-5/2}\sum x_{t-1}\sum e_t \\
&+(c-\tilde{c})^2\delta^2T^{-1/2}\left(T^{-3/2}\sum x_{t-1}\right)^2 - T^{-3/2}\sum v_t\sum e_t \\
&+(c-\tilde{c})\delta T^{-5/2}\sum x_{t-1}\sum v_t \\
\Rightarrow&\sigma_v\sigma_e\tilde{Z},
\end{aligned}$$

where $\tilde{Z}$ is a standard normal variable. Collecting relevant terms gives

$$\begin{bmatrix} T(\hat{\beta}-\beta) \\ T^{1/2}(\hat{\delta}-\delta)\end{bmatrix} = \left(\Gamma_T^{-1}\mathcal{X}X\Gamma_T^{-1}\right)^{-1}\Gamma_T^{-1}\mathcal{X}'\epsilon \Rightarrow \begin{bmatrix} \frac{\sigma_e\int J_c^{\mu}(r)dG(r)}{\sigma_v\int J_c^{\mu}(r)^2dr} + (\tilde{c}-c)\delta \\ \frac{\sigma_e}{\sigma_v}\tilde{Z}\end{bmatrix}.$$

Taking into account the fact that $\hat{\sigma}_e^2 \to \sigma_e^2$, I obtain the convergence results for each t-statistic under the null the

$$t(\beta_0) = \frac{T^{-1}\sum x_{t-1}^{\mu}\epsilon_t}{\hat{\sigma}_e\left(T^{-2}\sum x_{t-1}^{\mu}x_{t-1}\right)^{1/2}}$$

$$\Rightarrow \frac{\sqrt{1-\theta^2}\int_0^1 J_c^\mu(r)dG(r) + (\tilde{c}-c)\theta\int_0^1 J_c^\mu(r)^2dr}{\sqrt{1-\theta^2}(\int_0^1 J_c^\mu(r)^2dr)^{1/2}}$$

$$= Z + \frac{(\tilde{c}-c)\theta}{(1-\theta^2)^{1/2}}\left(\int_0^1 J_c^\mu(r)^2dr\right)^{1/2},$$

$$t(\delta_0) = \frac{T^{-1/2}\sum \tilde{v}_t{}^\mu \epsilon_t}{\hat{\sigma}_e(T^{-1}\sum(\tilde{v}_t^\mu \tilde{v}_t))^{1/2}}$$

$$\Rightarrow \tilde{Z}.$$

□

A.3 Proof of Lemma 1.4.3

Following White (1958) and Satchell (1984), the quadratic form $\sum_{t=1}^T x_{t-1}x_t - \alpha\sum_{t=1}^T x_{t-1}^2$ and $\sum x_{t-1}^2$ can be writen as $x'Ax$ and $x'Bx$, respectively, where the $T \times T$ matrices A and B are defined as

$$A = -\frac{1}{2}\begin{bmatrix} 2\alpha & -1 & \dots & 0 & 0 \\ -1 & 2\alpha & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 2\alpha & -1 \\ 0 & 0 & \dots & -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Moreover, $\sum_{t=1}^T (x_t - \alpha x_{t-1})^2 = x'y - \alpha^2 x'Bx - 2\alpha x'Ax$. Let D_T be the $T \times T$ matrix

$$D_T = \begin{bmatrix} p & q & \dots & 0 & 0 \\ q & p & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & p & q \\ 0 & 0 & \dots & q & 1 \end{bmatrix}.$$

The joint moment generating function of $x'Ax$ and $x'By$

$$M(u,v) = E\left(\exp(ux'Ax + vx'Bx)\right)$$

$$= (2\pi)^{-T/2} \int \exp\left(-x' D_T y/2\right) dy = (\det D_T)^{-1/2},$$

with $p = 1 + \alpha^2 + 2\alpha u - 2v$ and $q = -(\alpha + u)$. Notice that

$$\det D_T = D(T) = pD(T-1) - q^2 D(T-2).$$

Taking into account the fact that $D(1) = 1$ and $D(2) = p - q^2$, I obtain

$$D(T) = \frac{1-s}{r-s} r^T + \frac{1-r}{s-r} s^T,$$

where r and s are roots of the equation $x^2 - px + q^2$. That is,

$$r = \frac{p + \sqrt{p^2 - 4q^2}}{2}, \quad s = \frac{p - \sqrt{p^2 - 4q^2}}{2},$$

with

$$\begin{aligned} p^2 - 4q^2 &= (1-\alpha^2)^2 - 4\alpha(1-\alpha^2)u - 4(1-\alpha^2)u^2 - 4(1+\alpha^2)v - 8\alpha uv + 4v^2 \\ &= (1-\alpha^2)\left(1 - \alpha^2 - 4\alpha u - 4u^2 - \frac{4(1+\alpha^2)}{1-\alpha^2} v - \frac{8\alpha}{1-\alpha^2} uv + \frac{4}{1-\alpha^2} v^2\right), \end{aligned}$$

where the second equation holds if $|\alpha| \neq 1$. □

A.4 Proof of Lemma 1.4.4

The joint moment generating funtion of $x'AxT^{-1}$ and $x'BxT^{-2}$ can be obtained using $m_T(u, v) = M(uT^{-1}, vT^{-2})$. For $\alpha = \exp(cT^{-1})$, we have

$$\begin{aligned} p &= 2\left(1 + (c+u)T^{-1} + (c^2 + cu - v)T^{-2}\right) + O(T^{-3}), \\ q &= -1 - (c+u)T^{-1} - \left(c^2/2\right) T^{-2} + O(T^{-3}), \\ p^2 - 4q^2 &= 4(c^2 + 2cu - 2v)T^{-2} + O(T^{-3}). \end{aligned}$$

Denote $\kappa = (c^2 + 2cu - 2v)^{1/2}$. It follows that

$$r = 1 + (c + u + \kappa)T^{-1} + O(T^{-2}),$$
$$s = 1 + (c + u - \kappa)T^{-1} + O(T^{-2}),$$

and thus the joint moment generating function of $\left(\int_0^1 J_c(r)dW(r), \int_0^1 J_c(r)^2 dr\right)$

$$\begin{aligned} m(u,v) &= \lim_{T\to\infty} m_T(u,v) \\ &= \left\{ \frac{1}{2\kappa} \exp(c+u)[(\kappa - (c+u))\exp(\kappa) + (\kappa + (c+u))\exp(-\kappa)] \right\}^{-1/2} \\ &= \exp\left(-\frac{1}{2}(\kappa + c + u)\right) \left\{ \frac{1}{2\kappa}[\kappa - (c+u) + (\kappa + (c+u))\exp(-2\kappa)] \right\}^{-1/2}, \end{aligned}$$

as given in Phillips (1987). Moreover,

$$\begin{aligned} \kappa &= |c| \left(1 - \frac{2u}{|c|} - \frac{2v}{c^2}\right)^{1/2} \\ &= |c| \left(1 - \frac{u}{|c|} - \frac{v}{c^2} - \frac{u^2}{2c^2} + O(|c|^{-3})\right) \\ &= -c - u + \frac{v}{c} + \frac{u^2}{2c} + O(|c|^{-2}), \end{aligned}$$

which implies

$$\exp\left(-\frac{1}{2}(\kappa + c + u)\right) = \exp\left(-\frac{v}{2c} - \frac{u^2}{4c} + O(|c|^{-2})\right).$$

Taking into account the fact that

$$\begin{aligned} &\lim_{c\to-\infty} \left\{ \frac{1}{2\kappa}[\kappa - (c+u) + (\kappa + (c+u))\exp(-2\kappa)] \right\}^{-1/2} \\ &= \lim_{c\to-\infty} \left\{ \frac{\kappa - (c+u)}{2\kappa} + \frac{\kappa + (c+u)}{2\kappa \exp(2\kappa)} \right\}^{-1/2} = 1, \end{aligned}$$

we have

$$L(u,v) = \lim_{c\to-\infty} m[(-2c)^{1/2}u, (-2c)v] = \exp\left(\frac{u^2}{2} + v\right).$$

□

A.5 Proof of Corollary 1.4.5

Enough to show as $c \to -\infty$,

$$(-2c)^{1/2}\left(\int_0^1 J_c(r)dr\right) \to 0.$$

Since $J_c(r)$ is Gaussian, it is easy to show

$$\int_0^1 J_c(r)dr \equiv N(0,v),$$

where $v = 1/c^2 + (1/2c^3)(e^{2c} - 4e^c + 3)$. Hence $(-2c)^{1/2}v \to 0$ as $c \to -\infty$, and the result follows. □

A.6 Proof of Theorem 1.4.6

Note for $c \leq 0$ and $\theta \leq 0$,

$$t(\beta_0) \Rightarrow Z - c\frac{\theta}{(1-\theta^2)^{1/2}}\left(\int_0^1 J_c^\mu(r)^2 dr\right)^{1/2} \leq Z.$$

Moreover, as $c \to -\infty$,

$$Z + \left(\frac{|c|}{2}\right)^{1/2}\frac{\theta}{(1-\theta^2)^{1/2}}\left(-2c\int_0^1 J_c^\mu(r)^2 dr\right)^{1/2} \to -\infty.$$

It must be true that

$$\{t(\beta_0) > x\} \subseteq \{Z > x\},$$

and thus $\Pr(t(\beta_0) > x) \leq \Pr(Z > x)$. □

A.7 Proof of Theorem 1.4.7

Since $\beta = \beta_0 + b/T$, now we have

$$T\left(\hat{\beta} - \beta\right) = T\left(\hat{\beta} - \beta_0\right) + b + o_p(1).$$

It follows that

$$\begin{aligned} t(\beta_0) &= \frac{T^{-1}\sum x_{t-1}^{\mu}\epsilon_t}{\hat{\sigma}_e\left(T^{-2}\sum x_{t-1}^{\mu}x_{t-1}\right)^{1/2}} + \frac{b}{\hat{\sigma}_e\left(T^{-2}\sum x_{t-1}^{\mu}x_{t-1}\right)^{-1/2}} \\ &\Rightarrow Z + \left(\frac{(\tilde{c}-c)\theta}{(1-\theta^2)^{1/2}} + \frac{b}{(1-\theta^2)^{1/2}}\right)\left(\int_0^1 J_c^{\mu}(r)^2 dr\right)^{1/2}. \end{aligned}$$

□

A.8 Proof of Lemma 1.5.1

Recall the joint weak convergence

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} v_t \\ T^{-1/2}\sum_{t=1}^{T} e_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_v W(1) \\ \sigma_e G(1) \end{bmatrix}.$$

By the continuous mapping theorem, the result follows. □

A.9 Proof of Theorem 1.5.2

For $k = 1, \ldots, m+1$, we have for each segment

$$T^{-1} \sum_{t=T_{k-1}+1}^{T_k} x_{t-1}^{\mu_k} \epsilon_t \Rightarrow \sigma_e \sigma_v \int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r) dG(r) - \sigma_v^2 \delta \int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2 dr,$$

$$T^{-2} \sum_{t=T_{k-1}+1}^{T_k} x_{t-1}^{\mu_k 2} \Rightarrow \sigma_v^2 \int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2 dr,$$

where $J_c^{\mu_k}(r) = J_c(r) - (\lambda_k - \lambda_{k-1})^{-1} \int_{\lambda_{k-1}}^{\lambda_k} J_c(s) ds$. By Continuous Mapping Theorem,

$$\begin{aligned} t_k(\beta_0) &\Rightarrow \frac{\sqrt{1-\theta^2} \int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r) dG(r) - c\theta \int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2 dr}{\sqrt{1-\theta^2} \left(\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2 dr \right)^{1/2}} \\ &\leq \frac{\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r) dG(r)}{\left(\int_{\lambda_{k-1}}^{\lambda_k} J_c^{\mu_k}(r)^2 dr \right)^{1/2}} = q_k, \end{aligned}$$

where the inequality holds almost sure because $0 \leq \theta \leq 1$ and $c \leq 0$. □

Appendix B

Proof of Chapter 2

B.1 Proof of Theorem 2.3.1

Let $\overline{V}_j = (Tr_j)^{-1} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} V_t$, $\mu_{2,j}(u) = \mathbb{E}(V_{\lfloor Tu \rfloor})^2$ for $T_{j-1}^0 \leq Tu \leq T_j^0$ and $\overline{\mu}_{2,j} = r_j^{-1} \int_{\lambda_{j-1}^0}^{\lambda_j^0} \mu_{2,j}(u)\, du$. By Assumption 2.2.1-2.2.2-(i), the latter implying ergodicity, it follows that for fixed $k \geq 0$ that

$$
\begin{aligned}
\widehat{\Gamma}(k) &= \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1+k}^{\lfloor T\lambda_j^0 \rfloor} V_t V_{t-k} - \left(\sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} V_t \right)^2 + o_{\text{a.s.}}(1) \\
&= \sum_{j=1}^{m_0+1} \int_{\lambda_{j-1}^0}^{\lambda_j^0} c(u,\, k)\, du + \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1+k}^{\lfloor T\lambda_j^0 \rfloor} \mathbb{E}(V_t)\, \mathbb{E}(V_{t-k}) \\
&\quad - \left(\sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} V_t \right)^2 + O\left(T^{-1}\right) + o_{\text{a.s.}}(1) \\
&= \int_0^1 c(u,\, k)\, du + \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1+k}^{\lfloor T\lambda_j^0 \rfloor} \mathbb{E}(V_t)\, \mathbb{E}(V_{t-k}) \\
&\quad - \left(\sum_{j=1}^{m_0+1} r_j \overline{V}_j \right)^2 + O\left(T^{-1}\right) + o_{\text{a.s.}}(1) \\
&= \int_0^1 c(u,\, k)\, du + \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1+k}^{\lfloor T\lambda_j^0 \rfloor} \mu^2(t/T) - \left(\sum_{j=1}^{m_0+1} r_j \overline{V}_j \right)^2 \\
&\quad + O\left(T^{-1}\right) + o_{\text{a.s.}}(1),
\end{aligned}
$$

where we have used $\mathbb{E}(V_{t-k}) - \mathbb{E}(V_t) = O(k/T)$ by local stationarity in the third equality. Note that by ergodicity and an approximation to Riemann sums, we have

$$\begin{aligned}\sum_{j=1}^{m_0+1} r_j \overline{V}_j - \sum_{j=1}^{m_0+1} r_j \overline{\mu}_j &= \sum_{j=1}^{m_0+1} r_j \overline{V}_j - \sum_{j=1}^{m_0+1} r_j \mathbb{E}\left(\overline{V}_j\right) + \sum_{j=1}^{m_0+1} r_j \mathbb{E}\left(\overline{V}_j\right) - \sum_{j=1}^{m_0+1} r_j \overline{\mu}_j \\ &= o_{\text{a.s.}}(1) + O\left(T^{-1}\right). \end{aligned} \tag{B.1}$$

Basic manipulations show that

$$\begin{aligned}&\sum_{j_2 \neq j_1} r_{j_1} r_{j_2} \left(\overline{\mu}_{j_2} - \overline{\mu}_{j_1}\right)^2 \\ &\quad = \sum_{j_2 \neq j_1} r_{j_1} r_{j_2} \left(\overline{\mu}_{j_2}^2 + \overline{\mu}_{j_1}^2 - 2\overline{\mu}_{j_2}\overline{\mu}_{j_1}\right) \\ &\quad = \sum_{1 \leq j_2 \leq m_0+1} r_{j_2} \overline{\mu}_{j_2}^2 (1 - r_{j_2}) + \sum_{1 \leq j_1 \leq m_0+1} r_{j_1} \overline{\mu}_{j_1}^2 (1 - r_{j_1}) - 2 \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} \overline{\mu}_{j_2} \overline{\mu}_{j_1} \\ &\quad = 2 \sum_{1 \leq j \leq m_0+1} r_j \overline{\mu}_j^2 - 2 \sum_{1 \leq j \leq m_0+1} r_j^2 \overline{\mu}_j^2 - 2 \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} \overline{\mu}_{j_2} \overline{\mu}_{j_1}. \end{aligned} \tag{B.2}$$

Note that by Cauchy–Schwarz inequality,

$$(Tr_j - k) \sum_{t=\left\lfloor T\lambda_{j-1}^0 \right\rfloor + 1 + k}^{\left\lfloor T\lambda_j^0 \right\rfloor} \mu^2(t/T) \geq \left(\sum_{t=\left\lfloor T\lambda_{j-1}^0 \right\rfloor + 1 + k}^{\left\lfloor T\lambda_j^0 \right\rfloor} \mu(t/T) \right)^2. \tag{B.3}$$

Thus,

$$\begin{aligned}&\sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \sum_{t\left\lfloor T\lambda_{j-1}^0 \right\rfloor + 1 + k}^{\left\lfloor T\lambda_j^0 \right\rfloor} \mu^2(t/T) \\ &\quad = \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j(Tr_j - k)} (Tr_j - k) \sum_{t=\left\lfloor T\lambda_{j-1}^0 \right\rfloor + 1 + k}^{\left\lfloor T\lambda_j^0 \right\rfloor} \mu^2(t/T) \\ &\quad \geq \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j(Tr_j - k)} \left(\sum_{t=\left\lfloor T\lambda_{j-1}^0 \right\rfloor + 1 + k}^{\left\lfloor T\lambda_j^0 \right\rfloor} \mu(t/T) \right)^2 \end{aligned}$$

$$= \sum_{1 \leq j \leq m_0+1} r_j \overline{\mu}_j^2 + o(1). \tag{B.4}$$

Using (B.1)-(B.4) we have,

$$\begin{aligned}
\widehat{\Gamma}(k) &= \int_0^1 c(u,\, k)\, du + \sum_{j=1}^{m_0+1} r_j \frac{1}{T r_j} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor + 1 + k}^{\lfloor T\lambda_j^0 \rfloor} \mu^2(t/T) - \left(\sum_{j=1}^{m_0+1} r_j \overline{V}_j \right)^2 + o_{\text{a.s.}}(1) \\
&\geq \int_0^1 c(u,\, k)\, du + \sum_{j=1}^{m_0+1} r_j \overline{\mu}_{2,j} - \left(\sum_{j=1}^{m_0+1} r_j \overline{V}_j \right)^2 + O\left(T^{-1}\right) + o_{\text{a.s.}}(1) \\
&= \int_0^1 c(u,\, k)\, du + 2^{-1} \sum_{j_1 \neq j_2} r_{j_1} r_{j_2} \left(\overline{\mu}_{j_2} - \overline{\mu}_{j_1} \right)^2 + O\left(T^{-1}\right) + o_{\text{a.s.}}(1). \qquad \text{(B.5)}
\end{aligned}$$

The claim that $\widehat{\Gamma}(k) \geq d$ $\mathbb{P}$-a.s. as $k \to \infty$ follows from Assumption 2.2.2-(i) since this implies that $c(u,\, k) \to 0$ as $k \to \infty$ and from the fact that the second term on the right-hand side of (B.5) does not depend on k. If in addition it holds that $\mu_j(t/T) = \mu_j$ for $j = 1, \ldots, m_0 + 1$, then (B.3) holds with equality and the result follows as a special case of (B.5). □

B.2 Proof of Theorem 2.3.2

Lemma B.2.1 *Assume that $\{V_{t,T}\}$ satisfies Definition 2.2.1. Under Assumption 2.2.1-2.2.2 and 2.3.1-(ii),*

$$\begin{aligned}
&\sum_{j_1 \neq j_2} \frac{1}{T} \sum_{t=\lfloor T\lambda_{j_1-1}^0 \rfloor + 1}^{\lfloor T\lambda_{j_1}^0 \rfloor} \sum_{s=\lfloor T\lambda_{j_2-1}^0 \rfloor + 1}^{\lfloor T\lambda_{j_2}^0 \rfloor} \mathbb{E}\left(\left(V_t - \mu(t/T) \right) \left(V_s - \mu(s/T) \right) \right) \exp\left(-i\omega_l (t-s) \right) \\
&= o(1).
\end{aligned}$$

Proof. Let $\overline{r}_{j_1,j_2} = \max\{r_{j_1},\, r_{j_2}\}$ and $\underline{r}_{j_1,j_2} = \min\{r_{j_1},\, r_{j_2}\}$. We consider the case of adjacent regimes (i.e., $j_2 = j_1 + 1$) which also provides an upper bound for non-adjacent regimes due to the short memory property. For any $k = s - t =$

$1, \ldots, \left\lfloor T\underline{r}_{j_1,j_2} \right\rfloor$ there are k pairs in the above sum. The double sum above (over t and s) can be split into

$$T^{-1} \sum_{k=1}^{\lfloor CT^{\kappa} \rfloor} \left| \Gamma_{\{1:\lfloor CT^{\kappa} \rfloor\}} (\cdot,\, k) \right| + T^{-1} \sum_{k=\lfloor CT^{\kappa} \rfloor + 1}^{\lfloor hT \rfloor} \left| \Gamma_{\{\lfloor CT^{\kappa} \rfloor + 1:\lfloor hT \rfloor\}} (\cdot,\, k) \right| \tag{B.6}$$
$$+ T^{-1} \sum_{k=\lfloor hT \rfloor + 1}^{\left\lfloor T\underline{r}_{j_1,j_2} \right\rfloor - 1} \left| \Gamma_{\left\{\lfloor hT \rfloor + 1:\left\lfloor T\underline{r}_{j_1,j_2} \right\rfloor - 1\right\}} (\cdot,\, k) \right| + T^{-1} \sum_{k=\left\lfloor T\underline{r}_{j_1,j_2} \right\rfloor}^{\left\lfloor T\overline{r}_{j_1,j_2} \right\rfloor} \left| \Gamma_{\left\{\underline{r}_{j_1,j_2}:\overline{r}_{j_1,j_2}\right\}} (\cdot,\, k) \right|$$

where $C > 0$, $0 < h < 1$ with $\lfloor hT \rfloor < \left\lfloor T\underline{r}_{j_1,j_2} \right\rfloor - 1$, and $\Gamma_S (\cdot,\, k)$ is the sum of the autocovariances at lag k computed at the time points corresponding to $k \in S$. Note that the term $|\exp(-i\omega_l(\pm k))|$ can be bounded by some constant. The sums run over only $k > 0$ because by symmetry $\Gamma_u(k) = \Gamma_{u-k/T}(-k)$. Consider the first sum in (B.6). This is of order $O(T^{-1}T^{2\kappa})$ which goes to zero given $\kappa < 1/2$. The second sum is also negligible using the following arguments. By Assumption 2.3.1-(ii), $|\Gamma(u,\, k)| = C_{u,k}k^{-m}$ with $m > 2$ and choosing C large enough yields that the second sum of (B.6) converges to zero. In the third sum, the number of summands grows at rate $O(T)$ and for each lag k there are $O(T)$ autocovariances. However, by Assumption 2.3.1-(ii) each autocovariance is $O(T^{-m})$. Thus, the bound is $O(T^{-1}T^{2-m})$ which goes to zero as $T \to \infty$. The difference between the arguments used for the third sum and fourth sums is that now we do not have $O(T)$ autocovariances for each lag k. Thus, the bound for the fourth sum cannot be greater than the bound for the third sum. Thus, the fourth sum also converges to zero. □

Proof of Theorem 2.3.2. We have,

$$I_T(\omega_l) = \left| \frac{1}{\sqrt{T}} \sum_{j=1}^{m_0+1} \sum_{t=\left\lfloor T\lambda_{j-1}^0 \right\rfloor + 1}^{\left\lfloor T\lambda_j^0 \right\rfloor} \exp(-i\omega_l t) V_t \right|^2$$

$$= \left| \frac{1}{\sqrt{T}} \sum_{j=1}^{m_0+1} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} (X_t - \mu(t/T)) \exp(-i\omega_l t) \right.$$
$$\left. + \frac{1}{\sqrt{T}} \sum_{j=1}^{m_0+1} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} \mu(t/T) \exp(-i\omega_l t) \right|^2 .$$

From Assumption 2.3.1,

$$\left| \sum_{j=1}^{m_0+1} \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} \mu(t/T) \exp(-i\omega_l t) \right|^2$$
$$\geq \left| \sum_{j=1}^{m_0+1} B_j \sum_{t=\lfloor T\lambda_{j-1}^0 \rfloor+1}^{\lfloor T\lambda_j^0 \rfloor} \exp(-i\omega_l t) \right|^2$$
$$= \left| \sum_{j=1}^{m_0+1} B_j \exp\left(-i\omega_l \left(\lfloor T\lambda_{j-1}^0 \rfloor + 1\right)\right) \sum_{t=0}^{\lfloor T\lambda_j^0 \rfloor - \lfloor T\lambda_{j-1}^0 \rfloor - 1} \exp(-i\omega_l t) \right|^2$$
$$= \left| \frac{\exp(-i\omega_l)}{1-\exp(-i\omega_l)} \sum_{j=1}^{m_0+1} B_j \exp\left(-i\omega_l \left(\lfloor T\lambda_{j-1}^0 \rfloor\right)\right) \right|^2$$
$$\left|1 - \exp\left(-i\omega_l \left(\lfloor T\lambda_j^0 \rfloor - \lfloor T\lambda_{j-1}^0 \rfloor\right)\right)\right|^2$$
$$= \left| \frac{\exp(-i\omega_l)}{1-\exp(-i\omega_l)} \sum_{j=1}^{m_0+1} B_j \left(\exp\left(-i\omega_l \left(\lfloor T\lambda_{j-1}^0 \rfloor\right)\right) - \exp\left(-i\omega_l \lfloor T\lambda_j^0 \rfloor\right)\right) \right|^2 ,$$

using the formula for the first n-th terms of a geometric series $\sum_{k=0}^{n-1} ar^k = a\sum_{k=0}^{n-1} r^k = a(1-r^n)/(1-r)$. Then, using summation by parts,

$$\frac{\exp(-i\omega_j)}{1-\exp(-i\omega_j)} \sum_{j=1}^{m_0+1} B_j \left(\exp\left(-i\omega_l \left(\lfloor T\lambda_{j-1}^0 \rfloor\right)\right) - \exp\left(-i\omega_l \lfloor T\lambda_j^0 \rfloor\right)\right)$$
$$= \frac{\exp(-i\omega_j)}{1-\exp(-i\omega_j)} \left[B_1 - B_{m_0+1} - \sum_{j=1}^{m_0} (B_j - B_{j+1}) \exp\left(-i\omega_l \lfloor T\lambda_j^0 \rfloor\right) \right].$$

By Lemma B.2.1, it is sufficient to consider the cross-products within each regime j,

$$
\begin{aligned}
&\mathbb{E}\left(I_T\left(\omega_l\right)\right) \\
&\geq \sum_{j=1}^{m_0+1} r_j \frac{1}{Tr_j} \mathbb{E} \sum_{t=\left\lfloor T\lambda_{j-1}^0 \right\rfloor+1}^{\left\lfloor T\lambda_j^0 \right\rfloor} \sum_{s=\left\lfloor T\lambda_{j-1}^0 \right\rfloor+1}^{\left\lfloor T\lambda_j^0 \right\rfloor} \left(V_t - \mu\left(t/T\right)\right)\left(V_s - \mu\left(s/T\right)\right) \exp\left(-i\omega_l\left(t-s\right)\right) \\
&+ \sum_{j_1 \neq j_2} \sum \frac{1}{T} \mathbb{E} \sum_{t=\left\lfloor T\lambda_{j_1-1}^0 \right\rfloor+1}^{\left\lfloor T\lambda_{j_1}^0 \right\rfloor} \sum_{s=\left\lfloor T\lambda_{j_2-1}^0 \right\rfloor+1}^{\left\lfloor T\lambda_{j_2}^0 \right\rfloor} \left(V_t - \mu\left(t/T\right)\right)\left(V_s - \mu\left(s/T\right)\right) \exp\left(-i\omega_l\left(t-s\right)\right) \\
&+ \left| \frac{1}{\sqrt{T}} \frac{\exp\left(-i\omega_l\right)}{1-\exp\left(-i\omega_l\right)} \sum_{j=1}^{m_0+1} B_j \left(\exp\left(-i\omega_l\left(\left\lfloor T\lambda_{j-1}^0 \right\rfloor\right)\right) - \exp\left(-i\omega_l\left\lfloor T\lambda_j^0 \right\rfloor\right)\right) \right|^2 + o\left(1\right) \\
&= \sum_{j=1}^{m_0+1} \left(\mathbb{E} \frac{1}{T} \sum_{t=\left\lfloor T\lambda_{j-1}^0 \right\rfloor+1}^{\left\lfloor T\lambda_j^0 \right\rfloor} \left(V_t - \mu\left(t/T\right)\right)^2 \right. \\
&\left. + \frac{2}{Tr_j} \sum_{k=1}^{\lfloor Tr_j \rfloor - 1} \sum_{t=\left\lfloor T\lambda_{j-1}^0 \right\rfloor+k+1}^{\left\lfloor T\lambda_j^0 \right\rfloor} \Gamma_{t/T}\left(k\right) \exp\left(-i\omega_l k\right) \right) \\
&+ \left| \frac{1}{\sqrt{T}} \frac{\exp\left(-i\omega_l\right)}{1-\exp\left(-i\omega_l\right)} \sum_{j=1}^{m_0+1} B_j \left(\exp\left(-i\omega_l\left(\left\lfloor T\lambda_{j-1}^0 \right\rfloor\right)\right) - \exp\left(-i\omega_l\left\lfloor T\lambda_j^0 \right\rfloor\right)\right) \right|^2 + o\left(1\right).
\end{aligned}
$$

Next, using the definition of $f\left(u,\, \omega_l\right)$, $e^{-2i\omega_l} = 1$ by Euler's formula and letting $\omega_l \to 0$ we have,

$$
\begin{aligned}
&\mathbb{E}\left(I_T\left(\omega_l\right)\right) \\
&\geq \sum_{j=1}^{m_0+1} \left(\int_{\lambda_{j-1}^0}^{\lambda_j^0} c\left(u,\, 0\right) du + 2 \sum_{k=1}^{\infty} \int_{\lambda_{j-1}^0}^{\lambda_j^0} c\left(u,\, k\right) \exp\left(-i\omega_l k\right) du \right) \\
&+ \frac{1}{T} \frac{1}{\left|1-\exp\left(-i\omega_l\right)\right|^2} \left| \left[B_1 - B_{m_0+1} - \left(1+o\left(1\right)\right) \sum_{j=1}^{m_0} \left(B_j - B_{j+1}\right) \exp\left(-2\pi i l \lambda_j^0\right) \right] \right|^2 \\
&+ o\left(1\right)
\end{aligned}
$$

$$
\begin{aligned}
&= 2\pi \sum_{j=1}^{m_0+1} \int_{\lambda_{j-1}^0}^{\lambda_j^0} f\left(u,\, \omega_l\right) du \\
&\quad + \frac{1}{T} \frac{1}{\left|1 - \exp\left(-i\omega_l\right)\right|^2} \left| \left[B_1 - B_{m_0+1} - \left(1 + o\left(1\right)\right) \sum_{j=1}^{m_0} \left(B_j - B_{j+1}\right) \exp\left(-2\pi i l \lambda_j^0\right) \right] \right|^2 \\
&\quad + o\left(1\right) \\
&= 2\pi \int_0^1 f\left(u,\, \omega_l\right) du + \frac{1}{T\omega_l^2} \left| \left[B_1 - B_{m_0+1} - \sum_{j=1}^{m_0} \left(B_j - B_{j+1}\right) \exp\left(-2\pi i l \lambda_j^0\right) \right] \right|^2 + o\left(1\right).
\end{aligned}
\tag{B.7}
$$

By Assumption 2.2.1-(ii), the first term of (B.7) is bounded for all frequencies ω_j. Since $B_1, \ldots, B_{m_0+1}$ are fixed, if $T\omega_l^2 \to 0$ then the order of the second term of (B.7) is $O((T\omega_l^2)^{-1})$. Note that as $\omega_l \to 0$ there are some values of l for which the corresponding term involving $|\cdot|^2$ on the right-hand side of (B.7) is equal to zero [see the argument in Mikosch and Stărică (2004)]. In such a case, $\mathbb{E}\left(I_T\left(\omega_l\right)\right) \geq 2\pi \int_0^1 f\left(u,\, \omega_l\right) du > 0$. For the other values of $\{l\}$ as $\omega_l \to 0$, the second term of (B.7) diverges to infinity. The outcome is that there are frequencies close to $\omega_l = 0$ for which $\mathbb{E}\left(I_T\left(\omega_l\right)\right) \to \infty$. □

B.3 Proof of Theorem 2.3.3

We consider the case $k \geq 0$. The case $k < 0$ follows similarly. Consider any $u \in (0,\, 1)$ such that $T_j^0 \notin [Tu + k/2 - n_{2,T}/2 + 1, \ldots, Tu + n_{2,T}/2]$ for all $j = 1, \ldots, m_0$. Theorem S.B.3 in Casini (2019) showed that

$$
\mathbb{E}\left[\widehat{c}_T\left(u,\, k\right)\right] = c\left(u_0,\, k\right) + \frac{1}{2}\left(n_{2,T}/T\right)^2 \left[\frac{\partial^2}{\partial^2 u} c\left(u,\, k\right)\right] + o\left(\left(n_{2,T}/T\right)^2\right) + O\left(1/n_{2,T}\right).
\tag{B.8}
$$

Since $n_{2,T} \to \infty$ and $n_{2,T}/T \to 0$, $\mathbb{E}\left[\widehat{c}_T\left(u,\, k\right)\right] = c\left(u_0,\, k\right) + o\left(1\right)$. The same aforementioned theorem shows that $n_{2,T}\mathrm{Var}\left[\widehat{c}_T\left(u,\, k\right)\right] = O_{\mathbb{P}}\left(1\right)$. This combined with (B.8) yields part (i) of the theorem.

Next, we consider case (ii-a) with $(T_j^0 - (Tu + k/2 - n_{2,T}/2 + 1))/n_{2,T} \to \gamma \in (0,\, 1)$. We have,

$$
\begin{aligned}
&\widehat{c}_T\left(u,\, k\right) \\
&= n_{2,T}^{-1} \sum_{s=0}^{n_{2,T}} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1 - k} - \left(n_{2,T}^{-1} \sum_{s=0}^{n_{2,T}} V_{\lfloor Tu \rfloor - n_{2,T}/2 + s + 1} \right)^2 \\
&= n_{2,T}^{-1} \sum_{s=0}^{T_j^0 - \left(\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1\right)} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1 - k} \\
&\quad + n_{2,T}^{-1} \sum_{s = T_j^0 - \left(\lfloor Tu \rfloor + k/2 - n_{2,T}/2\right)}^{n_{2,T}} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1 - k} \\
&\quad - \left(n_{2,T}^{-1} \sum_{s=0}^{T_j^0 - \left(\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1\right)} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1} \right. \\
&\quad \left. + n_{2,T}^{-1} \sum_{s = T_j^0 - \left(\lfloor Tu \rfloor + k/2 - n_{2,T}/2\right)}^{n_{2,T}} V_{\lfloor Tu \rfloor - n_{2,T}/2 + s + 1} \right)^2 \\
&= n_{2,T}^{-1} \sum_{s=0}^{T_j^0 - \left(\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1\right)} \left(V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1 - k} \right. \\
&\quad \left. - \mathbb{E}\left(V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1}\right) \mathbb{E}\left(V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1 - k}\right) \right) \\
&\quad + n_{2,T}^{-1} \sum_{s = T_j^0 - \left(\lfloor Tu \rfloor + k/2 - n_{2,T}/2\right)}^{n_{2,T}} \left(V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1} V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1 - k} \right. \\
&\quad \left. - \mathbb{E}\left(V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1}\right) \mathbb{E}\left(V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1 - k}\right) \right) \\
&\quad + n_{2,T}^{-1} \sum_{s=0}^{T_j^0 - \left(\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + 1\right)} \mathbb{E}\left(V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1}\right) \mathbb{E}\left(V_{\lfloor Tu \rfloor + k/2 - n_{2,T}/2 + s + 1 - k}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+ n_{2,T}^{-1} \sum_{s=T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2\right)}^{n_{2,T}} \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right)\mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1-k}\right) \\
&- \left(n_{2,T}^{-1} \sum_{s=0}^{T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2+1\right)} V_{\lfloor Tu\rfloor-n_{2,T}/2+s+1}\right. \qquad \text{(B.9)} \\
&\left.+ n_{2,T}^{-1} \sum_{s=T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2\right)}^{n_{2,T}} V_{\lfloor Tu\rfloor-n_{2,T}/2+s+1}\right)^2 + o_{\mathbb{P}}(1) \\
&\geq \gamma c\left(\lambda_j^0,\, k\right) + (1-\gamma)\, c\left(u,\, k\right) + \gamma\mu_j\left(\lambda_j^0\right)^2 + (1-\gamma)\,\mu_{j+1}\left(u\right)^2 \\
&\quad - \left(\gamma\mu_j\left(\lambda_j^0\right) + (1-\gamma)\,\mu_{j+1}\left(u\right)\right)^2 + o_{\mathbb{P}}(1) \\
&= \gamma c\left(\lambda_j^0,\, k\right) + (1-\gamma)\, c\left(u,\, k\right) + \gamma\,(1-\gamma)\left(\mu_j\left(\lambda_j^0\right) - \mu_{j+1}\left(u\right)\right)^2 + o_{\mathbb{P}}(1). \qquad \text{(B.10)}
\end{aligned}
$$

Consider the case (ii-b) with $(T_j^0 - (\lfloor Tu\rfloor + k/2 - n_{2,T}/2 + 1))/n_{2,T} \to 0$. The other sub-case follows by symmetry. Eq. (B.9) continues to hold. The first term, third term and the first summation of the last term on the right-hand side of (B.9) are negligible. Thus, using ergodicity, implied by Assumption 2.2.1-2.2.2-(i),

$$
\begin{aligned}
\widehat{c}_T\left(u,\, k\right) &= c\left(u,\, k\right) + \\
&\qquad n_{2,T}^{-1} \sum_{s=T_j^0-\left(\lfloor Tu\rfloor+k/2-n_{2,T}/2\right)}^{n_{2,T}} \mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right)\mathbb{E}\left(V_{\lfloor Tu\rfloor+k/2-n_{2,T}/2+s+1}\right) \\
&\qquad - \mu\left(u\right)^2 + o_{\mathbb{P}}(1) \\
&= c\left(u,\, k\right) + \mu_{j+1}\left(u\right)^2 - \mu_{j+1}\left(u\right)^2 + o_{\mathbb{P}}(1) = c\left(u,\, k\right) + o_{\mathbb{P}}(1),
\end{aligned}
$$

where we have used the smoothness of $\mathbb{E}(X_t)$ implied by local stationarity. The second claim of the lemma follows from Assumption 2.2.2-(i) since this implies that $\sup_{u\in[0,\,1]} c\left(u,\, k\right) \to 0$ as $k \to \infty$ and the fact that the third term on the right-hand side of (B.10) does not depend on k. Thus, $\widehat{\Gamma}_{\mathrm{DK}}\left(k\right) \geq d_T^* + o_{\mathbb{P}}(1)$ where $d_T^* = (n_T/T)\,\gamma\,(1-\gamma)\left(\mu_j\left(\lambda_j^0\right) - \mu_{j+1}\left(u\right)\right)^2 > 0$ and $d_T^* \to 0$ since $n_T/T \to 0$. □

B.4 Proof of Theorem 2.3.4

Consider first any $u \in (0,\, 1)$ such that $T_j^0 \notin [\lfloor Tu \rfloor - n_T/2 + 1, \ldots, \lfloor Tu \rfloor + n_T/2]$ for all $j = 1, \ldots, m_0$. Theorem 3.3 in Casini and Perron (2020a) shows that

$$\mathbb{E}\left(I_{\mathrm{L},T}\left(u,\, \omega_l\right)\right) \tag{B.11}$$

$$= \left| \frac{1}{\sqrt{n_T}} \sum_{s=0}^{n_T-1} V_{\lfloor Tu \rfloor - n_T/2+s+1,T} \exp\left(-i\omega_l s\right) \right|^2$$

$$= f\left(u,\, \omega_l\right) + \frac{1}{6}\left(\frac{n_T}{T}\right)^2 \frac{\partial^2}{\partial u^2} f\left(u,\, \omega_l\right) + o\left(\left(\frac{n_T}{T}\right)^2\right) + O\left(\frac{\log\left(n_T\right)}{n_T}\right). \tag{B.12}$$

By Assumption 2.2.1 the absolute value of the first term on the right-hand side is bounded for all frequencies ω_l. By Assumption 2.3.2-(iii) $\left|\left(\partial^2/\partial u^2\right) f\left(u,\, \omega_l\right)\right|$ is bounded and, since $n_T/T \to 0$, the second term converges to zero. Similarly, the third and fourth terms are negligible. Thus, $\mathbb{E}\left(I_{\mathrm{L},T}\left(u,\, \omega_l\right)\right)$ is bounded below by $f\left(u,\, \omega_l\right) > 0$ as $\omega_l \to 0$ which establishes part (i). Now we consider to part (ii). We begin with case (a). We only focus on the sub-case $\left(T_j^0 - \left(\lfloor Tu \rfloor - n_T/2 + 1\right)\right)/n_T \to \gamma$ with $\gamma \in (0,\, 1)$. We have

$$\begin{aligned}
&I_{\mathrm{L},T}\left(\omega_l\right) \\
&= \left| \frac{1}{\sqrt{n_T}} \left(\sum_{s=0}^{T_j^0 - \left(\lfloor Tu \rfloor - n_T/2+1\right)} V_{\lfloor Tu \rfloor - n_T/2+s+1,T} \exp\left(-i\omega_l s\right) + \right.\right. \\
&\left.\left. \sum_{s=T_j^0 - \left(\lfloor Tu \rfloor - n_T/2\right)}^{n_T-1} V_{\lfloor Tu \rfloor - n_T/2+s+1,T} \exp\left(-i\omega_l s\right) \right) \right|^2 \\
&= \frac{1}{n_T} \left| \sum_{s=0}^{T_j^0 - \left(\lfloor Tu \rfloor - n_T/2+1\right)} \left(V_{\lfloor Tu \rfloor - n_T/2+s+1,T} - \mu\left(\left(\lfloor Tu \rfloor - n_T/2 + s + 1\right)/T\right)\right) \exp\left(-i\omega_l s\right) \right. \\
&\quad \left. + \sum_{s=T_j^0 - \left(\lfloor Tu \rfloor - n_T/2\right)}^{n_T-1} \left(V_{\lfloor Tu \rfloor - n_T/2+s+1,T} - \mu\left(\left(\lfloor Tu \rfloor - n_T/2 + s + 1\right)/T\right)\right) \exp\left(-i\omega_l s\right) \right.
\end{aligned}$$

$$+\sum_{s=0}^{n_T-1}\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\exp\left(-i\omega_l s\right)\Bigg|^2. \tag{B.13}$$

Using Assumption 2.3.2, we have

$$\left|\sum_{s=0}^{n_T-1}\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\exp\left(-i\omega_l s\right)\right|^2 \geq$$
$$\left|B_j\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\exp\left(-i\omega_l s\right)+B_{j+1}\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\exp\left(-i\omega_l s\right)\right|^2. \tag{B.14}$$

Note that

$$\begin{aligned}
&B_j\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\exp\left(-i\omega_l s\right)+B_{j+1}\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\exp\left(-i\omega_l s\right)\\
&\quad=B_j\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\exp\left(-i\omega_l s\right) \qquad\qquad (B.15)\\
&\qquad+B_{j+1}\exp\left(-i\omega_l\left(T_j^0-\left(\lfloor Tu\rfloor-n_T/2\right)\right)\right)\sum_{s=0}^{n_T-1-\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)}\exp\left(-i\omega_l s\right).
\end{aligned}$$

Focusing on the second term on the right-hand side above,

$$\begin{aligned}
&n_T^{-1}\left|B_{j+1}\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\exp\left(-i\omega_l s\right)\right|^2\\
&\quad=n_T^{-1}\left|B_{j+1}\exp\left(-i\omega_l\left(T_j^0-\left(\lfloor Tu\rfloor-n_T/2\right)\right)\right)\sum_{s=0}^{n_T-1-\left(T_j^0-(\lfloor Tu\rfloor-n_T/2)\right)}\exp\left(-i\omega_l s\right)\right|^2\\
&\quad=n_T^{-1}\Bigg|B_{j+1}\exp\left(-i\omega_l\left(T_j^0-\left(\lfloor Tu\rfloor-n_T/2\right)\right)\right)\\
&\qquad\frac{1-\exp\left(-i\omega_l\left(n_T-\left(T_j^0-\left(\lfloor Tu\rfloor-n_T/2\right)\right)\right)\right)}{1-\exp\left(-i\omega_l\right)}\Bigg|^2
\end{aligned}$$

$$= n_T^{-1} \left| B_{j+1} \frac{\exp\left(-i\omega_l \left(T_j^0 - (\lfloor Tu \rfloor - n_T/2)\right)\right) - \exp\left(-i\omega_l n_T\right)}{1 - \exp\left(-i\omega_l\right)} \right|^2 . \tag{B.16}$$

We show that the above equation diverges to infinity as $\omega_l \to 0$ with $n_T\omega_l^2 \to 0$. If $n_T\omega_l \to a \in (0, \infty)$ then $\mathrm{Re}\left(\exp\left(-i\omega_l n_T\right)\right) \neq 1$ and the order is determined by the denominator. As in the proof of Theorem 2.3.2, $|1 - \exp(-i\omega_l)|^2 = \omega_l^2$. Since $n_T\omega_l^2 \to 0$, the right-hand side above diverges. If $n_T\omega_l \to 0$, we apply L'Hôpital's rule to obtain

$$\begin{aligned}
& n_T^{-1} \left| B_{j+1} \frac{-i\left(T_j^0 - (\lfloor Tu \rfloor - n_T/2)\right) + in_T}{i} \right|^2 \\
& \quad = n_T^{-1} B_{j+1}^2 \left(-\left(T_j^0 - (\lfloor Tu \rfloor - n_T/2)\right)^2 + n_T^2 - \left(T_j^0 - (\lfloor Tu \rfloor - n_T/2)\right) n_T \right) \\
& \quad = O\left(n_T^2/n_T\right) = O\left(n_T\right),
\end{aligned}$$

which shows that the right-hand side of (B.16) diverges. A similar argument can be applied to the first term on the right-hand side of (B.15) and to the product of the latter term and the complex conjugate of the second term on the right-hand side of (B.16).

It remains to consider case (b) and the sub-case $T_j^0 - (\lfloor Tu \rfloor - n_T/2 + 1)/n_T \to 0$. The other sub-case follows by symmetry. We have (B.13) and (B.14). Note that,

$$\begin{aligned}
& \left| \frac{1}{\sqrt{n_T}} B_{j+1} \sum_{s=T_j^0 - (\lfloor Tu \rfloor - n_T/2)}^{n_T - 1} \exp\left(-i\omega_l s\right) \right|^2 \\
& \quad = \left| \frac{1}{\sqrt{n_T}} B_{j+1} \sum_{s=0}^{n_T - 1} \exp\left(-i\omega_l s\right) - \frac{1}{\sqrt{n_T}} B_{j+1} \sum_{s=0}^{T_j^0 - (\lfloor Tu \rfloor - n_T/2) - 1} \exp\left(-i\omega_l s\right) \right|^2 \\
& \quad = \left| -\frac{1}{\sqrt{n_T}} B_{j+1} \sum_{s=0}^{T_j^0 - (\lfloor Tu \rfloor - n_T/2) - 1} \exp\left(-i\omega_l s\right) \right|^2 \to 0.
\end{aligned}$$

Thus, we have

$$
\begin{aligned}
&\mathbb{E}\left(I_{\mathrm{LT}}\left(\omega_l\right)\right) \\
&= \frac{1}{n_T}\Bigg|\Bigg(\sum_{s=0}^{T_j^0-(\lfloor Tu\rfloor-n_T/2+1)}\Bigg(V_{\lfloor Tu\rfloor-n_T/2+s+1,T} \\
&-\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\Bigg)\exp\left(-i\omega_l s\right)\Bigg) \\
&+\sum_{s=T_j^0-(\lfloor Tu\rfloor-n_T/2)}^{n_T-1}\left(V_{\lfloor Tu\rfloor-n_T/2+s+1,T}-\mu\left(\left(\lfloor Tu\rfloor-n_T/2+s+1\right)/T\right)\right)\exp\left(-i\omega_l s\right)\Bigg|^2 \\
&+o\left(1\right).
\end{aligned}
$$

Note that the first sum above involves at most $C<\infty$ summands. So the first term is negligible. The expectation of the product of the first term and the conjugate of the second term is negligible by using arguments similar to the proof in Lemma B.2.1 with n_T in place of T. Thus, the limit of $\mathbb{E}\left(I_T\left(\omega_l\right)\right)$ is equal to the right-hand side of (B.12) plus additional $o\left(1\right)$ terms. □

References

Altissimo, F. and V. Corradi (2003) "Strong Rules for Detecting the Number of Breaks in a Time Series," *Journal of Econometrics*, Vol. 117, No. 207–244.

Altug, S. (1989) "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, Vol. 30, No. 4, pp. 889–920.

Amihud, Y. and C. M. Hurvich (2004) "Predictive Regressions: a Reduced-bias Estimation Method," *Journal of Financial and Quantitative Analysis*, Vol. 39, No. 4, pp. 813–841.

An, S. and F. Schorfheide (2007) "Bayesian Analysis of DSGE Models," *Econometric Reviews*, Vol. 26, No. 2-4, pp. 113–172.

Andrews, D. W. K. (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, Vol. 59, pp. 817–858.

——— (1993) "Tests for Parameter Instability and Structural Change with Unknown Change-Point," *Econometrica*, Vol. 61, No. 4, pp. 821–56.

Andrews, D. W. K. and J. C. Monahan (1992) "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, Vol. 60, No. 4, pp. 953–966.

Andrieu, C., A. Doucet, and R. Holenstein (2010) "Particle Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 72, No. 3, pp. 269–342.

Bai, J. and P. Perron (1998) "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, Vol. 66, No. 1, pp. 47–78.

Belotti, F., A. Casini, L. Catania, S. Grassi, and P. Perron (2021) "Simultaneous Bandwidths Determination for DK-HAC Estimators and Long-Run Variance Estimation in Nonparametric Settings."

Bhattacharya, R., V. Gupta, and E. Waymire (1983) "The Hurst Effect Under Trends," *Journal of Applied Probability*, Vol. 20, No. 3, pp. 649–662.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994) "ARCH models," *Handbook of Econometrics*, Vol. 4, pp. 2959–3038.

Cai, Z. (2007) "Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors," *Journal of Econometrics*, Vol. 136, No. 1, pp. 163–188.

Cai, Z. and Y. Wang (2014) "Testing Predictive Regression Models with Nonstationary Regressors," *Journal of Econometrics*, Vol. 178, pp. 4–14.

Campbell, J. Y. and R. J. Shiller (1988) "Stock prices, Earnings, and Expected Dividends," *The Journal of Finance*, Vol. 43, No. 3, pp. 661–676.

Campbell, J. Y. and M. Yogo (2006) "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, Vol. 81, No. 1, pp. 27–60.

Canova, F. (2014) "Bridging DSGE Models and the Raw Data," *Journal of Monetary Economics*, Vol. 67, pp. 1–15.

Casini, A. (2018) "Tests for Forecast Instability and Forecast Failure under a Continuous Record Asymptotic Framework," *arXiv preprint arXiv:1803.10883*.

——— (2019) "Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models," *Unpublished manuscript, Department of Economics and Finance, University of Rome Tor Vergata*.

Casini, A. and P. Perron (2019) "Structural Breaks in Time Series," in *Oxford Research Encyclopedia of Economics and Finance*: Oxford University Press.

——— (2020a) "Change-Point Analysis of Time Series with Evolutionary Spectra," *Unpublished manuscript, Department of Economics and Finance, University of Rome Tor Vergata*.

——— (2020b) "Continuous Record Asymptotics for Structural Change Models," *arXiv preprint arXiv:1803.10881*.

——— (2020c) "Continuous Record Laplace-based Inference in Structural Change Models," *Journal of Econometrics*, forthcoming.

——— (2020d) "Generalized Laplace Inference in Multiple Change-Points Models," *arXiv preprint arXiv:1803.10871*.

——— (2020e) "Minimax MSE Bounds and Nonparametric Nonlinear VAR Prewhitened for Double-Kernel HAC Estimation and for Long-Run Variance Estimation," *Unpublish manuscript, Department of Economics and Finance, University of Rome Tor Vergata*.

——— (2020) "Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models," *Unpublished Manuscript*.

Casini, A., T. Deng, and P. Perron (2020) "Theory of Low Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference," *Unpublished Manuscript*.

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995) "Inference in Models with Nearly Integrated Regressors," *Econometric theory*, Vol. 11, No. 5, pp. 1131–1147.

Chan, K. W. (2020) "Mean-Structure and Autocorrelation Consistent Covariance Matrix Estimation," *Journal of Business and Economic Statistics, forthcoming*.

Chang, S. Y. and P. Perron (2018) "A Comparison of Alternative Methods to Construct Confidence Intervals for the Estimate of a Break Date in Linear Regression Models," *Econometric Reviews*, Vol. 37, No. 6, pp. 577–601.

Chen, B. and Y. Hong (2012) "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression," *Econometrica*, Vol. 80, No. 3, pp. 1157–1183.

Christensen, B. J. and R. T. Varneskov (2017) "Medium Band Least Squares Estimation of Fractional Cointegration in the Presence of Low-frequency Contamination," *Journal of Econometrics*, Vol. 197, No. 2, pp. 218–244.

Clarida, R., J. Gali, and M. Gertler (2000) "Monetary Policy Rules and Macroeconomic Stability: Evidence and some Theory," *The Quarterly Journal of Economics*, Vol. 115, No. 1, pp. 147–180.

Crainiceanu, C. M. and T. J. Vogelsang (2007) "Nonmonotonic Power for Tests of a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation*, Vol. 77, No. 6, pp. 457–476.

Creal, D. (2012) "A Survey of Sequential Monte Carlo Methods for Economics and Finance," *Econometric Reviews*, Vol. 31, No. 3, pp. 245–296.

Dahlhaus, R. (1997) "Fitting Time Series Models to Nonstationary Processes," *Annals of Statistics*, Vol. 25, pp. 1–37.

Deng, A. and P. Perron (2006) "A Comparison of Alternative Asymptotic Frameworks to Analyse a Structural Change in a Linear Time Trend," *Econometrics Journal*, Vol. 9, No. 3, pp. 423–447.

Dickey, D. A. and W. A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American statistical association*, Vol. 74, No. 366a, pp. 427–431.

——— (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica: journal of the Econometric Society*, pp. 1057–1072.

Diebold, F. and A. Inoue (2001) "Long Memory and Regime Switching," *Journal of Econometrics*, Vol. 105, No. 1, pp. 131–159.

Diebold, F. and R. Mariano (1995) "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, Vol. 13, No. 3, pp. 253–63.

Dou, L. (2019) "Optimal HAR Inference," *Unpublished manuscript, Department of Economics, Princeton University*.

Doucet, A. and A. Johansen (2011) "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," *The Oxford Handbook of Nonlinear Filtering*, Vol. 24, pp. 656–704.

Elliott, G. and U. K. Müller (2007) "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics*, Vol. 141, No. 2, pp. 1196–1218.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996) "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, Vol. 64, No. 4, pp. 813–836.

Elliott, G. and J. H. Stock (1994) "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown," *Econometric theory*, Vol. 10, No. 3-4, pp. 672–700.

Engle, R. F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

Evans, G. and N. Savin (1981) "Testing for Unit Roots: 1," *Econometrica: Journal of the Econometric Society*, pp. 753–779.

——— (1984) "Testing for Unit Roots: 2," *Econometrica (pre-1986)*, Vol. 52, No. 5, p. 1241.

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007) "Estimating Macroeconomic Models: A likelihood Approach," *The Review of Economic Studies*, Vol. 74, No. 4, pp. 1059–1087.

Giacomini, R. and B. Rossi (2009) "Detecting and Predicting Forecast Breakdowns," *Review of Economic Studies*, Vol. 76, pp. 669–705.

——— (2010) "Forecast Comparisons in Unstable Environments," *Journal of Applied Econometrics*, Vol. 25, No. 4, pp. 595–620.

——— (2015) "Forecasting in Nonstationary Environments: What Works and What Doesn't in Reduced-Form and Structural Models," *Annual Review of Economics*, Vol. 7, pp. 207–229.

Giacomini, R. and H. White (2006) "Tests of Conditional Predictive Ability," *Econometrica*, Vol. 74, No. 6, pp. 1545–1578.

Gonçalves, S. and T. Vogelsang (2011) "Block Bootstrap HAC Robust Tests: The Sophistication of the Naïve Bootstrap," *Econometric Theory*, Vol. 27, No. 4, pp. 745–791.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993) "Novel Approach to Nonlinear/Non-gaussian Bayesian State Estimation," in *IEE Proceedings F (Radar and Signal Processing)*, Vol. 140, pp. 107–113, IET.

Granger, C. W. J. and N. Hyung (2004) "Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns," *Journal of Empirical Finance*, Vol. 11, pp. 399–421.

Guerron-Quintana, P. A. (2010) "What You Match Does Matter: The Effects of Data on DSGE Estimation," *Journal of Applied Econometrics*, Vol. 25, No. 5, pp. 774–804.

Hamilton, J. (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, Vol. 57, No. 2, pp. 357–384.

Herbst, E. and F. Schorfheide (2015) *Bayesian Estimation of DSGE models*: Princeton University Press.

——— (2019) "Tempered Particle Filtering," *Journal of Econometrics*, Vol. 210, No. 1, pp. 26–44.

Hillebrand, E. (2005) "Neglecting Parameter Changes in GARCH models," *Journal of Econometrics*, Vol. 129, No. 1, pp. 121–138.

Ibragimov, R. and U. K. Müller (2010) "t-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business & Economic Statistics*, Vol. 28, No. 4, pp. 453–468.

Ireland, P. N. (2004) "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, Vol. 28, No. 6, pp. 1205–1226.

Jansson, M. (2004) "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, Vol. 72, No. 3, pp. 937–946.

de Jong, R. and J. Davidson (2000) "Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices," *Econometrica*, Vol. 68, No. 2, pp. 407–423.

Juhl, T. and Z. Xiao (2009) "Testing for Changing Mean with Monotonic Power," *Journal of Econometrics*, Vol. 148, No. 1, pp. 14–24.

Kendall, M. G. (1954) "Note on Bias in the Estimation of Autocorrelation," *Biometrika*, Vol. 41, No. 3-4, pp. 403–404.

Kiefer, N. and T. J. Vogelsang (2002) "Heteroskedasticity-Autocorrelation Robust Standard Errors using the Bartlett Kernel without Truncation," *Econometrica*, Vol. 70, No. 5, pp. 2093–2095.

——— (2005) "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, Vol. 21, No. 6, pp. 1130–1164.

Kiefer, N., T. J. Vogelsang, and H. Bunzel (2000) "Simple Robust Testing of Regression Hypotheses," *Econometrica*, Vol. 69, No. 3, pp. 695–714.

Kim, D. and P. Perron (2009) "Assessing the Relative Power of Structural Break Tests Using a Framework Based on the Approximate Bahadur Slope," *Journal of Econometrics*, Vol. 149, No. 1, pp. 26–51.

Kim, S., N. Shephard, and S. Chib (1998) "Stochastic Volatility: Likelihood Inference and Comparison with ARCH models," *The Review of Economic Studies*, Vol. 65, No. 3, pp. 361–393.

Lamoureux, C. G. and W. D. Lastrapes (1990) "Persistence in Variance, Structural Change, and the GARCH," *Journal of Business and Economic Statistics*.

Lazarus, E., D. J. Lewis, and J. H. Stock (2017) "The Size-Power Tradeoff in HAR Inference," *Unpublished manuscript, Department of Economics, Harvard University*.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018) "HAR Inference: Recommendations for Practice," *Journal of Business and Economic Statistics*, Vol. 36, No. 4, pp. 541–559.

Lettau, M. and S. Van Nieuwerburgh (2008) "Reconciling the Return Predictability Evidence," *The Review of Financial Studies*, Vol. 21, No. 4, pp. 1607–1652.

Lewellen, J. (2004) "Predicting returns with financial ratios," *Journal of Financial Economics*, Vol. 74, No. 2, pp. 209–235.

Lindsay, B. G. (1988) "Composite Likelihood Methods," *Contemporary Mathematics*, Vol. 80, No. 1, pp. 221–239.

Lubik, T. A. and F. Schorfheide (2004) "Testing for Indeterminacy: An Application to US Monetary Policy," *American Economic Review*, Vol. 94, No. 1, pp. 190–217.

Mankiw, N. G. and M. D. Shapiro (1986) "Do We Reject Too Often?: Small Sample Properties of Tests of Rational Expectations Models," *Economics Letters*, Vol. 20, No. 2, pp. 139–145.

Martins, L. and P. Perron (2016) "Improved Tests for Forecast Comparisons in the Presence of Instabilities," *Journal of Time Series Analysis*, Vol. 37, No. 5, pp. 650–659.

McCloskey, A. and J. B. Hill (2017) "Parameter Estimation Robust to Low Frequency Contamination," *Journal of Business and Economic Statistics*, Vol. 35, No. 4, pp. 3598–610.

Mikosch, T. and C. Stărică (2004) "Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects," *Review of Economic and Statistics*, Vol. 86, No. 1, pp. 378–390.

Müller, U. K. (2007) "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, Vol. 141, No. 2, pp. 1331–1352.

——— (2014) "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business and Economic Statistics*, Vol. 32, No. 3, pp. 311–322.

Newey, W. K. and K. D. West (1987) "A Simple Positive Semidefinite, Heteroskedastic and Autocorrelation Consistent Covariance Matrix," *Econometrica*, Vol. 55, pp. 703–708.

——— (1994) "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, Vol. 61, No. 4, pp. 631–653.

Ng, S. (1996) "Looking for evidence of speculative stockholding in commodity markets," *Journal of Economic Dynamics and Control*, Vol. 20, No. 1-3, pp. 123–143.

Ng, S. and J. H. Wright (2013) "Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling," *Journal of Economic Literature*, Vol. 51, No. 4, pp. 1120–54.

Paye, B. S. and A. Timmermann (2006) "Instability of Return Prediction Models," *Journal of Empirical Finance*, Vol. 13, No. 3, pp. 274–315.

Perron, P. (1989) "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, Vol. 57, pp. 1361–1401.

——— (1991) "A Continuous Time Approximation to the Unstable First-order Autoregressive Process: the Case without an Intercept," *Econometrica*, Vol. 59, No. 1, pp. 211–236.

Perron, P. and Z. Qu (2010) "Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices," *Journal of Business and Economic Statistics*, Vol. 28, No. 2, pp. 275–290.

——— (1990) "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics*, Vol. 8, No. 2, pp. 153–162.

Perron, P. and Y. Yamamoto (2019) "Testing for Changes in Forecast Performance," *Journal of Business and Economic Statistics*, Vol. forthcoming.

Phillips, P. C. B. (1987) "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, Vol. 74, No. 3, pp. 535–547.

——— (2005) "HAC Estimation by Automated Regression," *Econometric Theory*, Vol. 21, No. 1, pp. 116–142.

Pitt, M. K. and N. Shephard (1999) "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, Vol. 94, No. 446, pp. 590–599.

Politis, D. (2011) "Higher-Order Accurate, Positive Semidefinite Estimation of Large-Sample Covariance and Spectral Density Matrices," *Econometric Theory*, Vol. 27, No. 4, pp. 703–744.

Pötscher, B. M. and D. Preinerstorfer (2018) "Controlling the Size of Autocorrelation Robust Test," *Journal of Econometrics*, Vol. 207, No. 2, pp. 406–431.

——— (2019) "Further Results on Size and Power of Heteroskedasticity and Autocorrelation Robust Tests, with an Application to Trend Testing," *Electronic Journal of Statistics*, Vol. 13, No. 2, pp. 3893–3942.

Preinerstorfer, D. and B. M. Pötscher (2016) "On Size and Power of Heteroskedasticity and Autocorrelation Robust Tests," *Econometric Theory*, Vol. 32, No. 2, pp. 261–358.

Qu, Z. (2007) "Searching for Cointegration in a Dynamic System," *The Econometrics Journal*, Vol. 10, No. 3, pp. 580–604.

——— (2018) "A Composite Likelihood Framework for Analyzing Singular DSGE Models," *Review of Economics and Statistics*, Vol. 100, No. 5, pp. 916–932.

Qu, Z. and P. Perron (2007) "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica*, Vol. 75, No. 2, pp. 459–502.

Qu, Z. and F. Zhuo (2021) "Likelihood Ratio-Based Tests for Markov Regime Switching," *The Review of Economic Studies*, Vol. 88, No. 2, pp. 937–968.

Rabanal, P. and J. F. Rubio-Ramírez (2005) "Comparing New Keynesian models of the Business cycle: A Bayesian approach," *Journal of Monetary Economics*, Vol. 52, No. 6, pp. 1151–1166.

Rapach, D. E. and M. E. Wohar (2006) "Structural Breaks and Predictive Regression Models of Aggregate US Stock Returns," *Journal of Financial Econometrics*, Vol. 4, No. 2, pp. 238–274.

Robinson, P. (1998) "Inference-Without Smoothing in the Presence of Nonparametric Autocorrelation," *Econometrica*.

Roueff, F. and R. von Sachs (2011) "Locally Stationary Long Memory Estimation," *Stochastic Processes and their Applications*, Vol. 121, No. 4, pp. 813–844.

Sargent, T. J. (1989) "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy*, Vol. 97, No. 2, pp. 251–287.

Satchell, S. E. (1984) "Approximation to the Finite Sample Distribution for Nonstable First Order Stochastic Difference Equations," *Econometrica (pre-1986)*, Vol. 52, No. 5, p. 1271.

Schorfheide, F. (2000) "Loss Function-based Evaluation of DSGE models," *Journal of Applied Econometrics*, Vol. 15, No. 6, pp. 645–670.

Shiller, R. J. (2016) *Irrational Exuberance: Revised and Expanded Third Edition*: Princeton University Press.

Sims, C. A. (2002) "Solving Linear Rational Expectations Models," *Computational Economics*, Vol. 20, No. 1-2, p. 1.

Smets, F. and R. Wouters (2007) "Shocks and Frictions in US Business Cycles: A Bayesian DSGE approach," *American Economic Review*, Vol. 97, No. 3, pp. 586–606.

Sobel, M. (1963) "Multivariate Hermite Polynomials, Gram-Charlier Expansions and Edgeworth Expansions,"Technical report, University of Minnesota.

Stambaugh, R. F. (1986) "Predictive Regressions," *Working Paper, University of Chicago*.

——— (1999) "Predictive Regressions," *Journal of Financial Economics*, Vol. 54, No. 3, pp. 375–421.

Stock, J. H. (1999) *A Class of Tests for Integration and Cointegration*, pp. 135–167, Oxford: Oxford University Press.

Stock, J. H. and M. W. Watson (1996a) "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, Vol. 14, No. 1, pp. 11–30.

——— (1996b) "Evidence on Structural Stability in Macroeconmic Time Series," *Journal of Business and Economic Statistics*, Vol. 14, pp. 11–30.

Sun, Y. (2013) "Heteroscedasticity and Autocorrelation Robust F Test using Orthonormal Series Variance Estimator," *Econometrics Journal*, Vol. 16, No. 1, pp. 1–26.

——— (2014a) "Fixed-Smoothing Asymptotics in a Two-Step GMM Framework," *Econmetrica*, Vol. 82, No. 6, pp. 2327–2370.

——— (2014b) "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference," *Journal of Econometrics*, Vol. 178, No. 3, pp. 659–677.

Sun, Y., P. Phillips, and S. Jin (2008) "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, Vol. 76, No. 1, pp. 175–194.

Velasco, C. and P. M. Robinson (2001) "Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean," *Econometric Theory*, No. 17, pp. 497–539.

Vogeslang, T. (1999) "Sources of Nonmonotonic Power when Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics*, Vol. 88, No. 2, pp. 283–299.

West, K. D. (1996) "Asymptotic Inference about Predictive Ability," *Econometrica*, Vol. 64, No. 5, pp. 1067–84.

White, J. S. (1958) "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *The Annals of Mathematical Statistics*, pp. 1188–1197.

——— (1961) "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient," *Biometrika*, Vol. 48, No. 1-2, pp. 85–94.

Zhang, X. and X. Shao (2013) "Fixed-Smoothing Asymptotics for Time Series," *Annals of Statistics*, Vol. 41, No. 3, pp. 1329–1349.

CURRICULUM VITAE