DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Consistent Inference for Predictive Regressions in Persistent VAR Economies

Torben G. Andersen and Rasmus T. Varneskov

CREATES Research Paper 2018-9

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Consistent Inference for Predictive Regressions in Persistent VAR Economies*

Torben G. Andersen[†] **Rasmus T. Varneskov**[‡]

February 20, 2018

Abstract

This paper studies the properties of standard predictive regressions in model economies, characterized through persistent vector autoregressive dynamics for the state variables and the associated series of interest. In particular, we consider a setting where all, or a subset, of the variables may be fractionally integrated, and note that this induces a spurious regression problem. We then propose a new inference and testing procedure – the local spectrum (LCM) approach – for the joint significance of the regressors, which is robust against the variables having different integration orders. The LCM procedure is based on (semi-)parametric fractional-filtering and band spectrum regression using a suitably selected set of frequency ordinates. We establish the asymptotic properties and explain how they differ from and extend existing procedures. Using these new inference and testing techniques, we explore the implications of assuming VAR dynamics in predictive regressions for the realized return variation. Standard least squares predictive regressions indicate that popular financial and macroeconomic variables carry valuable information about return volatility. In contrast, we find no significant evidence using our robust LCM procedure, indicating that prior conclusions may be premature. In fact, if anything, our results suggest the reverse causality, i.e., rising volatility predates adverse innovations to key macroeconomic variables. Simulations are employed to illustrate the relevance of the theoretical arguments for finite-sample inference.

Keywords: Endogeneity Bias, Fractional Integration, Frequency Domain Inference, Hypothesis Testing, Spurious Inference, Stochastic Volatility, VAR Models.

JEL classification: C13, C14, C32, C52, C53, G12

*We wish to thank Ivan Shaliastovich, Fabio Trojani, Daniela Osterrieder, George Tauchen along with participants at various seminars and conferences for helpful comments and suggestions. Note that some of this material in this manuscript supersede a former manuscript entitled "Inference in Intertemporal Asset Pricing Models with Stochastic Volatility and the Variance Risk Premium". Financial support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation, is gratefully acknowledged.

[†]Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; NBER, Cambridge, MA; and CREATES, Aarhus, Denmark; e-mail: t-andersen@northwestern.edu.

[‡]Department of Finance, Copenhagen Business School, 2000 Frederiksberg, Denmark; CREATES, Aarhus, Denmark; Multi Assets at Nordea Asset Management, Copenhagen, Denmark; e-mail: rtv.fi@cbs.dk.

1 Introduction and Literature Review

Regressions involving persistent variables have a long history in economics, econometrics and finance. In particular, as many time series display slowly decaying autocorrelations, non-stationarities or both, researchers have studied the properties of regressions that include such variables and identified a number of pitfalls. Most prominently, in the spurious regressions problem, e.g., Granger & Newbold (1974) and Phillips (1986), where one persistent unit root, or $I(1)$, process is projected onto another independent $I(1)$ process, standard significance tests display large size distortions. A second canonical example is the prediction of a noisy, and possibly persistent, variable using a highly persistent regressor. This setting is motivated by, among others, stock return or volatility predictability in financial economics along with associated empirical puzzles such as unstable predictive relations, e.g., Peseran & Timmermann (1995) and Welch & Goyal (2008), or predictive biases, e.g., Stambaugh (1999).

In this paper, we consider inference for predictive regressions within a VAR economy, where all included variables may be highly persistent, suggesting that standard inference is subject to problems akin to those mentioned above. Specifically, the variables may display different degrees of fractional integration, that is, be $I(d)$ processes, where d may take different non-integer values across series. This nests standard short memory and integrated VARs, if all variables have $d = 0$ or $d = 1$, respectively, but also accommodates many intermediate cases of varying degrees of fractional integration, thus comprising a very flexible setting. The variables are (asymptotically) stationary if $0 \leq d < 1/2$, and non-stationary if $d \geq 1/2$.

Our analysis builds on important prior contributions studying different aspects of the scenarios outlined above. First, for spurious regressions, the theory in Phillips (1986) is extended to the fractionally integrated case by Tsay & Chung (2000), demonstrating how many basic insights carry over to regressions with independent, fractionally integrated variables. Second, Ferson, Sarkissian & Simin (2003), Valkanov (2003), Torous, Valkanov & Yan (2005) and Deng (2014) document that similar issues arise when predicting a noisy stationary variable with a persistent regressor in a local-to-unity (unit root) setting and, in particular, for long-horizon return regressions. In the latter case, the inference may also be distorted by a bias caused by correlation between innovations to the (stationary) returns and the persistent predictor. Stambaugh (1999) shows that this bias can be corrected for a stationary predictor, but Phillips & Lee (2013) and Phillips (2014) demonstrate that this, more generally, is infeasible, when the regressor displays local-to-unity, unit-root or explosive persistence. In summary, it is widely acknowledged that standard inference techniques encounter serious size problems, when one or more variables of the system are strongly persistent.

Several alternative procedures have been developed to accommodate spurious inference problems and issues with predictive biases. For example, standard and fractional cointegration frameworks facilitate inference on general linear relations, whose error is purged of (some of) the persistence of the original processes, see, e.g., Johansen & Nielsen (2012) for modeling and parametric inference, and Robinson & Marinucci (2003) and Christensen & Nielsen (2006) for the semiparametric case. Moreover, various robust inference procedures have been proposed for predictive regressions when the persistent regressor follow local-to-unity dynamics, e.g., using Bonferroni corrections in Cavanagh, Elliott & Stock (1995) and Campbell & Yogo (2006), a conditional likelihood approach in Jansson & Moreira (2006), and a class of nearly optimal tests in Elliott, Müller & Watson (2015). Recent contributions such as Phillips & Lee (2013) and Phillips (2014) discuss drawbacks of these approaches, e.g., the lack of power of the Bonferroni procedure, if the persistence of the regressor is actually stationary, not local-to-unity, and the lack of extendability to multivariate regressions. As an alternative, and in conjunction with Magdalinos & Phillips (2009), Kostakis, Magdalinos & Stamatogiannis (2015) and Phillips & Lee (2016), they develop the IVX methodology. It applies generally to persistent autoregressive processes (stationary, local-to-unity, unity and explosive) as well as to multivariate testing problems. Finally, Sizova (2013) consider predictive regressions for stock returns using stationary fractionally integrated variables ($I(d)$ processes with $0 < d < 1/2$), including historical volatility, thus analyzing the same fundamental problem, but from a different asymptotic perspective.

Despite the fact that inference in systems with persistent variables has received considerable attention, there are presently no results pertaining to our flexible predictive regression framework, where each variable may display fractional integration of different orders, covering stationary and non-stationary values. In the cointegration literature, the underlying assumption of (fractional) cointegration is violated under the null hypothesis of no predictability. Moreover, in the stationary-persistent variable prediction literature, such as the IVX methodology, the regressand is only weakly dependent, i.e., not fractionally integrated. Finally, the setting with diverse values of d puts us squarely outside the frequently-applied local-to-unity framework.

In this paper, we fill an important gap in the literature by developing inference and testing procedures for prediction of an $I(d)$ variable within a system of fractionally integrated variables of potentially different orders. Our methodology is robust to stationary and non-stationary variables. Such uniformity across persistence regimes is highly desirable, allowing us to retain statical power, while letting the user, a priori, remain agnostic about the stationarity of the system. Specifically, we propose a two-step inference procedure – the local spectrum (LCM) approach. The first step uses a (semi-)parametric fractional filter to purge the variables of their long memory, while retaining the coherence among the variables in the filtered series. The second step relies on band-spectrum regressions, using carefully selected frequency ordinates to account for any slippage from the mean (or initial value, if $d \geq 1$) and first-stage filtering errors. We establish the asymptotic properties, showing that, the LCM inference is asymptotically Gaussian and the associated test statistic for joint significance is χ^2-distributed, and thus readily implementable. These results apply in spite of the original variables consisting of a mixture of stationary and non-stationary variables. Furthermore, in settings where predictive biases occur in the fractionally integrated VAR system, we show that the LCM procedure, at most, incurs a second-order impact, regardless of the persistence in the system. This contrasts Phillips & Lee (2013), who find the corresponding bias to be of first order and uncorrectable in the local-to-unity case.

The methodological contributions closest to ours are those by Shao (2009), Maynard, Smallwood & Wohar (2013), Christensen & Varneskov (2017), and Müller & Watson (2017). In Shao (2009), the test for independence of two fractionally integrated processes is bivariate, and it neither accommodate non-zero means (or initial values) of the series nor non-stationary long memory ($d \geq 1/2$). The two-stage approach in Maynard et al. (2013) has similar drawbacks. In addition, their asymptotic distribution theory differs under the null (no predictability) and alternative hypothesis, and it depends on the first-stage filtering. Müller & Watson (2017) rely on the estimation of a bivariate spectral density, whose confidence intervals are computed numerically, while LCM approach delivers closed-form inference. Moreover, the former methodology is not readily extendable to the multivariate case, and is designed only to extract information about the "long-run" coherence between two series from a fixed number of frequencies in the vicinity of the origin.[1] In contrast, the LCM procedure captures information about the coherence from a wider range of the spectrum, allowing for more general statements about predictability. Finally, Christensen & Varneskov (2017) propose a band-spectrum regression estimator in a stationary fractional cointegration setting, resembling the one used in the second-step LCM analysis. However, their results pertain only to the $0 < d < 1/2$ case. In fact, a direct application of their medium-band least squares estimator generally generates inconsistent estimates of predictive power. Hence, both steps of our LCM approach are crucial for reliable inference and testing.

We apply the LCM approach to study predictive regressions for realized volatility. Going back to Schwert (1989), it has been debated whether financial and macroeconomic variables aid in the prediction of aggregate stock market volatility. Recently, interest in the topic has surged with the adoption of stochastic volatility in macro-finance models for the purpose of explaining consumption innovations and cross-sectional asset pricing, e.g., Bansal, Kiku, Shaliastovich & Yaron (2014) and Campbell, Giglio, Polk & Turley (2017). Hence, recent studies explore alternative ways of improving realized volatility forecasts using various state variables, e.g., Christiansen, Schmeling & Schrimpf (2012), Paye (2012), Conrad & Loch (2014), Mittnik, Robinzonov & Spindler (2015), Dew-Becker, Giglio, Le & Rodriguez (2017), and Nonejad (2017). Although some studies use sophisticated econometric techniques, many still rely on standard regressions to discern whether financial and macroeconomic variables boost the predictability of return volatility. Moreover, none of the methods account suitably for the persistence in the state variables, nor do they explicitly account for long memory in the realized market variance. The latter is particularly problematic, as this feature is a stylized fact within the financial econometrics literature.[2] In fact, Paye (2012, p. 533) and Nonejad (2017, p. 135) recognize that the state variables are very persistent, having first-order autocorrelations of the magnitude that generate size distortions according to Ferson et al. (2003, 2009), so alternative inference procedures may well be warranted.

For specificity, we focus on three state variables – the default spread, three-month U.S. treasury bills, and price-earnings ratio – whose prowess for realized variance forecasting is highlighted by, among others, Campbell et al. (2017). We first document that all involved variables may be characterized as fractionally integrated processes; realized variance as a persistent stationary process, the state

[1] A similar methodology is used to construct long-run predictive sets in Müller & Watson (2016).

[2] See, for example, Baillie, Bollerslev & Mikkelsen (1996), Comte & Renault (1998), Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001, 2003), Bandi & Perron (2006), Christensen & Nielsen (2006), Andersen, Bollerslev & Diebold (2007), Corsi (2009), and Bollerslev, Osterrieder, Sizova & Tauchen (2013).

variables as non-stationary ones. Second, we confirm evidence from the extant literature, showing that all three state variables are, seemingly, significant predictors of realized variance based on least squares and HAC inference. Third, we demonstrate through realistically calibrated simulations that standard least squares inference procedures suffer from large size distortions, when the variables in the VAR system are persistent in the sense of fractional integration. This complements the comprehensive simulation study in Ferson et al. (2003), demonstrating that size distortions can be very severe, up to 70% for joint significance tests, in our general setting. Our LCM test for joint significance, on the other hand, has excellent size and power in finite samples. Fourth, when testing the predictive ability of the state variables using our robust LCM procedure, we fail to find significant predictive power. Finally, we test for a reversal of the predictive relation. That is, we explore whether the realized stock market variance is informative about future realizations of the state variables. Indeed, for the given sample, elevated volatility serves as a strong predictor for a widening the default spread and a drop in the price-earnings ratio. We conclude that standard least squares techniques are likely to generate spurious results regarding realized variance predictions, while the LCM inference procedure may help uncover new predictive relations, with direct implications for current research themes in both the macroeconomic and finance areas.

The paper proceeds as follows. Section 2 provides the setup and identifies problems associated with standard predictive regressions. The local spectrum (LCM) approach is introduced in Section 3. Section 4 describes the data and establishes baseline least squares evidence, and Section 5 contains the simulation study. Section 6 explores realized variance prediction using the LCM procedure, and Section 7 provides robust checks and considers the reverse predictive relations. Finally, Section 8 concludes, and the Appendix contains assumptions, proofs, data details and additional theory.

2 Predictive Regressions with Persistent Variables

This section presents the regression framework, that we use throughout to study the predictive power of regressors in persistent VAR systems, where all variables may exhibit long memory of different orders. Inspired by our findings, we discuss the possibility of drawing spurious inference in such systems, which further motivates our design of the new LCM testing procedure in Section 3.

2.1 Predictive Regressions

We observe a $(\mathcal{K}+1) \times 1$ vector $\boldsymbol{Z}_t = (y_t, \boldsymbol{X}_t')'$ at times $t = 1, \ldots, n$, where $\boldsymbol{X}_t$ may include lagged values of y_t. Our main interest is to draw inference on the $\mathcal{K} \times 1$ vector $\boldsymbol{B}$ from the regression,

$$y_t \;=\; a + \boldsymbol{B}'\boldsymbol{X}_{t-1} + u_t\,, \tag{1}$$

under mild, general assumptions on $\boldsymbol{Z}_t$ and u_t, including, importantly, allowing for the variables to display varying degrees of persistence and flexible dynamics. If we impose short memory ARMA dynamics on $\boldsymbol{Z}_t$, with sufficiently well-behaved persistence parameters, then inference on $\boldsymbol{B}$ is standard.[3] On the other hand, if some variables display higher degrees of persistence, inference becomes more complex and subtle. As detailed below, not only will standard least squares generally fail, but the risk of, spuriously, detecting predictive power grows, almost regardless of how the persistence is modeled.

As noted, the vector $\boldsymbol{X}_{t-1}$ may contain the history of y_t up to and including time $t-1$, e.g., a finite order of its lags. Letting $\boldsymbol{x}_{t-1}$ denote the $k \times 1$ subvector of $\boldsymbol{X}_{t-1}$ that excludes past values of y_t, we also consider regressions of the form,

$$\widetilde{y}_t \;=\; \alpha + \boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \tilde{u}_t, \tag{2}$$

where $\widetilde{y}_t$ has been *pre-whitened* using its own historical values. By the Frisch-Waugh theorem, if pre-whitening is performed exploiting the exact same set of historical values of y_t included in $\boldsymbol{X}_{t-1}$, then testing for the joint significance of $\boldsymbol{x}_{t-1}$ using either equation (1) or (2) is equivalent. This result motivates the design of our robust spectrum inference procedure below.

We primarily focus on testing for significance of the external predictors $\boldsymbol{x}_{t-1}$, using regressions of the form (1) or (2), in cases where $\boldsymbol{x}_{t-1}$ as well as the dependent variable y_t may display dynamics incompatible with short-memory ARMA processes. Specifically, the variables in the system are allowed to be fractionally integrated processes, $I(d)$, whose integration order, d, may differ from variable to variable, nesting short memory and unit root processes as special cases for $d = 0$ and $d = 1$, respectively. As mentioned in the introduction, several studies develop robust inference for (long-horizon) asset return regressions, where the regressors are local-to-unity processes. Our setting differs not only because we consider an alternative description of persistence – fractional integration – but also because we face the additional challenge that the entire VAR system, including the dependent variable, may be persistent to different degrees. We stress that this is, in fact, the typical case in applications with multiple financial or macroeconomic regressors.

2.2 Fractional Integration and Spurious Regressions

To set the stage, we assume $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t')'$ obeys the following general dynamics,

$$\boldsymbol{D}(L)(\boldsymbol{z}_t - \boldsymbol{\mu}) \;=\; \boldsymbol{v}_t\, \mathbf{1}_{\{t \geq 1\}} \tag{3}$$

where $\boldsymbol{\mu}$ is a $(k+1) \times 1$ vector of nonrandom unknown finite numbers, either the means or initial values of the vector process, $\boldsymbol{D}(L) = \text{diag}[(1-L)^{d_1}, \ldots, (1-L)^{d_{k+1}}]$, with $(1-L)^d$ being a generic fractional filter, defined as,

$$(1-L)^d \;=\; \sum_{i=0}^{\infty} \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}\, L^i, \tag{4}$$

[3] Regular strong mixing conditions on $\boldsymbol{Z}_t$ and mild summability conditions on its moments imply that HAC inference procedures, as developed in Newey & West (1987) and Andrews (1991), are readily amenable for least squares.

and $\Gamma(\cdot)$ is the gamma function.[4] Moreover, letting "$\sim$" signify that the ratio of the left- and right-hand-side tends to one in the limit, element-wise, we assume that the fractionally differenced process in equation (3), $\boldsymbol{v}_t$, is covariance stationary with a spectral density satisfying,

$$\boldsymbol{f}_{vv}(\lambda) \sim \boldsymbol{G}_{vv}, \qquad \text{as} \quad \lambda \to 0^+, \tag{5}$$

and that it admits a Wold representation,

$$\boldsymbol{v}_t \;=\; \sum_{j=0}^{\infty} \boldsymbol{A}_j \boldsymbol{\epsilon}_{t-j}, \qquad \sum_{j=0}^{\infty} \|\boldsymbol{A}_j\|^2 < \infty, \tag{6}$$

with innovations $\boldsymbol{\epsilon}_t$ satisfying $\mathbb{E}[\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\mathcal{F}_{t-1}] = \boldsymbol{\Sigma}$, where $\mathcal{F}_t = \sigma(\boldsymbol{\epsilon}_s, s \leq t)$ is the σ-field generated by the innovations $\boldsymbol{\epsilon}_s$, $s \leq t$. Assumptions (3)-(6) are quite general. They only parameterize the frequency domain behavior of the vector $\boldsymbol{z}_t$ in a local neighborhood of the origin, $\lambda \to 0^+$, leaving the "higher-frequency" dynamics unspecified. For example, they allow $\boldsymbol{z}_t$ to obey a vector fractional ARIMA process, thus nesting commonly adopted VAR dynamics as a special case with $d_1 = \cdots = d_{k+1} = 0$.[5] Moreover, beyond accommodating different degrees of persistence among the variables, conditions (3)-(6) allow for elaborate lag- and predictive structures.[6]

The properties of the individual variables in $\boldsymbol{z}_t$ depend on the respective integration orders, d_i for $i = 1, \ldots, k+1$. First, if $0 \leq d_i < 1/2$, the ith variable is stationary with long memory, whenever $d > 0$. As conveyed by equations (3) and (5), long memory processes feature hyperbolically decaying auto-covariances, contrary to the geometric decay of short memory processes ($d = 0$). Second, if $d_i \geq 1/2$, the variable is non-stationary, but it possesses a well-defined mean, if $d_i < 1$. Finally, if $-1/2 < d_i < 0$, the ith variable is anti-persistent, a typical characteristic of an over-differenced process. In addition to these univariate time series characteristics, the memory properties of the variables in $\boldsymbol{z}_t$ have important implications for inference and interpretation of the predictive regressions (1) and (2). Specifically, if y_t and $\boldsymbol{x}_{t-1}$ have strictly positive integration orders, standard regression techniques typically lead to spurious inference.

To highlight the spurious inference concerns, consider the simplest case (1), where $\boldsymbol{x}_{t-1}$ is univariate (written x_{t-1}), fractionally integrated of order $I(d_x)$, independent of $y_t \in I(d_y)$, both have zero means, and we run regressions of the form,

$$y_t \;=\; \alpha \,+\, \beta x_{t-1} \,+\, u_t\,. \tag{7}$$

[4] The assumptions stated in the main text are mostly expositional and will need to be strengthened to derive the asymptotic results. We state the assumptions formally in Appendix A, noting that these are standard in the literature.

[5] First-order VAR dynamics is used in, e.g., Campbell & Vuolteenaho (2004), Bansal et al. (2014), Campbell et al. (2017), many of the references therein, and the textbook by Campbell & Viceira (2002).

[6] Note that assumptions similar to those imposed in equations (3)-(6), see also Appendix A, often are used to study fractional cointegration, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006), and Christensen & Varneskov (2017), or for the estimation of multivariate fractional models, e.g., Shimotsu (2007) and Nielsen (2015).

Then, OLS estimation results for equation (7) will be nonstandard, with properties that reflect the relative integration orders d_y and d_x. Specifically, let t_α and t_β denote the usual t-statistics associated with the OLS estimates $\hat{\alpha}$ and $\hat{\beta}$, respectively, $\mathcal{DW}$ denote the Durbin-Watson test for residual autocorrelation, and $\rho_y(1)$ be the first-order autocorrelation of y_t, then by combining results from Theorems 1, 2 and 4 of Tsay & Chung (2000), we obtain,[7]

$$\begin{cases} (\mathbf{a}) \text{ if } d_y \in (0,1/2) \text{ and } d_x \in (1/2,3/2): \ \hat{\alpha} = O_p(n^{d_y-1/2}), \quad \hat{\beta} = O_p(n^{d_y-d_x}), \\ \qquad t_\alpha = O_p(n^{d_y}), \ t_\beta = O_p(n^{d_y}), \ R^2 = O_p(n^{2d_y-1}), \text{ and } \mathcal{DW} \xrightarrow{\mathbb{P}} 2(1-\rho_y(1)). \\ (\mathbf{b}) \text{ if } d_y \in (1/2,3/2) \text{ and } d_x \in (1/2,3/2): \ \hat{\alpha} = O_p(n^{d_y-1/2}), \quad \hat{\beta} = O_p(n^{d_y-d_x}), \\ \qquad t_\alpha = O_p(n^{1/2}), \ t_\beta = O_p(n^{1/2}), \ R^2 = O_p(1), \text{ and } \mathcal{DW} \xrightarrow{\mathbb{P}} 0. \\ (\mathbf{c}) \text{ if } d_y, d_x \in (0,1/2) \text{ and } d_y + d_x > 1/2: \ \hat{\alpha} = O_p(n^{d_y-1/2}), \quad \hat{\beta} = O_p(n^{d_y+d_x-1}), \\ \qquad t_\alpha = O_p(n^{d_y}), \ t_\beta = O_p(n^{d_y+d_x-1/2}), \ R^2 = O_p(n^{2(d_y+d_x-1)}), \text{ and } \mathcal{DW} \xrightarrow{\mathbb{P}} 2(1-\rho_y(1)). \end{cases} \tag{8}$$

The results in equation (8) are quite alarming. If the predictive power of non-stationary variables ($d_i > 1/2$) is to be tested *and* the regressand possesses long memory, $d_y > 0$, then standard least squares techniques fail. They deliver inconsistent t-statistics, and the associated R^2 measures are biased, Case **(a)**, or inconsistent, Case **(b)**. Moreover, inconsistency also occurs when both variables are stationary, yet sufficiently persistent to generate the collective non-stationary memory of Case **(c)**.[8] Furthermore, the results for the Durbin-Watson test statistic illustrate that residual-based specification tests may be misleading. Finally, we stress that the rate of (in)consistency of $\hat{\alpha}$ and $\hat{\beta}$ depends on the, a priori unknown, memory of y_t and x_{t-1}. In particular, if $d_y \in (1/2, 3/2)$, Case **(b)** shows that $\hat{\alpha}$ always is inconsistent, while $\hat{\beta}$ is either inconsistent, if $d_y \geq d_x$, or subject to very slow convergence if, for example, $d_y = d_x - \epsilon$ for some small $\epsilon > 0$. When y_t is stationary, Cases **(a)** and **(c)**, the parameter estimates are consistent, albeit at a slow, memory-dependent rate.

For multivariate regressions with independent, but persistent regressors, equation (8) in conjunction with Chung (2002, Theorem 1) suggest that all parameters, depending on the memory properties of the corresponding regressors, will display different rates of convergence and feature inconsistent t-statistics, while the regression R^2 will be biased or inconsistent. In the context of our empirical example below, these results and the compelling evidence of long memory in the realized return variation (see the introduction and Table 1) imply that we, almost inevitably, face a spurious inference problem, when using standard least squares methods to forecast volatility based on popular financial and macroeconomic variables, as the latter, as shown in Section 4, typically are well-characterized as

[7] Note that Tsay & Chung (2000) derive their asymptotic results under a Type I model of fractional integration, unlike the Type II model in (3), and they restrict the short memory dynamics to be i.i.d. The latter, however, may easily be relaxed following the results of Chung (2002). Moreover, the differences between Type I and II models of fractional integration are detailed in, e.g., Shimotsu & Phillips (2006). For ease of exposition in the main text, we will abstain from distinguishing between Type I and II processes and refer to specific papers for details.

[8] If, on the other hand, the variables are stationary and persistent with collective memory $d_y + d_x < 1/2$, Chung (2002, Corollaries 2 and 3) show that β is amenable to standard inference, but inference for α remains non-standard.

non-stationary long memory processes.[9] Moreover, these concerns extend to other economic relations, like the predictive modeling of, e.g., purchasing power parity, exchange rate dynamics, and bond yields, for which, among others, Cheung & Lai (1993), Baillie & Bollerslev (1994) and Duecker & Startz (1998), respectively, document long-memory persistence.

In sum, the concerns about least squares inference raised by equation (8) highlight the need for the development of a new reliable methodology for tests regarding the (joint) statistical significance of predictor variables in diverse, persistent economic systems. We turn towards that task next.

3 The Local Spectrum Approach

This section introduces the local spectrum (LCM) inference and testing procedure and establishes its asymptotic properties. First, we motivate our approach using the spectral density decomposition of $\boldsymbol{z}_t$, before describing its two-step implementation in Sections 3.2 and 3.3. Section 3.4 provides the asymptotic theory results. Finally, Section 3.5 explores robustness to the regressor endogeneity bias.

3.1 Motivation

The intuition behind the local spectrum inference and testing procedure is best conveyed by considering the spectral density of $\boldsymbol{z}_t$, which is given by,

$$\boldsymbol{f}_{zz}(\lambda) \sim \boldsymbol{\Lambda}\, \boldsymbol{G}_{vv}\, \bar{\boldsymbol{\Lambda}}, \qquad \lambda \to 0^+, \tag{9}$$

where $\bar{\boldsymbol{\Lambda}}$ is the complex conjugate of the $(k+1) \times (k+1)$ matrix $\boldsymbol{\Lambda}$, defined as,

$$\boldsymbol{\Lambda} \;=\; \text{diag}\left[(1 - e^{\mathrm{i}\lambda})^{d_1}, \ldots, (1 - e^{\mathrm{i}\lambda})^{d_{k+1}}\right],$$

and $\mathrm{i} = \sqrt{-1}$. We rely on an exact spectral density representation around the origin in equation (9), that is, the entries of $\boldsymbol{\Lambda}$ are of the form $(1 - e^{\mathrm{i}\lambda})^d$ rather than the usual approximation λ^{-d}. Not only does this accommodate richer dynamics of $\boldsymbol{z}_t$, c.f. Shimotsu (2007) and Robinson (2008), it also allows the LCM inference and testing procedure to apply over a wider range of d, covering both stationary and non-stationary values. In particular, as discussed by Shimotsu & Phillips (2005), λ^{-d} provides a good approximation, when $d < 1/2$, but it deteriorates when d belongs to the non-stationary range, eventually generating inconsistent estimators of the integration order.

Importantly, equation (9) highlights that the co-dependence structure among the variables of the VAR system, including predictive relations, depend on whether the off-diagonal elements of $\boldsymbol{G}_{vv}$ are non-zero, not on their respective long memory properties, which are captured by the diagonal matrix $\boldsymbol{\Lambda}$. Hence, writing e_t for the first element of $\boldsymbol{v}_t$ and $\boldsymbol{u}_t$ for the remaining $k \times 1$ vector, corresponding to the innovations of the predictors $\boldsymbol{x}_t$, such that $\boldsymbol{v}_t = (e_t, \boldsymbol{u}_t')'$, then testing significance of $\boldsymbol{x}_{t-1}$ for y_t

[9] We remark that the inference problems in equation (8) are generic and generally will **not** be alleviated by applying other least squared based estimation procedures such as WLS instead of OLS.

is equivalent to testing significance in the latent regression relation,

$$e_t \;=\; \boldsymbol{\mathcal{B}}'\boldsymbol{u}_{t-1} \,+\, \eta_t\,, \tag{10}$$

where $\boldsymbol{\mathcal{B}}$ is a constant parameter vector. That is, we may construct a consistent testing procedure based on (long memory) filtered variables instead of the original series. The main obstacles, however, are that we observe $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t')'$ and $\boldsymbol{\Lambda}$ is unknown. To accommodate these issues, we put forth a two-step procedure, where $\boldsymbol{z}_t$ is *fractionally-filtered* in a first step, and then a consistent frequency domain estimator of $\boldsymbol{\mathcal{B}}$ that explicitly accounts for the filtering errors is applied in a second step.[10]

Remark 1. *Drawing inference in equations* (7) *and* (10) *is analogous to estimating the cross-spectrum* $\boldsymbol{f}_{xx}^{-1}(\lambda)\boldsymbol{f}_{xy}(\lambda)$*. However, due to their strong persistence, one cannot readily apply standard local band-spectrum estimators, either. To see this, let, again,* $y_t \in I(d_y)$ *and* $\boldsymbol{x}_t \in I(d_x)$*, then as* $\boldsymbol{f}_{xx}^{-1}(\lambda)\boldsymbol{f}_{xy}(\lambda) = \boldsymbol{G}_{uu}^{-1}\boldsymbol{G}_{ue}(1-e^{i\lambda})^{d_y-d_x}$ *with* $(1-e^{i\lambda}) = O(\lambda^{-1})$ *as* $\lambda \to 0^+$*, such procedures are generally inconsistent unless* $d_y = d_x$*. This highlights the usefulness of a two-step procedure, where the first step accounts for* $(1-e^{i\lambda})^{d_y-d_x}$*, and the second step is designed to ensure robust inference.*

Remark 2. *Despite Assumptions* (3)-(6) *only parameterizing the low-frequency part of the spectrum (as* $\lambda \to 0^+$*), we emphasize that the (latent) test for predictability in equation* (10) *is not confined to persistent (or lower frequency) components in the filtered series* e_t *and* $\boldsymbol{u}_{t-1}$*; even white noise processes have constant spectral densities in the vicinity of the origin.*

3.2 Step 1: Fractional Filtering

To accommodate a wide range of alternative procedures, we do not adopt a specific estimator of the fractional integration orders, but rather assume we have an estimator, $\hat{d}_i$ for $i = 1, \ldots, k+1$, available, which satisfies mild consistency requirements. This is formalized through the following assumption,

Assumption F. *Let* $m_d \asymp n^{\varrho}$ *be a sequence of integers where* $0 < \varrho \leq 1$*, then, for all* $i = 1, \ldots, k+1$ *elements of* $\boldsymbol{z}_t$*, we assume to have an estimator with the property,*

$$\hat{d}_i - d_i \;=\; O_p\left(1/\sqrt{m_d}\right), \quad \textit{and we then let,} \quad \hat{\boldsymbol{D}}(L) \;=\; \operatorname{diag}\left[(1-L)^{\hat{d}_1}, \ldots, (1-L)^{\hat{d}_{k+1}}\right].$$

Assumptions F is very mild, essentially only requiring the existence of an estimator which, under the assumptions in Appendix A, is consistent. Of course, we need such consistency to hold for a wide range of d_i. To simplify the further analysis, we impose another mild restriction,

$$0 \leq d_i < 2, \quad \text{for } i = 1, \ldots, k+1, \quad \text{and we then define,} \quad \underline{d} = \min_{i=1,\ldots,k+1} d_i. \tag{11}$$

[10] Frisch-Waugh equivalence between equations (1) and (2) rests on the complete elimination of memory in y_t. For inference regarding the predictive relations in equation (10), we likewise must purge the variables of their long memory persistence.

This restriction is innocuous, in the sense that it is satisfied by most economic series, including all the macroeconomic and financial variables considered in our empirical analysis. In addition, the upper bound, $d_i < 2$, shortens the proofs considerably by allowing us to invoke periodogram bounds from Shimotsu (2010), and $0 \leq d_i$ enables us to provide a unified set of trimming and bandwidth conditions for the medium-band least squares estimator in the second estimation step in Section 3.3.

Once we have obtained $\hat{\boldsymbol{D}}(L)$, estimates of the innovations, $\boldsymbol{v}_t$, are given by,

$$\hat{\boldsymbol{v}}_t \equiv (\hat{e}_t, \hat{\boldsymbol{u}}_t')' = \hat{\boldsymbol{D}}(L)\boldsymbol{z}_t, \tag{12}$$

that is, without accounting for the mean, or initial value, in $\boldsymbol{z}_t$. Rather than treating "de-meaning" of the series on a case-by-case basis, depending on d_i, we account for the residual impact of the mean component, $\hat{\boldsymbol{D}}(L)\boldsymbol{\mu}$, in a unified manner during the second stage estimation.

We conclude this section with a couple of examples of memory parameter estimators, one semi-parametric and one parametric procedure, which are accommodated in our framework:

Example 1 (Exact local Whittle)**.** *The semiparametric exact local Whittle (ELW) by Shimotsu & Phillips (2005) and, in particular, the mean and trend-robust version in Shimotsu (2010) are accommodated by Assumption F, where the rate of convergence is restricted through the condition $\varrho < 4/5$, when the spectral density is sufficiently smooth ($\varpi = 2$ in Assumption D1 of Appendix A).*

Example 2 (ARFIMA filter)**.** *A parametric alternative to ELW estimation is fitting (possibly, long) ARFIMA(p, d, q) models, using, e.g., information criteria to determine p and q, and obtain estimates of d, relying on asymptotic results from Hualde & Robinson (2011) and Nielsen (2015). This procedure also requires $\varpi = 2$, and it achieves the optimal rate of convergence $\varrho = 1$.*[11]

3.3 Step 2: Medium Band Least Squares

After having computed $\hat{\boldsymbol{v}}_t$, we estimate the parameter vector $\boldsymbol{\mathcal{B}}$ in equation (10) in a second step using a new frequency-domain least squares estimator. To this end, we let,

$$\boldsymbol{w}_h(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \boldsymbol{h}_t \, e^{\mathrm{i}t\lambda_j}, \qquad \boldsymbol{I}_{hk}(\lambda_j) = \boldsymbol{w}_h(\lambda_j)\, \bar{\boldsymbol{w}}_k(\lambda_j), \tag{13}$$

be the discrete Fourier transform and cross-periodogram, respectively, where $\boldsymbol{h}_t$ and $\boldsymbol{k}_t$ are generic (and compatible) vector time series, and $\lambda_j = 2\pi j/n$ denotes the Fourier frequencies. Moreover, we denote by $\boldsymbol{I}_{hk}(\lambda_j) = \Re(\boldsymbol{I}_{hk}(\lambda_j)) + \mathrm{i}\Im(\boldsymbol{I}_{hk}(\lambda_j))$ the real and imaginary decomposition of $\boldsymbol{I}_{hk}(\lambda_j)$. Finally, define

[11] Assumptions D1-D3 in Appendix A mirror the corresponding assumptions in Shimotsu & Phillips (2005, Assumptions 1'-3'), but we need to impose slightly stronger differentiability assumptions in D3 to satisfy the (still very) mild conditions of Nielsen (2015, Assumption C). We refer to the latter for details.

the trimmed discretely averaged co-periodogram (TDAC) as,

$$\hat{\boldsymbol{F}}_{hk}(\ell,m) \;=\; \frac{2\pi}{n}\sum_{j=\ell}^{m}\Re(\boldsymbol{I}_{hk}(\lambda_j)), \qquad 1\le \ell\le m\le n, \tag{14}$$

where $\ell=\ell(n)$ and $m=m(n)$ are trimming and bandwidth functions, respectively. Then, we can write the TDAC of $\hat{\boldsymbol{u}}_{t-1}$ as $\hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m)$ and, similarly, of $\hat{\boldsymbol{u}}_{t-1}$ and $\hat{e}_t$ as $\hat{\boldsymbol{F}}_{\hat{u}\hat{e}}(\ell,m)$, and use these to define the medium band least squares (MBLS) estimator as,

$$\widehat{\boldsymbol{\mathcal{B}}}(\ell,m) \;=\; \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m)^{-1}\,\hat{\boldsymbol{F}}_{\hat{u}\hat{e}}(\ell,m), \tag{15}$$

for which $\ell,m\to\infty$ and $\ell/m+m/n\to 0$, as $n\to\infty$.[12] The MBLS estimator has some distinct advantages. First, we avoid being fully parametric about the dynamics of $\boldsymbol{z}_t$, needing only structure of the spectrum as $\lambda\to 0^+$. Second, by utilizing trimming and a bandwidth $m/n\to 0$, we asymptotically annihilate first-stage estimation errors from the filtering procedure. Specifically, the trimming component eliminates any slippage from the means, or initial values, $\hat{\boldsymbol{D}}(L)\boldsymbol{\mu}$ occurring at lower frequencies and, in conjunction with the bandwidth, the estimation errors in Assumptions F. The combination of these features ensures robust testing for the predictive power of the regressors in equation (10).

Christensen & Varneskov (2017) introduces the generic structure of the MBLS estimator (15) to analyze fractional co-integration among stationary long-memory processes in the presence of structural breaks and other low-frequency contaminants. Despite these similarities, this estimator differs from ours in equation (15) in important ways. Not only is the setting and objective different, but we perform estimation using fractionally filtered variables and impose new conditions on the trimming and bandwidth functions ℓ and m. Moreover, by using filtering in combination with the exact representation (9), we accommodate non-stationary variables, whereas Christensen & Varneskov (2017) require stationarity for all variables *and* cointegration between y_t and $\boldsymbol{x}_t$. This renders their approach ill-suited for predictive testing, as the latter condition is violated under the null hypothesis of no predictive ability, and the former condition rules out many series (with $d_i\ge 1/2$). Furthermore, direct applicability of their estimator is subject to the issues outlined in Remark 1, namely, if the variables of the system have different (and unknown) integration orders, the estimator is inconsistent for $\boldsymbol{\mathcal{B}}$. Even for the stationary case, with $d_i=d\in(0,1/2)$, $i=1,\dots,k+1$, the differences behind the procedures will, as detailed below, have a first-order impact on the distribution theory. Drawing an analogy with the differences between the LW and ELW estimators, their procedure is reminiscent of the former and ours of the latter. This explains why, as described in the following section, our proofs bear a strong resemblance to those in, e.g., Shimotsu & Phillips (2005) and Shimotsu (2010).[13]

---

[12] We have suppressed dependence on the (lagged) time t indicator in $\hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m)$ and $\hat{\boldsymbol{F}}_{\hat{u}\hat{e}}(\ell,m)$ to ease exposition. We will, however, explicate the dependence on time when necessary, e.g., when establishing Lemmas B.1-B.3 and B.10.

[13] Nonetheless, the relation to the MBLS estimator in Christensen & Varneskov (2017) suggests some inherent robustness to outliers, structural breaks, deterministic trends, etc., which are known to contaminate co-periodograms at frequency ordinates close to the origin. We do not formally analyze those effects here.

3.4 Asymptotic Theory and Inference

The development of the asymptotic theory for our two-step estimator requires additional assumptions.

Assumption T. *Let the bandwidth* $m \asymp n^{\kappa}$ *and* $\ell \asymp n^{\nu}$ *with* $0 < \nu < \kappa < \varrho \leq 1$*. Moreover, recall that the parameter* $\varpi \in (0, 2]$ *measures the smoothness of the spectral density in Assumption D1 of Appendix A. Then, the following cross-restrictions are imposed on* ℓ*,* m*,* m_d *and* n*,*

$$\frac{m^{1+2\varpi}}{n^{2\varpi}} + \frac{\ell^{1+\varpi}}{n^{\varpi} m^{1/2}} + \frac{n^{1/2}}{m_d^{1/2}\ell} + \frac{n^{1-2\underline{d}}}{m^{1/2-2\underline{d}}\ell^{2}} \to 0, \qquad \text{as} \quad n \to \infty.$$

The trimming conditions in Assumption T are mild. The first term is standard for semiparametric estimation in the frequency domain, e.g., Robinson (1995) and Lobato (1999). In our setting, this condition is needed, as we only impose local exogeneity in the spectrum (as $\lambda \to 0^+$) between the regressors, $\boldsymbol{u}_{t-1}$, in equation (10), and the regression residuals, η_t, rather than global exogeneity. The local exogeneity assumption is mild and relates to the Stambaugh (1999) bias, which we discuss further in Section 3.5. We stress, however, that our second stage MBLS estimator does not suffer from an asymptotic bias, regardless of the persistence of the VAR system (3). Note also that $\varpi = 2$ holds for the empirically relevant vector ARFIMA process, implying that $\kappa < 4/5$ must be satisfied.

The last three conditions in Assumption T are new to our MBLS estimator, imposing mild upper and lower bounds on the trimming rate, ν. Specifically, conditions two and four imply,

$$\nu < (\varpi + \kappa/2)/(\varpi + 1) \quad \text{and} \quad (1 - \kappa/2 - 2\underline{d}\,(1 - \kappa))/2 < \nu, \tag{16}$$

respectively, to restrict the loss of information from trimming frequencies and eliminate the low-frequency bias from mean-slippage after the first-stage fractional filtering. For the empirically relevant vector ARFIMA process (with $\varpi = 2$), and if we select κ arbitrarily close to its upper bound 4/5, we have $(3/5 - 2\underline{d}/5)/2 < \nu < 4/5$. Since the lower bound is strictly decreasing in $\underline{d}$, the highest lower bound is attained for $\underline{d} = 0$ and equals 3/10. The last condition, $0 < \nu < \kappa < \varrho \leq 1$, as well as the trimming restriction, $(1 - \varrho)/2 < \nu$, are needed to eliminate additional errors stemming from the estimation of integration orders, d_i, $i = 1, \ldots, k+1$, in the first stage. If we adopt a parametric estimator of d_i, $\varrho = 1$, and the restriction is trivial. If the estimator is semi-parametric, and we select $\kappa < \varrho$ as well as ϱ arbitrarily close to 4/5, the lower bound becomes $1/10 < \nu$.

We are now ready to state the distribution theory for the two-step MBLS estimator.

Theorem 1. *Let Assumptions F, T, and D1-D3 of Appendix A hold. Moreover, suppose that* $0 \leq d_i < 2$*,* $i = 1, \ldots, k+1$*, and* $\max(0, (1 - 3\kappa/2)/(1 + \kappa/2)) < \varpi \leq 2$*, then,*

$$\sqrt{m}\left(\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) - \boldsymbol{\mathcal{B}}\right) \xrightarrow{\mathbb{D}} N\left(\mathbf{0}, G_{\eta\eta}\boldsymbol{G}_{uu}^{-1}/2\right).$$

Theorem 1 demonstrates that the MBLS estimator is correctly centered, so it is not subject to the persistent regressor biases described by Stambaugh (1999) and Phillips & Lee (2013), and it converges at the rate $\sqrt{m}$, well-known from semiparametric estimation in the frequency domain, e.g., Brillinger (1981, Chapters 7-8), Robinson (1995), and Shimotsu & Phillips (2005).[14] The Gaussian distribution theory is remarkable, given that Robinson & Marinucci (2001) and Christensen & Nielsen (2006) find the asymptotic distribution for the NBLS estimator ($\ell = 1$ in equation (3.3)) to be non-uniform in the persistence of $\boldsymbol{z}_t$ in a fractional co-integration context, being Gaussian for the stationary case and exhibiting different forms of non-Gaussianity in non-stationary cases. In contrast, our inference procedure is uniform in $0 \leq d_i < 2$, $i = 1, \dots, k+1$. Moreover, the asymptotic distribution for our MBLS estimator differs from the one in Christensen & Varneskov (2017, Theorem 3) by being independent of the integration order of the variables, d_i, and it holds for stationary as well as non-stationary variables. The asymptotic variance only depends on the noise-to-signal ratio, $G_{\eta\eta}\boldsymbol{G}_{uu}^{-1}$. Finally, we stress that the asymptotic distribution is independent of mean slippage, first-stage fractional filtering errors and the trimming parameter, ℓ, as long as Assumption T is satisfied.

As a last obstacle for feasible inference and testing, we must provide consistent estimators of the long-run covariance matrix $\boldsymbol{G}_{uu}$ and variance $G_{\eta\eta}$. Again, the main challenge is that we observe $\hat{\boldsymbol{v}}_t$, not $\boldsymbol{v}_t$. Similarly, the residuals η_t are latent and we estimated them as,

$$\hat{\eta}_t \;=\; \hat{e}_t \,-\, \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)'\,\hat{\boldsymbol{u}}_{t-1}\,. \tag{17}$$

Now, using these filtered and estimated series, we define a generic class of estimators,

$$\widehat{\boldsymbol{G}}_{hh}(\ell_G, m_G) \;=\; \frac{1}{m_G - \ell_G + 1}\sum_{j=\ell_G}^{m_G} \Re\left(\boldsymbol{I}_{hh}(\lambda_j)\right), \tag{18}$$

for some arbitrary vector $\boldsymbol{h}_t$, while $m_G = m_G(n)$ and $\ell_G = \ell_G(n)$ are other bandwidth and trimming functions. This class of long-run covariance estimators is akin to those used by Christensen & Varneskov (2017). This is natural; both procedures rely on local spectrum theory ($\lambda \to 0^+$) and trimming of frequency ordinates. However, equation (18) differs importantly by using fractionally filtered series as input and an exact spectrum representation (9), not an approximation valid only in the stationary case, $d_i < 1/2$, $i = 1, \dots, k+1$. These points mirror the distinction among the MBLS estimators discussed earlier. Now, from equation (18), the asymptotic variance of $\widehat{\boldsymbol{\mathcal{B}}}(\ell, m)$ is estimated as,

$$\widehat{\mathbf{AVAR}} \;=\; \widehat{G}_{\hat{\eta}\hat{\eta}}(\ell_G, m_G)\,\widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G, m_G)^{-1}/\,(2m). \tag{19}$$

Hence, writing general linear hypotheses on the parameters $\boldsymbol{\mathcal{B}}$ as $\mathcal{H}_0 : \boldsymbol{\mathcal{R}}\boldsymbol{\mathcal{B}} = \boldsymbol{r}$ for some $k \times k$ selection matrix $\boldsymbol{\mathcal{R}}$ and $k \times 1$ vector $\boldsymbol{r}$, we only need to impose conditions on the bandwidth m_G and trimming function ℓ_G to be ready to introduce, and study the properties of, the LCM test.

Assumption T-G. *Let $m_G \asymp n^{\kappa_G}$ and $\ell_G \asymp n^{\nu_G}$ with $0 < \nu_G < \kappa_G \leq \varrho \leq 1$ and $n/(m_G \ell_G^2) \to 0$.*

[14] The condition $\max(0, (1 - 3\kappa/2)/(1 + \kappa/2)) < \varpi \leq 2$ ensures that the rate restrictions on the trimming and bandwidth functions in Assumption T are mutually consistent for all values of $0 \leq \underline{d} < 2$.

Theorem 2. *Let the conditions of Theorem 1 and Assumption T-G hold. Then,*

$$\mathcal{LCM}(\ell, m) \;\equiv\; \left(\boldsymbol{\mathcal{R}}\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) \,-\, \boldsymbol{r}\right)' \left(\boldsymbol{\mathcal{R}}\widehat{\mathbf{AVAR}}\boldsymbol{\mathcal{R}}'\right)^{-1} \left(\boldsymbol{\mathcal{R}}\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) - \boldsymbol{r}\right) \;\xrightarrow{\mathbb{D}}\; \chi^2_k.$$

Theorem 2 provides a significance test for a vector of candidate predictors, valid under general forms of (long memory) persistence and short memory dynamics, parameterizing only the spectrum of the processes near the origin ($\lambda \to 0^+$). Interestingly, under appropriate and mild rate conditions on the functions ℓ, m, ℓ_G and m_G, the limiting χ^2-distribution of $\mathcal{LCM}(\ell, m)$ is independent of the tuning parameters, making the procedure easy to implement in practice.

Remark 3. *Shao (2009) provides a generalized portmanteau test for independence of two stationary long memory processes ($d_i < 1/2$) based on spectral density estimates. However, the test is bivariate, it relies on a spectrum approximation that is invalid in the non-stationary case, and it assumes that all series have zero means. Hence, the test does not readily extend to our general setting.*

Remark 4. *Maynard et al. (2013) propose a two-step rebalancing approach to test for predictive power in univariate regressions, when the variables may have long memory. The main difference between their approach and ours is the second step. While they use OLS after filtering, we rely on MBLS. The use of standard least squares generates a different limiting distribution theory depending on whether y_t, x_{t-1} or both display long memory. Moreover, the first-stage filtering error generally impacts the asymptotic distribution, and the latter differs depending on whether or not the slope coefficient is zero. Finally, the regressor, x_t, is assumed to have zero mean. Our second-step MBLS approach overcomes these drawbacks and applies generally for multivariate testing problems.*

Remark 5. *Utilizing results from Phillips & Shimotsu (2004) and Shimotsu (2010), and furthermore letting $y_t \in I(d_y)$ and $\boldsymbol{x}_t \in I(d_x)$, our proofs of Theorem 1 illustrate that,*

$$\begin{aligned}\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) \;=\;& \left(\frac{2\pi}{n}\sum_{j=\ell}^{m} D_n(e^{\mathrm{i}\lambda_j}; d_x)^2 \Re(\boldsymbol{I}_{xx}(\lambda_j))\right)^{-1} \\ & \times \left(\frac{2\pi}{n}\sum_{j=\ell}^{m} D_n(e^{\mathrm{i}\lambda_j}; d_x) D_n(e^{\mathrm{i}\lambda_j}; d_y)\Re(\boldsymbol{I}_{xy}(\lambda_j))\right) \;+\; \boldsymbol{\mathcal{E}}(\ell, m, m_d, n), \end{aligned} \tag{20}$$

with $D_n(e^{\mathrm{i}\lambda_j}; d_y)$ and $D_n(e^{\mathrm{i}\lambda_j}; d_x)$ being defined in Appendix B.4 and $\boldsymbol{\mathcal{E}}(\ell, m, m_d, n)$ denoting an approximation error. This shows that our LCM procedure is related to the frequency domain GLS estimators in Robinson & Hidalgo (1997) and Nielsen (2005), who weight periodograms in the numerator and denominator by (the same) functions of the form $\lambda_j^{2d_u}$, d_u being the integration order of the residuals in the original regression (1)*, in parametric and semi-parametric frameworks, respectively. Hence, our LCM estimator, as seen by equation* (20)*, differs importantly from their estimators by utilizing exact differencing, resulting in differential weights of the form $D_n(e^{\mathrm{i}\lambda_j}; d)$, in conjunction with trimming.*

As a result, we can draw inference for systems with variables having $0 \leq d_i < 2$ and non-zero means, whereas the use of their estimators is constrained to stationary long memory processes.

Remark 6. *In a sequence of papers, Magdalinos & Phillips (2009), Kostakis et al. (2015) and Phillips & Lee (2013, 2016) develop the IVX methodology, which is an inference procedure for regressions, where the variables may be stationary, local-to-unity, unit roots, or mildly explosive. This involves a somewhat less flexible setting than for our LCM estimator, requiring, for example, the dependent variable not to be long-range dependent. Moreover, the IVX inference is not Gaussian uniformly in the persistence regimes, and the correlation between the regressors and the error term is restricted to be instantaneous. We explore the robustness of the LCM inference towards endogeneity in Section 3.5. However, importantly, the IVX-based Wald significance test is robust to the specific form of persistence, and it may also differ across regressors. In this sense, our LCM methodology may be viewed as an analogue for the general class of multivariate fractionally integrated processes, possessing similarly desirable properties for testing in predictive regressions. We take advantage of this feature by implementing the IVX methodology as a robustness check on our empirical results in Section 7.*

Remark 7. *The null hypothesis that $\boldsymbol{x}_{t-1}$ contains no predictive information for y_t rules out (fractional) cointegration between them and, hence, no information is lost by fractional filtering. If there is, in fact, cointegration between them, one may utilize this in the inference procedure, rather than filtering, to obtain a faster rate of convergence (relative to $\sqrt{m}$) for frequency domain least squares, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Christensen & Varneskov (2017). However, as such procedures impose predictability a priori, they are generally not amenable for testing.*

3.5 Robustness to Endogeneity Bias

The local exogeneity condition in Assumption D1 accommodates a setting reminiscent of Stambaugh (1999), as long as the correlation between the innovation to $\boldsymbol{u}_t$ and η_t is not too persistent, that is, as long as $\boldsymbol{u}_t$ and η_t have a co-spectrum with $\boldsymbol{f}_{u\eta}(\lambda) \sim \boldsymbol{0}$, as $\lambda \to 0^+$. In this section, we allow for a stronger degree of endogeneity which, arguably, is more aligned with the spirit of Stambaugh (1999), as well as the imperfect predictor definition in Pastor & Stambaugh (2009).

Suppose we observe $\boldsymbol{x}_t^c = \boldsymbol{x}_t + \boldsymbol{c}_t$ where, as before, $\boldsymbol{x}_t = \boldsymbol{\mu}_x + \boldsymbol{D}_x(L)^{-1}\boldsymbol{u}_t$, with $\boldsymbol{D}_x(L)$ being the lower right $k \times k$ submatrix of $\boldsymbol{D}(L)$, and $\boldsymbol{\mu}_x$ contains the last k elements of $\boldsymbol{\mu}$. Moreover, we let $\boldsymbol{c}_t$ be a $k \times 1$ mean-zero error process with co-spectum $\boldsymbol{f}_{c\eta}(\lambda) \sim \boldsymbol{G}_{c\eta}$, as $\lambda \to 0^+$, where $\boldsymbol{G}_{c\eta}$ can be non-trivial, and the components $\boldsymbol{u}_t$ and $\boldsymbol{c}_t$ are independent (note that all assumptions on $\boldsymbol{c}_t$ are formalized in Appendix A). In this setting, the predictor, or signal, of interest, $\boldsymbol{x}_t$, is contaminated with errors, giving rise to endogenous regressor problems, that generate a bias similar to the one analyzed by Stambaugh (1999). To cleanly identify the impact of endogeneity, suppose that $\boldsymbol{D}(L)$ is known, $\boldsymbol{\mu} = \boldsymbol{0}$, and exact differencing is carried out, such that we observe e_t and $\boldsymbol{u}_{t-1}^c = \boldsymbol{u}_{t-1} + \tilde{\boldsymbol{c}}_{t-1}$ with $\tilde{\boldsymbol{c}}_t = \boldsymbol{D}_x(L)\boldsymbol{c}_t$. Then, we have the decomposition,

$$\hat{\boldsymbol{F}}_{ue}^c(\ell, m) \;=\; \hat{\boldsymbol{F}}_{uu}(\ell, m)\boldsymbol{\mathcal{B}} + \hat{\boldsymbol{F}}_{u\eta}(1, m) + (\hat{\boldsymbol{F}}_{u\eta}(\ell, m) - \hat{\boldsymbol{F}}_{u\eta}(1, m)) + \hat{\boldsymbol{F}}_{\tilde{c}e}(\ell, m). \tag{21}$$

Now, utilizing local exogeneity between u_{t-1} and η_t, we know, from Lemma B.2 and B.3 of the Appendix, that $\sqrt{m}\lambda_m^{-1}\hat{\boldsymbol{F}}_{u\eta}(1,m)$ induces the central limit theory in Theorem 1, i.e., an $O_p(1)$ limit purged of any asymptotic bias, and $\sqrt{m}\lambda_m^{-1}(\hat{\boldsymbol{F}}_{u\eta}(\ell,m)-\hat{\boldsymbol{F}}_{u\eta}(1,m)) = o_p(1)$. However, the endogeneity-generated bias term $\sqrt{m}\lambda_m^{-1}\hat{\boldsymbol{F}}_{\tilde{c}e}(\ell,m)$ is unknown and may distort inference. Of course, this setting is simplified as the mean, or initial values, generally are non-zero, and the integration orders are unknown and must be estimated in the fractional-filtering stage. Nonetheless, equation (21) reveals an additional source of complexity for inference on predictive relations in long-memory systems.

We now demonstrate that trimming of frequency ordinates is useful, not only for the asymptotic elimination of errors due mean-slippage and fractional-filtering, but also in terms of boosting robustness towards biases arising from endogenous regressors. To this end, we define the fractionally-filtered and contaminated regressors, $\hat{\boldsymbol{u}}_t^c = \hat{\boldsymbol{D}}_x(L)\boldsymbol{x}_t^c = \hat{\boldsymbol{D}}_x(L)\boldsymbol{x}_t + \hat{\boldsymbol{D}}_x(L)\boldsymbol{c}_t$, and we let $\hat{\boldsymbol{c}}_t = \hat{\boldsymbol{D}}_x(L)\boldsymbol{c}_t$. Specifically, we next establish asymptotic bounds for $\hat{\boldsymbol{F}}^c_{\hat{u}\hat{u}}(\ell,m) - \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m)$ and $\hat{\boldsymbol{F}}^c_{\hat{u}\hat{e}}(\ell,m) - \hat{\boldsymbol{F}}_{\hat{u}\hat{e}}(\ell,m)$, that is, the additional source of errors for the estimator (15), stemming from having an endogenous component embedded in the regressors. Moreover, to carry out testing using the LCM approach in Theorem 2, we define $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) = \hat{\boldsymbol{F}}^c_{\hat{u}\hat{u}}(\ell,m)^{-1}\hat{\boldsymbol{F}}^c_{\hat{u}\hat{e}}(\ell,m)$ and $\hat{\eta}_t^c = \hat{e}_t - \widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m)'\hat{\boldsymbol{u}}_t^c$, and we then obtain equivalent bounds for $\widehat{G}^c_{\hat{\eta}\hat{\eta}}(\ell_G,m_G) - \widehat{G}_{\hat{\eta}\hat{\eta}}(\ell_G,m_G)$ and $\widehat{\boldsymbol{G}}^c_{\hat{u}\hat{u}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G,m_G)$, that control the errors entering through the variance estimator (19).

Theorem 3. *Suppose the conditions of Theorems 1 and 2 as well as Assumption C hold. Moreover, suppose that $n^{1/2}/m \to 0$, $n^{1/2}/m_G \to 0$ and $\underline{d} > 0$, then, for some arbitrarily small $\epsilon > 0$,*

(a) $\lambda_m^{-1}(\boldsymbol{F}^c_{\hat{u}\hat{u}}(\ell,m) - \boldsymbol{F}_{\hat{u}\hat{u}}(\ell,m)) = O_p((m/n)^{\underline{d}}/\ell^{1+\epsilon})$,

(b) $\sqrt{m}\lambda_m^{-1}(\boldsymbol{F}^c_{\hat{u}\hat{e}}(\ell,m) - \boldsymbol{F}_{\hat{u}\hat{e}}(\ell,m)) = O_p((m/n)^{\underline{d}}m^{1/2}/\ell^{1+\epsilon})$,

(c) $\widehat{\boldsymbol{G}}^c_{\hat{u}\hat{u}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G,m_G) \le O_p((m_G/n)^{\underline{d}}/\ell_G^{1+\epsilon})$,

(c) $\widehat{\boldsymbol{G}}^c_{\hat{\eta}\hat{\eta}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\hat{\eta}\hat{\eta}}(\ell_G,m_G) \le O_p((m_G/n)^{\underline{d}}/\ell_G^{1+\epsilon}) + O_p((m/n)^{\underline{d}}/\ell^{1+\epsilon})$.

Theorem 3 provides several interesting insights. First, from (a) and (b), we observe that trimming is instrumental for the elimination of the endogenous regressor bias. In fact, if $\ell = O(1)$, we need to impose $\kappa < \underline{d}/(1/2+\underline{d})$ to avoid that the bias has a first-order (or larger) asymptotic impact on the inference. This likely will hurt the efficiency of the LCM inference severely, in particular for small $\underline{d}$. Second, if we, on the other hand, let $\ell \to \infty$, as $n \to \infty$, and impose $\kappa/2 - \underline{d}(1-\kappa) < \nu$ in addition to the conditions in Assumption T, we can readily utilize trimming to eliminate the endogeneity bias, asymptotically, for all values of κ, thereby retaining the asymptotic efficiency of the LCM procedure reported in Theorem 1 and 2, obtained without endogenous components in the regressors. Since this bound is strictly decreasing in $\underline{d}$, the worst case applies for $\underline{d}$ arbitrarily close to 0, which, in conjunction with selecting κ arbitrarily close to its upper bound 4/5 for efficiency, implies $2/5 < \nu$. Hence, quite intuitively, we require stronger trimming to retain the same asymptotic efficiency in the presence of endogenous regressors, if the minimal persistence of the system is small. Third, we impose $\underline{d} > 0$ to separate the signal in the predictive regressors from its noise, asymptotically, in analogy to the approach in Pastor & Stambaugh (2009). Moreover, importantly, this restriction may be relaxed to only require $\min_{2,\ldots,k+1} d_i > 0$, that is, we do not need y_t to be fractionally integrated for this identification, and thereby Theorem 3, to hold. This ensures that the procedure is applicable for predictive regressions for stock returns with fractionally integrated regressors, which may be stationary or non-stationary and contain endogenous innovations. Fourth, the conditions $n^{1/2}/m \to 0$ and $n^{1/2}/m_G \to 0$ on the bandwidths are imposed to simplify exposition and avoid stronger cross-restrictions on the tuning parameters. These may be relaxed.

Corollary 1. *Suppose that the conditions of Theorem 3 hold and $\kappa/2 - \underline{d}(1-\kappa) < \nu$. The asymptotic limits in Theorem 1 and 2 then still apply with $\boldsymbol{x}_t^c$ in place of $\boldsymbol{x}_t$.*

Corollary 1 demonstrates that, utilizing trimming, we can ensure that the endogenous regressor bias does not affect the first-order asymptotic theory for the LCM approach. Instead, it will be of second or smaller order, depending on $\underline{d}$. This provides a sharp contrast to the corresponding results of Stambaugh (1999), who shows that the bias is of second order if the regressors are stationary, and Phillips & Lee (2013), who document that an uncorrectable bias enters the asymptotic distribution if the regressors are local-to-unity. The LCM approach successfully eliminates such concerns, uniformly across the empirically relevant long-memory regimes.

In order for these observations to be applicable in practice, the first step fractional-filtering procedure must also be robust to the presence of noise in the series. This is, indeed, the case.

Remark 8. *Despite the latent regressor signal being perturbed by noise, as $\boldsymbol{x}_t^c = \boldsymbol{x}_t + \boldsymbol{c}_t$, it follows from, e.g., Deo & Hurvich (2001), Arteche (2004) and Frederiksen, Nielsen & Nielsen (2012), that standard semi-parametric estimators of d_i, $i = 1, \ldots, k$, remain valid, albeit suffering from a higher-order bias. Similarly, parametric methods may still be utilized by carefully choosing the lag structure of the short-memory component of the filter, utilizing equivalent representations of an ARMA(p,q) process with measurement noise and ARMA$(p, \max(p,q))$ models, see, e.g., Granger & Morris (1976).*

4 Empirical Illustration: Forecasting Equity Market Volatility

The prediction of future realized equity market volatility using financial and macroeconomic indicators in first-order VAR systems has gained renewed attention, as illustrated by the studies of, e.g., Christiansen et al. (2012), Paye (2012), Bansal et al. (2014), Campbell et al. (2017), and Dew-Becker et al. (2017). This section replicates the qualitative evidence generated by these papers through an empirical illustration in which we rely on standard least squares techniques. However, we note that the series display pronounced persistence, pointing towards potential problems with the inference. Hence, in Section 6, we reassess the evidence using our robust LCM procedure.

4.1 Data Description

We employ two separate data sets of monthly observations for realized volatility of the aggregate U.S. stock market, proxied by the S&P 500. The first spans the period from February 1960 through March 2015 and exploits realized variance measures constructed from daily data. This time span mimics those covered by prior studies in the literature.[15] Since high-frequency data are available for the last part of the sample period, we undertake an additional analysis, covering January 1990 through March 2015, using a more accurate measure of the realized (log-)return variance.

We first introduce the two realized variance measures. We let $r_{t,i}$ denote the daily log-return on the S&P 500 for trading days $i = 1, \dots, n_t$ in months $t = 1, \dots, n$, and then,

$$\tilde{V}_t \; = \; \sum_{i=1}^{n_t} r_{t,i}^2 \,, \tag{22}$$

comprises our low-frequency (LF) realized variance measure.[16] Such return variance measures over a fixed (here, monthly) horizon computed from intermediately sampled data have been widely used in financial econometrics since the work of, e.g., Andersen & Bollerslev (1998), Barndorff-Nielsen & Shephard (2002), and Andersen, Bollerslev, Diebold & Labys (2003).

Next, for the period where an intra-daily price record is available, we construct an alternative return variation measure. We use high-frequency return data for the CME Group E-mini S&P 500 futures to construct accurate trading day measures, and then add the squared close-to-open returns to obtain an overall variation measure. Since market microstructure frictions induce unwarranted serial correlation in high-frequency returns, we rely on the flat-top realized kernel of Varneskov (2016, 2017) during the trading day. This approach is robust to general forms of microstructure noise and possesses desirable asymptotic properties and good finite sample performance. Since our construction otherwise follows standard procedures, we relegate the details to Section C of the appendix.

The macroeconomic and financial indicators consists of monthly series for the default spread (DS), three-month U.S. Treasury bills (TB), and price-earnings ratio (PE). They have all been found to be successful predictors of equity-index return volatility in recent studies. We follow the literature in defining DS as the difference between the logarithmic percentage yield on Moody's BAA and AAA bonds, the Treasury bill rates are log-transformed, and the PE is constructed as the logarithm of the ratio of the S&P 500 index to the ten-year trailing moving average of aggregate earnings on the S&P 500 index constituents.[17] The source of the different series is provided in Section C of the appendix.

[15] However, existing work often relies on lower frequency series. Among our references, only Christiansen et al. (2012), Paye (2012) and Dew-Becker et al. (2017) use monthly data; Bansal et al. (2014) use yearly data and Campbell et al. (2017) use quarterly. We adopt monthly sampling to increase the power of the statistical significance tests and facilitate the study of causality and directional predictability. The latter is discussed in Section 7.

[16] We label it a LF measure since no intra-daily, i.e., "high-frequency" financial data are used in its construction.

[17] Originally, we used the term spread, defined as the difference between the logarithmic rates on 10-year U.S. Treasury constant maturity bonds and 3-month U.S. Treasury bills instead of simply the 3-month U.S. Treasury bill series, and the results were qualitatively identical to those presented below. The change was motivated by a desire to match the set of state variables employed in the final version of Campbell et al. (2017).

Table 1 presents full-sample summary statistics for the four series, as well as square-root and log-transformations of the LF realized variance.[18] Beyond the usual unconditional measures, we report a set of statistics that speak to the time series properties of the series. We provide estimates for the degree of fractional integration, using both the local Whittle (LW) estimator, cf. Künsch (1987), and the local polynomial Whittle with noise (LPWN) estimator, see Frederiksen et al. (2012), which is more robust against measurement errors and short-memory dependencies. Further, we report the MZ unit root test statistic of Ng & Perron (2001), which is correctly sized and has good power properties against local alternatives, and we include the KPSS test statistic of Kwiatkowski, Phillips, Schmidt & Shin (1992) for an $I(0)$ process against the alternative of an $I(1)$.

Table 1 documents that the realized variance (RV) distribution is positively skewed and has fat tails. These features are mitigated by the concave square-root or logarithmic transformations. Similar comments apply to the DS, whereas TB and PE are closer to being Gaussian. The most noteworthy results, however, concerns the conditional properties of the series. Specifically, at standard levels of significance, we reject that the realized variance is either $I(0)$ or $I(1)$. Instead, the series is best characterized as fractionally integrated with d in the 0.25-0.6 range, depending on the transformation. This is consistent with the findings of the comprehensive literature referenced in Section 1. Furthermore, the larger estimates for the fractional integration order, obtained as we apply more concave transformations to realized variance, are consistent with the findings of, e.g., Haldrup & Nielsen (2007), who show that outliers, as reflected in the skewness, may bias various estimators of d downwards.

We also reject the null hypothesis of the state variables being $I(0)$ processes and, from the MZ test, we similarly reject the DS and TB series being $I(1)$. The LW and LPWN estimates corroborate these findings, suggesting that the DS is fractionally integrated with $d \simeq 0.8$, while TB is slightly more persistent with $d \simeq 0.9$.[19] Finally, we cannot reject that PE is a unit root process.

For visualization, we complement the estimates of d and the unit root test by depicting the autocorrelation functions (ACFs) in Figure 1. The slowly decaying ACFs are consistent with all four series being long-range dependent, with PE being most persistent, followed by TB, DS and RV. This corroborates our estimates of the (relative) size of the respective fractional integration orders.

4.2 Standard Predictive Regressions

The evidence in Table 1 and Figure 1 is consistent with the realized variance series being fractionally integrated with $0.25 < d < 0.6$, while the macroeconomic and financial state variables all have $d \geq 0.8$. Hence, equation (8) suggests that standard least squares procedures provide misleading inference, as either Case (a) or (b) apply in this context. Nonetheless, we shall initially ignore this issue, as we seek to establish a benchmark for the (apparent) predictive power obtained through the commonly adopted OLS procedures.

[18] The corresponding statistics for the subsample are very similar and omitted for brevity.

[19] For the LPWN estimator, the integration order d is restricted to the range $(0.01, 0.99)$, because the asymptotic results in Frederiksen et al. (2012) apply only for $d \in (0, 1)$. Hence, estimates reported as $d = 0.01$ or $d = 0.99$, reflect that the lower, respectively, upper bound has been hit.

An informal assessment may be drawn from Figure 2, which plots the state variables against the future realized variance over the last fifteen years of the sample, which is characterized by a particularly high degree of coherence among the series. We observe, in particular, that the realized variance and DS both are elevated in the fall of 2008, while PE drops sharply during the same time span. However, we also note that the spikes in the state variables often lag those in the market variance by a few months, raising questions regarding the direction of predictive causality. We return to this issue in Section 7, using the LCM approach. Notwithstanding this issue, the fact that the predictor variables all display abnormal variation during the turbulent market conditions surrounding the recent financial crisis suggest they may carry (important) information about the future realized variance.

Motivated by the extant literature and the evidence in Figure 2, we run predictive regressions for the realized variance assuming a first-order VAR system, where $\boldsymbol{X}_{t-1}$ in equation (1) consists of the lagged realized variance and the macroeconomic and financial state variables. The coefficient estimates, HAC standard errors and adjusted R^2 are reported in Table 2 along with LW estimates of the residual memory and a HAC-based Wald test for the joint significance of the three state variables. The results for both samples are similar: (1) DS seemingly predicts realized variance and its strength increases with the addition of PE and TB; (2) all state variables are individually significant (except TB in the subsample with high-frequency data) and the R^2 increases slightly with their inclusion; and (3) the Wald tests show joint significance at a 5% level ($\mathbb{P}$-Wald above 0.95). Hence, the full sample results corroborate prior findings by Christiansen et al. (2012), Paye (2012), Bansal et al. (2014), Campbell et al. (2017), and Dew-Becker et al. (2017). Moreover, the subsample results, utilizing a high-frequency measure of volatility, demonstrates that they are robust to the choice of realized variance proxy. However, as noted above, this OLS analysis does not account for the degree of persistence of the realized variance and the regressors which, as equation (8) shows, will distort the inference. Consequently, we next explore whether these caveats are relevant for the present setting via simulations, based on the time series properties of the variables conveyed by Table 1, and contrast them to the size and power properties of the LCM procedure.

5 Simulation Evidence: Size, Power and Spurious Inference

The section examines the properties of the OLS and LCM procedures in a setup that captures the persistence of the realized variance and the predictor variables, as summarized in the previous section. We first provide size and power results for the LCM procedure in a bivariate setup and then proceed to study the properties of the LCM and OLS procedures in a more general predictive setting.

5.1 Finite Sample Performance of the LCM Test

The size and power of the LCM test are analyzed in a bivariate setting, resembling the one in Hong (1996) and Shao (2009), but generalized to allow for non-stationary long memory. This entails simulating fractional ARMA$(1,0)$ processes for y_t and x_t, $t = 1, \dots, n$, in equation (7) as,

$$(1-L)^{d_y}(1-\phi_y L)y_t \;=\; \rho u_t + \sqrt{1-\rho^2}v_t, \qquad (1-L)^{d_x}(1-\phi_x L)x_t \;=\; u_t, \tag{23}$$

respectively, where u_t and v_t are i.i.d. standard Gaussian random variables. Despite its simplicity, this setup encompasses several distinct inference scenarios, depending on the identity of the parameter vector, $(\phi_y, \phi_x, \rho, d_y, d_x)$. First, we fix $\phi_y = \phi_x = 0.2$ and let either $\rho = 0$ or $\rho = 0.2$, when examining the size and power properties, respectively. Second, we vary the fractional integration orders d_y and d_x to consider different persistence regimes. In particular, **DGP 1** is configured with $(d_y, d_x) = (0.30, 0.45)$; **DGP 2** with $(d_y, d_x) = (0.30, 0.80)$; **DGP 3** with $(d_y, d_x) = (0.55, 0.45)$; and **DGP 4** with $(d_y, d_x) = (0.55, 0.80)$. These values are in line with the estimated memory parameters for the realized variance and DS in Table 1, and they capture all three inference cases in equation (8). Finally, we consider two different sample sizes $n = \{300, 650\}$, closely matching the respective size of the subsample ($n = 302$) and full sample ($n = 662$) in our empirical analysis.

Implementing the LCM test in Theorem 2 requires choosing an estimator for the first step fractional filtering and tuning parameters for the MBLS estimation in the second step. We estimate the memory parameters using a parametric fractional ARMA$(1,0)$ model, noting we may apply results from Hualde & Robinson (2011) and Nielsen (2015) to verify that Assumption F holds with $\varrho = 1$.[20] Moreover, we consider different tuning parameters for MBLS to analyze their finite sample impact on the LCM test. Specifically, we let $\nu = \{0.21, 0.25, 0.30\}$, $\kappa = \{0.70, 0.75, 0.799\}$, $\nu_G = 0.25$ and $\kappa_G = 0.9$. While the bandwidth rate κ generally is chosen close to its upper bound to boost the efficiency of the inference and satisfy the condition $n^{1/2}/m \to 0$ in Theorem 3, the selection of the trimming rate ν is guided by an assessment of empirically realistic lower bounds. In particular, using the estimate $\underline{d} \simeq 3/10$ from Table 1 (the memory of RV), the lower bound restriction in Assumption T implies that, if κ is close to $4/5$, then $(3/5 - 2\underline{d}/5)/2 < 1/4$. Similarly, since the lowest integration order of the state variables is $\min_{i=2,\dots,k+1} d_i \simeq 4/5$, the restriction imposed by a potential endogenous component in the regressors is similarly strictly less than $1/4$. Hence, the values $\nu = \{0.21, 0.25, 0.30\}$ capture realistic lower bounds for the trimming, as they are guided by our empirical application.[21] Finally, Assumption T-G is satisfied by selecting $\nu_G = 0.25$ and $\kappa_G = 0.9$. All tests are implemented with a nominal size of 5%, and the simulations are performed using 1,000 replications. The results are reported in Table 3.

Table 3 documents that the LCM test has excellent finite sample properties. For the small (sub-)sample size $n = 300$, the test is only slightly oversized, with rejection rates in the range 6-9% compared

[20] For the empirical implementation, we rely on fractional ARMA$(p,0)$ models with $p = \{0, 1, 1, 4\}$ for the realized variance, DS, PE and TB, respectively. These models fit the data well, as indicated by the residuals in Figure 3 and Section 6, and adding more lags barely increases the explanatory power. Again, Shao (2009), Hualde & Robinson (2011) and Nielsen (2015) provide results which may be used to verify that the estimates of d are consistent at rate $n^{-1/2}$. Finally, when implementing the fractional filter in the first step, we use 10 observations for initialization.

[21] We also implemented the LCM test with the worst case lower bound on the trimming rate implied by endogenous regressors, $2/5 < \nu$, both in the empirical analysis and simulation study. In spite of this very conservative choice, the persistent state variables imply that the numerical results remain very similar to those reported below, and they are omitted for brevity.

to the 5% nominal level, and the power is good, especially when selecting the wider bandwidth, $\kappa = 0.799$. For the larger sample size of $n = 650$, the LCM test shows both great size and power. Moreover, we note that the LCM test performs well across all DGPs, so it is robust to a variety of different, and empirically relevant, persistence scenarios. Finally, while the test is robust to the choice of the trimming rate ν, yielding similar results across the board, the power is uniformly higher for a larger bandwidth ($\kappa = 0.799$), and this is achieved without sacrificing the size properties. Consequently, we use $\kappa = 0.799$ throughout.

5.2 Inference on Predictive Ability: LCM versus OLS

This section explores the performance of LCM and least squares tests by examining their size properties in settings, where we aim to predict a persistent variable using persistent regressors. Specifically, we adopt a scenario similar to equation (23) with $\rho = 0$ and allow for additional exogenous processes of the same form as x_t. We consider regressions with up to three state variables, setting $d_x = \{0.8, 0.9, 1\}$, corresponding to the estimated fractional integration orders of DS, TB, and PE in Table 1, and we let $d_y = 0.3$ (**DGP M1**) or $d_y = 0.55$ (**DGP M2**) to capture the relevant range of persistence in the RV measures. The autoregressive parameter is fixed at 0.2 for all processes, and a larger sample size $n = 1000$ is included to help gauge the limiting properties of the testing procedures. Whereas the LCM test is implemented as described in the previous section, OLS inference is performed using Newey & West (1987) standard errors and Wald tests. In all regressions, we include a constant, the lagged realized variance, and one to three exogenous regressors. We report test power for lagged RV, test size for the individual state variable, the average adjusted R^2, and the size of a Wald test for joint significance of the exogenous regressors. The results are displayed in Table 4.

Table 4 contains several interesting findings. First, for **DGP M1**, all significance tests for individual coefficients of the persistent regressors are oversized, irrespective of the number of regressors being included, and, in fact, the size distortions only grow as the sample size increases. This corroborates the theoretical result in Case (a) of equation (8), and it shows that the size distortions are substantial, causing serious concerns about the applicability of OLS inference. For example, for simple regressions involving only one exogenous predictor with $d_x = 0.8$ (mimicking the DS), we see that nominal rejection rate rises from 27.7% for $n = 300$ to 33.5% for $n = 1000$, far exceeding the nominal 5% level. Second, sequentially adding exogenous processes, with $d_x = 1$ (PE) and $d_x = 0.9$ (TB) fails to restore the testing properties of DS, whose individual significance tests continue to be badly distorted. Third, adding predictors enhances the adjusted R^2. This is, again, readily explained by Case (a) of equation (8). Fourth, the OLS-based Wald test for joint significance of the predictors is severely distorted, and the size properties only worsen as more persistent predictors are introduced and the sample size increases. The nominal rejection rates range from 27.7% to 59.8%, underscoring the propensity for misleading empirical inference. Fifth, the size properties of the LCM test are excellent. Although the test is slightly oversized for $n = 300$, the size is very accurate for $n = 650$, and essentially perfect for $n = 1000$, regardless of the number of predictors included.

The results for **DGP M2** are very similar to those described above, with all qualitative conclusions being identical despite some minor numerical differences. Perhaps most notably, the size distortions are even greater for the individual and joint tests in this setting, which is consistent with the different divergence rates of the t-statistics indicated in equation (8) for Case (a) and (b), respectively.

Overall, the simulations demonstrate severe problems with OLS-based inference and testing, irrespective of the predictive ability of the persistent regressors. In contrast, our LCM procedure possesses very good size and power properties in finite samples, and it should allow for reliable inference in settings with general and diverse degrees of persistence among the variables in the system. Hence, the LCM test provides a rigorous basis for deciding whether our macroeconomic and financial state variables add auxiliary predictive power beyond the past volatility in forecasting future realized variance.

6 LCM Analysis of Predictive Power for Future Realized Variance

Section 5 documents that least squares predictive inference and testing procedures are unreliable, when the variables of the system are persistent. Hence, we now revisit the findings in Table 2 concerning the significant forecast power of macroeconomic and financial indicators for future realized variance, using the robust LCM approach. As described in the previous section, first-stage fractional filtering is based on estimates from (long) fractional ARMA$(p, 0)$ processes, as in Shao (2009). To gauge the suitability of the fractional ARMA models, we depict the ACFs of the model-implied residual series in Figure 3. Relative to Figure 1, the effectiveness of our parametric approach to "whiten" the variables of the system is evident.[22] Likewise, the MBLS estimation in stage 2 is implemented as described above, with the trimming rates $\nu = \{0.21, 0.25, 0.30\}$ and bandwidth parameter $\kappa = 0.799$. The results for the full sample and the subsample, exploiting the high-frequency data, are reported in Table 5.

First, from Panel A of Table 5, we observe that the coefficient estimates from the LCM procedure and the corresponding ones for OLS in Table 2 are of similar magnitude and sign. Second, and importantly, the LCM test for joint significance of the persistent state variables are now all insignificant, as the $\mathbb{P}$-Wald statistics are far below any conventional significance level. This contrasts sharply with the OLS results, which indicate that the state variables are jointly significant the 95% level, and sometimes at the 99% level. Third, the LCM results are robust to the selection of tuning parameters and the choice of RV measure, as seen by the subsample results in Panel B. In fact, the LCM tests for predictive power are even less significant in the subsample. Again, this is contrary to the OLS results for joint significance, which are stronger in the subsample (significant at the 99% level).

The stark difference between the LCM and OLS results is readily explained by our theoretical and simulation results. The theory in Section 2.2 and finite sample evidence in Section 5, combined with the empirical results in Table 1, documenting a strong degree of persistence for the realized variance as well as the financial and macroeconomic series, imply that the OLS-based tests are inconsistent. In fact, these procedures will, incorrectly, reject the null hypothesis of no predictability with probability

[22] The estimated persistence is similar to that conveyed by the results in Table 1.

approaching one, as the sample size grows. Our LCM procedure, on the other hand, remains valid irrespective of the persistence displayed by the variables of the VAR system. We conclude that there is no statistical evidence that any of the selected state variables contain relevant information for forecasting the future realized variance.

7 Robustness and Reverse Predictability

The LCM results in the previous section indicate that the selected financial and macroeconomic state variables do not possess significant predictive power for the future realized variance. This section provides a robustness check with respect to the modeling of the persistence in the VAR system, using the IVX approach of Kostakis et al. (2015) and Phillips & Lee (2016) reviewed in Appendix D. Next, using the LCM approach, we explore whether the realized volatility has predictive power for the macroeconomic indicators, i.e., in the reverse direction of what was tested so far.

7.1 IVX Analysis of Predictability for Realized Volatility

The long memory paradigm is widely acknowledged in the volatility modeling literature, e.g., Andersen et al. (2003). However, a different econometric framework is often adopted, when studying (long-horizon) return predictability using financial and macroeconomic state variables. This alternative approach assumes that the persistence is generated by local-to-unity processes, and robust inference methods have been developed for this setting. Hence, as robustness check, we now assume that the dynamics of the VAR system can be described as autoregressive with stationary, local-to-unity, unit root or locally explosive persistence, and then apply the IVX procedure, cf. Appendix D, to test realized variance predictions using the three persistent state variables. The results from IVX Wald tests are presented in Table 6.

The IVX results are fully consistent with those for the LCM procedure in Table 5. In particular, the coefficient estimates are similar (slightly different for the PE), and the IVX Wald tests for joint significance of state variables also fail to detect any significant evidence of predictive ability. Hence, the empirical discrepancy between the LCM and OLS testing procedures, as described in the previous section, is robust to the paradigm adopted for modeling the persistence of the system.

7.2 LCM Testing of Reverse Causality

The LCM test finds no significant evidence of predictive information for realized volatility in the three persistent state variables. This does not imply that they are unrelated. Specifically, as noted in Section 4.2, Figure 2 reveals that some major peaks in the DS and PE series trail the realized variance, not the other way around. Hence, this section explores the reverse predictive relation, that is, whether lagged realized variance carry information about the subsequent realization of the three state variables. For illustration, Figure 4 plots the lagged realized variance against the three state variables, both before and after the first-stage fractional filtering step of the LCM procedure, for the last 15 years of the sample. Two points stand out. First, the fractional filtering is successful in stripping the persistence from the variables, as intended. Second, the large spikes in the DS and PE ratio during the recent financial crisis occur contemporaneously with outliers in the *lagged* realized variance series, suggesting that the latter carries important information about the realized values of the state variables. The results from testing this hypothesis, using the LCM procedure, are presented in Panels A and B of Table 7 for the full sample and subsample with high-frequency return data, respectively.

The message from Table 7 is clear; realized variance predicts future changes in all three state variables, and the results hold for both samples, so it applies for the different RV measures. The predictive relations are strongly significant for all Wald tests (the $\mathbb{P}$-Wald measures exceed 0.95). The estimates imply that an increase in realized variance forecasts future elevations in DS and declines for TB and PE, consistent with the visual evidence in Figure 4. Obviously, the recent financial crisis is an extraordinary, yet important, economic event with an extraordinary impact on the inference, as is also evident from Figure 4. Hence, for robustness, we also implement the LCM test for the sample truncated in December 2007 ($n = 574$). Panel C of Table 7 reveals that positive shocks to realized variance remain significant predictors of future increases in DS and declines in PE.

Our LCM methodology is specifically designed to accommodate persistent variation in VAR systems, so it is well positioned to uncover low-frequency ties between the realized variance and macroeconomic variables. The above findings show that the market variance is related to important macroeconomic state variables, even if the latter do not forecast the former, but rather the reverse. These findings provide a challenge for Campbell et al. (2017) and Bansal et al. (2014), who assert, based on least squares inference, that such macroeconomic indicators do predict the market variance. Moreover, this point is important for their conclusion that shocks to the market variance are integral to understanding the role of diverse macroeconomic fluctuations in driving the cross-sectional pricing across distinct asset classes. Our results suggest it may be useful to revisit these studies and reassess the evidence. We defer an in-depth investigation of the implications for macroeconomic and realized variance prediction to future research.

8 Conclusion

This paper studies the properties of standard predictive regressions in persistent VAR economies and considers robust inference and testing in such systems. In particular, we analyze a setting, where all variables may be fractionally integrated of different orders and show that this induces a spurious regression problem for least squares methods. As a remedy, we propose a new inference and testing procedure – the local spectrum approach – for joint significance of the predictors, that is robust to the variables having different integration orders. The LCM procedure is based on (semi-)parametric fractional-filtering and band spectrum regressions (MBLS), using a carefully selected set of frequency ordinates. We establish the asymptotic properties of the coefficient estimates and the associated significance test, relying on an exact spectrum representation. The latter allows us to include both stationary ($0 \leq d < 1/2$) and non-stationary ($d \geq 1/2$) variance in the system. The theoretical analysis is supplemented with an empirically relevant simulation study, documenting that least squares inference methods suffer from large size distortions when the variables are persistent. In contrast, our LCM approach displays excellent finite sample size and power.

We use the LCM procedure to study the implications of assuming standard short-memory VAR dynamics for the economy in predictive regressions for the realized variance of the S&P 500 equity index. Focusing on three financial and macroeconomic state variables, whose forecasting ability have been widely appraised in the macro-finance literature, we confirm that least squares methods generate evidence supportive of highly significant forecast power. However, we find no such evidence using the LCM approach. We argue that this suggests that the standard least squares evidence is spurious, driven by the (ignored) strong persistence of the VAR economy. In fact, our robust LCM approach suggests that causality may run in the reverse direction, i.e., innovations to the realized variance may foreshadow future changes in the state variables. Overall, our findings carry implications for several areas in empirical macroeconomics and finance, including the choice of econometric tools for modeling.

Full Sample Summary Statistics						
Panel A:	Mean	S.D.	Max	Min	Skew	EKur
RV	0.0021	0.0047	0.0814	0.0001	11.088	159.06
Sqrt-RV	0.0396	0.0239	0.2853	0.0104	3.9174	27.513
Log-RV	-6.6972	0.9305	-2.5086	-9.1403	0.5268	1.0794
DS	13.462	5.6609	51.241	6.5329	2.4891	9.6242
TB	4.6805	2.9352	15.100	0.0100	0.5813	0.7491
PE	2.8975	0.4154	3.7887	1.8929	-0.3301	–0.4229
Panel B:	LW	LPWN	AR-ϕ	AR-R^2	KPSS	MZ
RV	0.2897	0.2755	0.4288	0.1839	0.6298*	-23.664**
Sqrt-RV	0.4458	0.4161	0.6468	0.4184	0.9976**	-19.022**
Log-RV	0.5223	0.5740	0.7034	0.4947	1.2525**	-16.269**
DS	0.8216	0.7811	0.9699	0.9373	1.1527**	-15.400**
TB	0.8993	0.9294	0.9926	0.9814	1.5911**	-8.2203*
PE	1.0610	0.9900	0.9969	0.9923	1.5836**	-4.9610

Table 1: Descriptive statistics. The summary statistics are provided for all variables using the full sample of monthly observations ($n = 662$). The variables are market realized variances (in levels, square-root, and logs), the default spread (DS), 3m T-bills (TB), and the price-earnings ratio (PE). Panel A shows unconditional summary statistics, whereas Panel B provides conditional summary statistics. Here, standard deviation, skewness, and excess kurtosis (relative to 3) are denoted “S.D.”, “Skew”, and “EKur”, respectively. The LW and LPWN semiparametric estimators of integration order d are implemented using bandwidths $m = n^{0.7}$ and $m = n^{0.9}$, respectively, following the recommendations in Frederiksen et al. (2012). If LPWN takes the values 0.0100 and 0.9900, this implies that the lower and upper bound of the parameter space have been hit. AR-ϕ and AR-R^2 are the estimated first-order autocorrelation coefficient and R^2. The MZ unit root test is based on GLS detrended data with the number of lags selected by the MAIC on OLS detrended data, as recommended by Perron & Qu (2007), see Ng & Perron (2001, Table 1) for tabulated critical values. The KPSS test for the processes being $I(0)$, that is, for obeying short memory dynamics, is implemented using the Newey & West (1994) automatic bandwidth selection procedure and the quadratic spectral kernel function, see Hobjin et al. (2004). It has 0.463 and 0.739 as 5% and 1% critical values. Finally, (*) and (**) denote rejection at a 5% and 1% significance level, respectively.

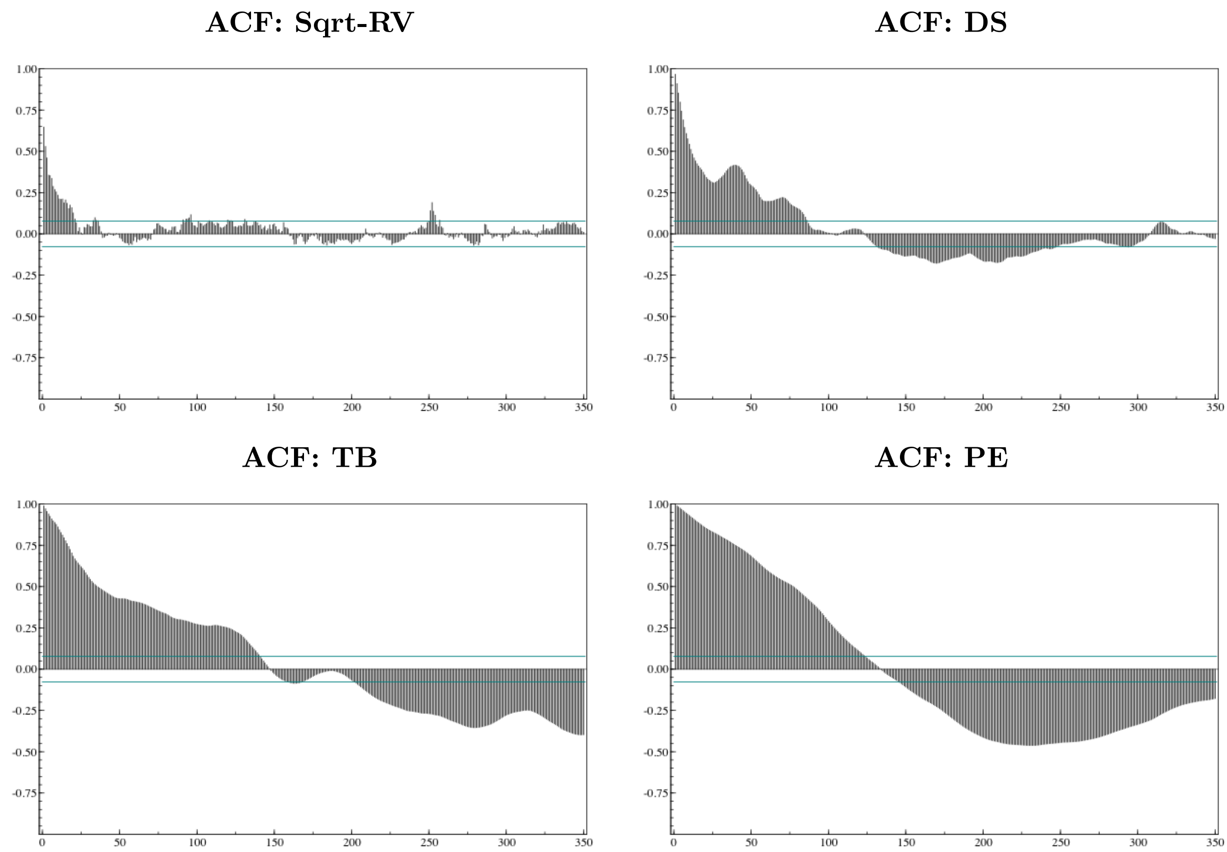

Figure 1: Autocorrelation functions. The sample autocorrelation functions are computed for the first 350 lags for each variable in the full sample, which spans the period from February 1960 through March 2015 ($n = 662$), where realized variance is estimated using daily log-returns. The variables are the square-root (sqrt) transformation of realized variance (RV), the default spread (DS), 3m T-bills (TB), and the price-earnings ratio (PE).

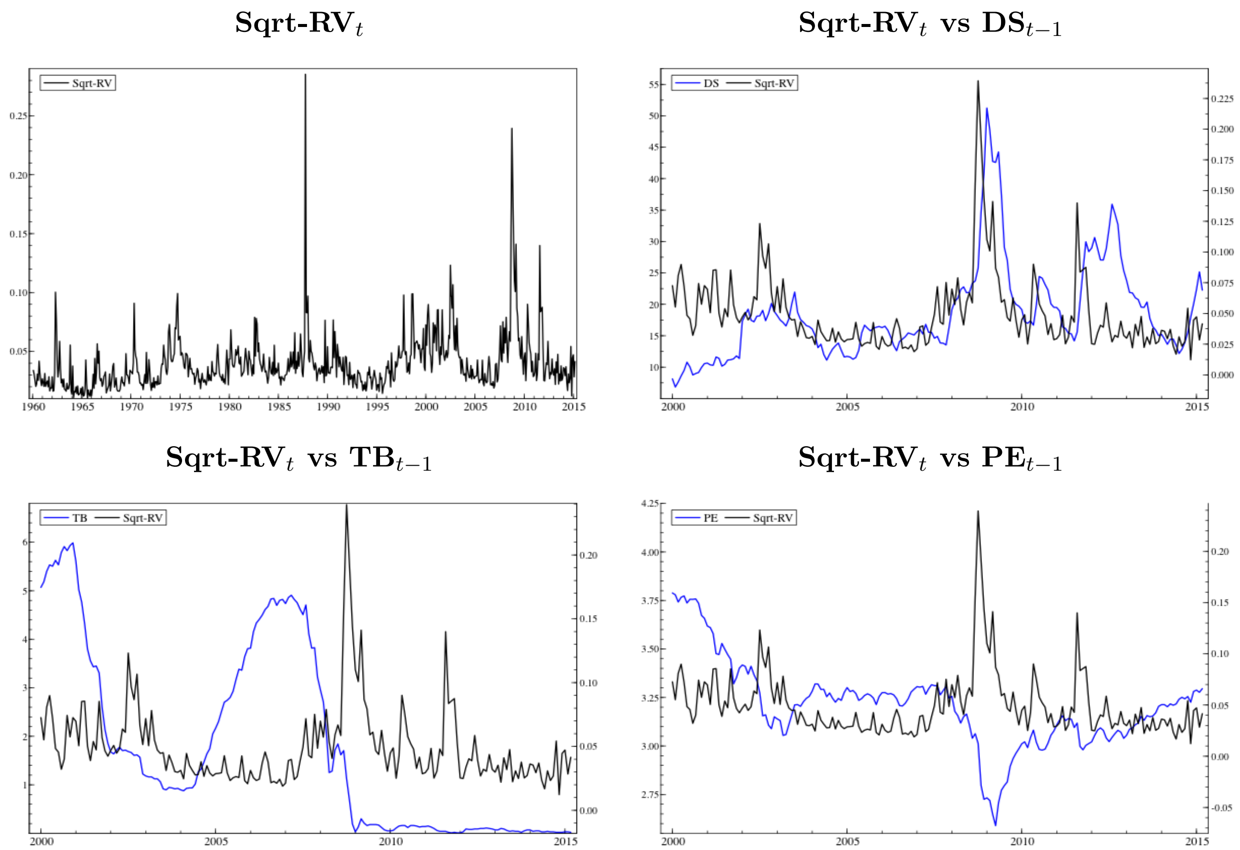

Figure 2: Plotted series. The upper left panel depicts the full sample of square-root transformed realized variance. The three remaining panels show the state variables (blue) along with the square-root RV estimates based on daily data (black) using the sample period January 2000 through March 2015 ($n = 183$). The left-hand scale is for the state variables, the right-hand scale for the square-root RV estimates.

Predictive RV Regressions								
	Panel A				Panel B			
Constant	0.0012	-0.0005	-0.0012	-0.0049	0.0011	0.0005	-0.0035	-0.0041
	(0.0004)	(0.0005)	(0.0005)	(0.0019)	(0.0002)	(0.0004)	(0.0013)	(0.0014)
RV_{t-1}	0.4288	0.3919	0.3918	0.3838	0.5939	0.5653	0.5628	0.5558
	(0.2005)	(0.1852)	(0.1849)	(0.1842)	(0.0679)	(0.0730)	(0.0712)	(0.0726)
DS_{t-1}	-	0.0099	0.0099	0.0148	-	0.0046	0.0062	0.0082
		(0.0047)	(0.0047)	(0.0059)		(0.0037)	(0.0036)	(0.0043)
TB_{t-1}	-	-	-	0.0179	-	-	-	0.0091
				(0.0083)				(0.0079)
PE_{t-1}	-	-	0.0385	0.1185	-	-	0.1186	0.1203
			(0.0220)	(0.0425)			(0.0400)	(0.0383)
Adj. R^2	0.1826	0.1945	0.1944	0.1979	0.3504	0.3527	0.3539	0.3527
$\hat{d}_u$	0.1876	0.1315	0.1299	0.1032	0.0826	0.0510	0.0283	0.0147
	(0.0516)	(0.0516)	(0.0516)	(0.0516)	(0.0680)	(0.0680)	(0.0680)	(0.0680)
Wald	-	4.4044	6.0757	8.7854	-	1.5673	11.282	12.573
$\mathbb{P}$-Wald	-	0.9642	0.9521	0.9677	-	0.7894	0.9965	0.9943

Table 2: OLS estimates and tests. We report least squares coefficient estimates and corresponding Newey & West (1987) standard errors or the variables along with the adjusted R^2, a local Whittle (LW) estimate of the residual memory parameter, and a Wald test and its associated $\mathbb{P}$-value for whether the state variables are jointly significant. Specifically, Panel A reports results from the full sample where realized variance is estimated using daily log-returns, and Panel B using a subsample from February 1990 through March 2015 ($n = 302$) where high-frequency data is utilized. The LW estimator is implemented using a bandwidth $m = \lfloor n^{0.7} \rfloor$. Note that the coefficients in front of the state variables have been scaled with 100.

Size and Power of the Local Spectrum Test								
	DGP 1				DGP 2			
	$n = 300$		$n = 650$		$n = 300$		$n = 650$	
Implementation:	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$
$(\nu, \kappa) = (0.21, 0.799)$	9.20	61.00	6.30	89.60	8.80	54.80	6.40	83.70
$(\nu, \kappa) = (0.25, 0.799)$	8.40	60.40	6.70	90.70	8.50	55.60	6.40	83.70
$(\nu, \kappa) = (0.30, 0.799)$	7.10	60.40	6.50	90.90	7.80	54.80	7.50	85.10
$(\nu, \kappa) = (0.25, 0.70)$	8.00	47.00	5.90	74.30	9.10	43.20	6.20	66.40
$(\nu, \kappa) = (0.25, 0.75)$	8.70	54.60	6.30	84.60	8.20	49.00	6.20	76.90
	DGP 3				DGP 4			
	$n = 300$		$n = 650$		$n = 300$		$n = 650$	
Implementation:	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$
$(\nu, \kappa) = (0.21, 0.799)$	7.60	62.90	6.00	90.10	8.70	62.00	5.70	85.50
$(\nu, \kappa) = (0.25, 0.799)$	7.70	63.00	6.30	91.10	8.90	63.00	6.30	85.20
$(\nu, \kappa) = (0.30, 0.799)$	7.50	63.40	5.60	91.40	8.00	63.50	6.90	86.60
$(\nu, \kappa) = (0.25, 0.70)$	6.60	45.40	5.70	72.40	7.70	44.90	5.40	66.80
$(\nu, \kappa) = (0.25, 0.75)$	7.20	54.70	5.70	84.10	7.90	53.80	6.00	77.20

Table 3: Size and power of the LCM test. This table displays the size ($\rho = 0$) and power ($\rho \neq 0$) of the proposed local spectrum test from Theorem 2, $\mathcal{LCM}(\ell, m)$, as a function of the MBLS trimming and bandwidth parameters, defined by $\ell = n^{\nu}$ and $m = n^{\kappa}$, respectively. As described in Section 5, the tuning parameters are fixed according to the asymptotic theory and the DGPs are simulated as in Hong (1996) and Shao (2009). Specifically, two (possibly, correlated) fractional ARMA(1,0) processes are simulated with $\phi_y = \phi_x = 0.2$ and varying fractional integration orders d_y and d_x. DGP 1 is configured with memory parameters $(d_y, d_x) = (0.30, 0.45)$; DGP 2 with $(d_y, d_x) = (0.30, 0.80)$; DGP 3 with $(d_y, d_x) = (0.55, 0.45)$; and DGP 4 with $(d_y, d_x) = (0.55, 0.80)$. The fractional filtering in the first step of the local spectrum procedure is based on ARFIMA parameter estimates of the memory parameter, where one AR lag has been included; see Hualde & Robinson (2011) and Nielsen (2015). All tests are implemented with $\nu_G = 0.25$ and $\kappa_G = 0.9$. Two sample sizes are considered, $n = \{300, 650\}$, corresponding well with the respective sizes of the subsample and full sample, see Tables 1 and 2. All tests are implemented with a nominal size of 5%. The simulations are carried out with 1,000 replications.

Test Size: LCM versus OLS												
DGP M1												
	$n = 300$				$n = 650$				$n = 1000$			
RV_{t-1}	100	100	100	100	100	100	100	100	100	100	100	100
DS_{t-1}	-	27.70	23.60	21.50	-	31.70	30.50	28.40	-	33.50	31.60	32.00
TB_{t-1}	-	-	-	10.80	-	-	-	12.90	-	-	-	14.00
PE_{t-1}	-	-	23.50	22.20	-	-	30.70	28.40	-	-	32.40	33.20
Adj. $\bar{R}^2$	30.80	31.25	31.62	31.93	32.74	33.03	33.29	33.53	33.29	33.48	33.69	33.89
OLS-Wald	-	27.70	37.80	44.30	-	31.70	45.40	54.20	-	33.50	47.40	59.80
$\mathcal{LCM}(0.21, 0.799)$	-	8.80	9.80	12.70	-	6.40	8.20	6.90	-	6.40	5.10	4.80
$\mathcal{LCM}(0.25, 0.799)$	-	8.50	9.80	13.10	-	6.40	9.00	8.70	-	6.40	5.50	5.40
$\mathcal{LCM}(0.30, 0.799)$	-	7.80	9.70	13.30	-	7.50	10.30	9.90	-	6.30	6.00	6.30
DGP M2												
	$n = 300$				$n = 650$				$n = 1000$			
RV_{t-1}	100	100	100	100	100	100	100	100	100	100	100	100
DS_{t-1}	-	33.50	30.40	26.70	-	37.30	37.00	36.10	-	39.90	39.70	41.00
TB_{t-1}	-	-	-	15.10	-	-	-	20.00	-	-	-	21.30
PE_{t-1}	-	-	34.00	29.80	-	-	40.30	39.60	-	-	42.10	41.80
Adj. $\bar{R}^2$	70.06	70.26	70.46	70.63	75.03	75.13	75.24	75.35	76.44	76.51	76.59	76.67
OLS-Wald	-	33.50	48.50	59.30	-	37.30	55.20	68.10	-	39.90	59.80	73.20
$\mathcal{LCM}(0.21, 0.799)$	-	8.70	9.80	10.50	-	5.70	7.40	6.90	-	6.30	4.20	4.70
$\mathcal{LCM}(0.25, 0.799)$	-	8.90	9.20	10.40	-	6.30	8.00	7.90	-	6.40	4.50	5.10
$\mathcal{LCM}(0.30, 0.799)$	-	8.00	9.40	11.10	-	6.90	9.20	9.20	-	6.50	5.20	5.90

Table 4: Size of LCM and OLS tests. This table displays power of the realized variance coefficient estimates (RV_{t-1}) using OLS and the corresponding size ($\rho = 0$) of the coefficient estimates for up to three exogenous predictor variables, which are simulated as x_t in equation (23). Moreover, the average adjusted R^2 is shown along with a Wald test for joint significance of the predictors. The inference and Wald tests are based on Newey & West (1987) covariance estimates. In addition, the the proposed local spectrum test from Theorem 2, $\mathcal{LCM}(\nu, \kappa)$, of the joint significance of the exogenous predictors is shown as a function of the trimming and bandwidth *rates*, defined by $\ell = n^\nu$ and $m = n^\kappa$, respectively. All processes are fractional ARMA$(1, 0)$ with an AR(1) parameter of 0.2 and varying fractional integration orders. The predictors are simulated with $d = 0.8$ (labelled by DS_{t-1}), $d = 1$ (PE_{t-1}), and $d = 0.9$ (TB_{t-1}). Moreover, realized variance is simulated with either $d_y = 0.3$ (DGP M1) or $d_y = 0.55$ (DGP M2). The fractional filtering in the first step of the local spectrum procedure is based on ARFIMA parameter estimates of the memory parameter, where one AR lag has been included; see Hualde & Robinson (2011) and Nielsen (2015). The LCM tests are further implemented with $\nu_G = 0.25$ and $\kappa_G = 0.9$. Three sample sizes are considered, $n = \{300, 650, 1000\}$, corresponding well with the respective sizes of the subsample and full sample, see Tables 1 and 2. All tests are implemented with a nominal size of 5%. The simulations are carried out with 1,000 replications.

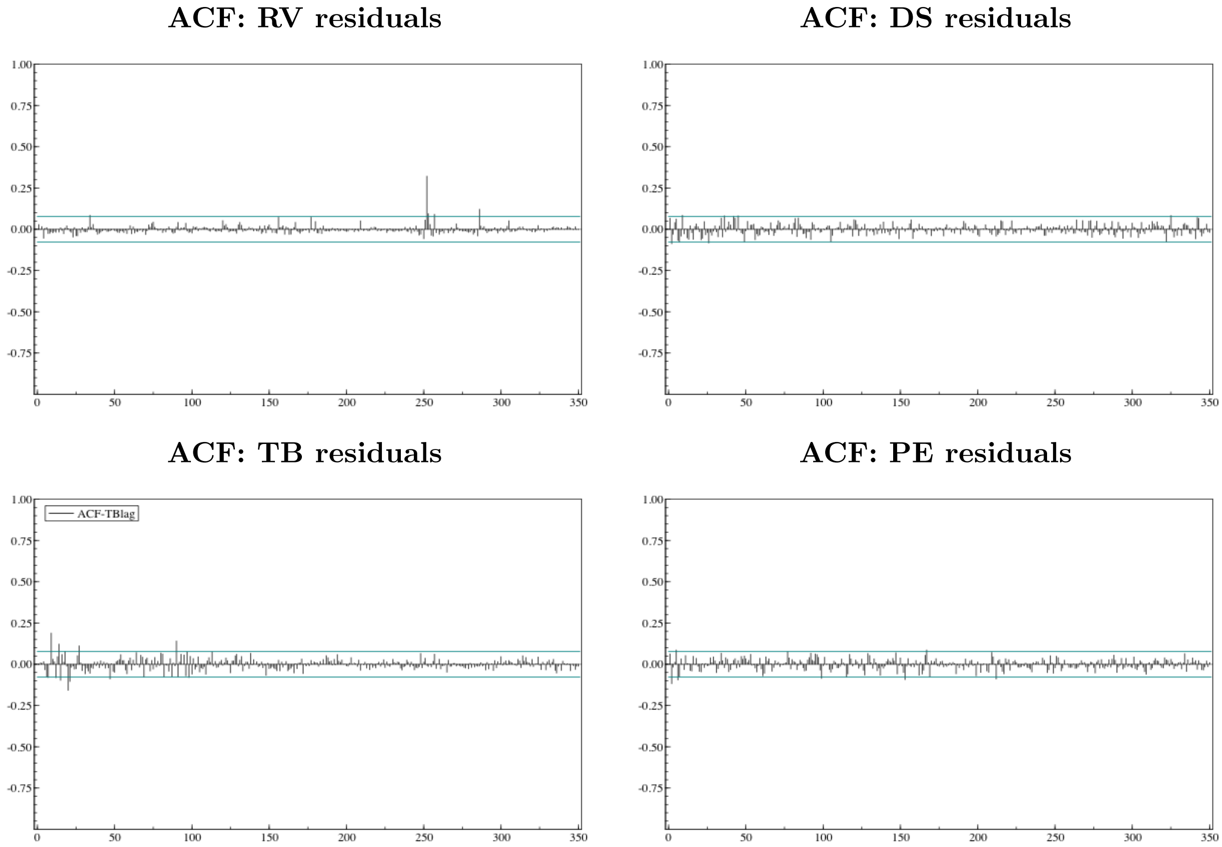

Figure 3: Autocorrelation functions. The sample autocorrelation functions are computed for the first 350 lags for the residual series of each variable after applying the ARFIMA filter to estimate the fractional integration order in the full sample, which spans the period from February 1960 through March 2015 ($n = 662$). The variables are the realized variance (RV), the default spread (DS), 3m T-bills (TB), and the price-earnings ratio (PE). Note that the fractional filter uses ten observations for initialization.

Local Spectrum Estimates and Tests: RV Predictions									
	$\mathcal{H}_1, \nu =$			$\mathcal{H}_2, \nu =$			$\mathcal{H}_3, \nu =$		
Panel A	0.21	0.25	0.30	0.21	0.25	0.30	0.21	0.25	0.30
DS_{t-1}	0.0120	0.0115	0.0101	0.0134	0.0136	0.0105	0.0142	0.0148	0.0116
TB_{t-1}	-	-	-	-	-	-	0.0123	0.0174	0.0141
PE_{t-1}	-	-	-	0.1008	0.1721	0.0321	0.0944	0.1628	0.0312
Wald	0.9252	0.8462	0.6534	1.1287	1.3660	0.6791	1.2019	1.5026	0.7795
$\mathbb{P}$-Wald	0.6639	0.6424	0.5811	0.4313	0.4949	0.2879	0.2475	0.3183	0.1456
	$\mathcal{H}_1, \nu =$			$\mathcal{H}_2, \nu =$			$\mathcal{H}_3, \nu =$		
Panel B	0.21	0.25	0.30	0.21	0.25	0.30	0.21	0.25	0.30
DS_{t-1}	0.0009	-0.0009	-0.0025	0.0030	0.0005	0.0013	-0.0003	-0.0015	0.0007
TB_{t-1}	-	-	-	-	-	-	-0.1294	-0.09978	-0.0286
PE_{t-1}	-	-	-	0.2178	0.1116	0.3694	0.2062	0.1245	0.3627
Wald	0.0041	0.0043	0.0304	0.3638	0.0989	1.0906	0.8880	0.4661	1.0609
$\mathbb{P}$-Wald	0.0510	0.0520	0.1384	0.1663	0.0483	0.4203	0.1717	0.0737	0.2135

Table 5: Local spectrum estimates and tests. We report coefficient estimates from the local spectrum procedure to predictability testing as well as corresponding Wald test statistics and $\mathbb{P}$-values. Specifically, Panel A reports results from the full sample where realized variance is estimated using daily log-returns, and Panel B using a subsample from February 1990 through March 2015 ($n = 302$) where high-frequency data is utilized. The LCM procedure is implemented using bandwidths determined by $\kappa = 0.799$ and $\kappa_G = 0.9$ as well as the trimming parameters $\nu = \{0.21, 0.25, 0.30\}$ and $\nu_G = 0.25$. The series are fractionally filtered using ARFIMA estimates of the fractional integration orders, which are consistent at rate $n^{-1/2}$. The selection of the ARMA polynomials and properties of the ARFIMA filters are discussed in the Sections 5-6 and Figure 3. The fractional filter uses ten observations for initialization. The three test statistics $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$ uses the DS, the DS and PE, or all three variables as predictors and test their joint predictive power. All parameter estimates are scaled with 100.

IVX Estimates and Tests: RV Predictions									
	$\mathcal{H}_1, C_z =$			$\mathcal{H}_2, C_z =$			$\mathcal{H}_3, C_z =$		
Panel A	-1	-5	-10	-1	-5	-10	-1	-5	-10
DS_{t-1}	0.0110	0.0122	0.0116	0.0119	0.0136	0.0124	0.0152	0.0122	0.0084
TB_{t-1}	-	-	-	-	-	-	0.0112	-0.0049	-0.0133
PE_{t-1}	-	-	-	0.0727	0.0333	0.0050	0.1178	-0.0018	-0.0761
Wald	0.6836	0.6263	0.4162	2.6633	0.8268	0.4959	3.0477	2.1215	3.1843
$\mathbb{P}$-Wald	0.5917	0.5713	0.4812	0.7360	0.3386	0.2196	0.6157	0.4524	0.6359
	$\mathcal{H}_1, C_z =$			$\mathcal{H}_2, C_z =$			$\mathcal{H}_3, C_z =$		
Panel B	-1	-5	-10	-1	-5	-10	-1	-5	-10
DS_{t-1}	0.0089	0.0114	0.0111	0.0118	0.0026	-0.0058	0.0127	-0.0082	-0.0214
TB_{t-1}	-	-	-	-	-	-	0.0335	0.1085	0.1315
PE_{t-1}	-	-	-	0.0906	-0.2853	-0.6180	0.0264	-0.8430	-1.5487
Wald	0.2959	0.3445	0.2455	0.2829	1.9290	2.2620	-0.1255	0.3235	0.2704
$\mathbb{P}$-Wald	0.4135	0.4428	0.3798	0.1319	0.6188	0.6773	0.0000	0.0445	0.0345

Table 6: IVX estimates and tests. We report the coefficient estimates from the IVX regression procedure to predictability testing as well as corresponding Wald test statistics and $\mathbb{P}$-values. Specifically, Panel A reports results from the full sample where realized variance is estimated using daily log-returns, and Panel B using a subsample from February 1990 through March 2015 ($n = 302$) where high-frequency data is utilized. The procedure is implemented using $\beta_z = 0.95$ and $C_z = \{-1, -5, -10\}$ as well as with a bias-correction and using HAC estimates of the asymptotic covariance matrix, as recommended by Kostakis et al. (2015) and Phillips & Lee (2016). Details are provided in Appendix D. The three test statistics $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$ uses the DS, the DS and PE, or all three variables as predictors and test their joint predictive power. All parameter estimates are scaled with 100.

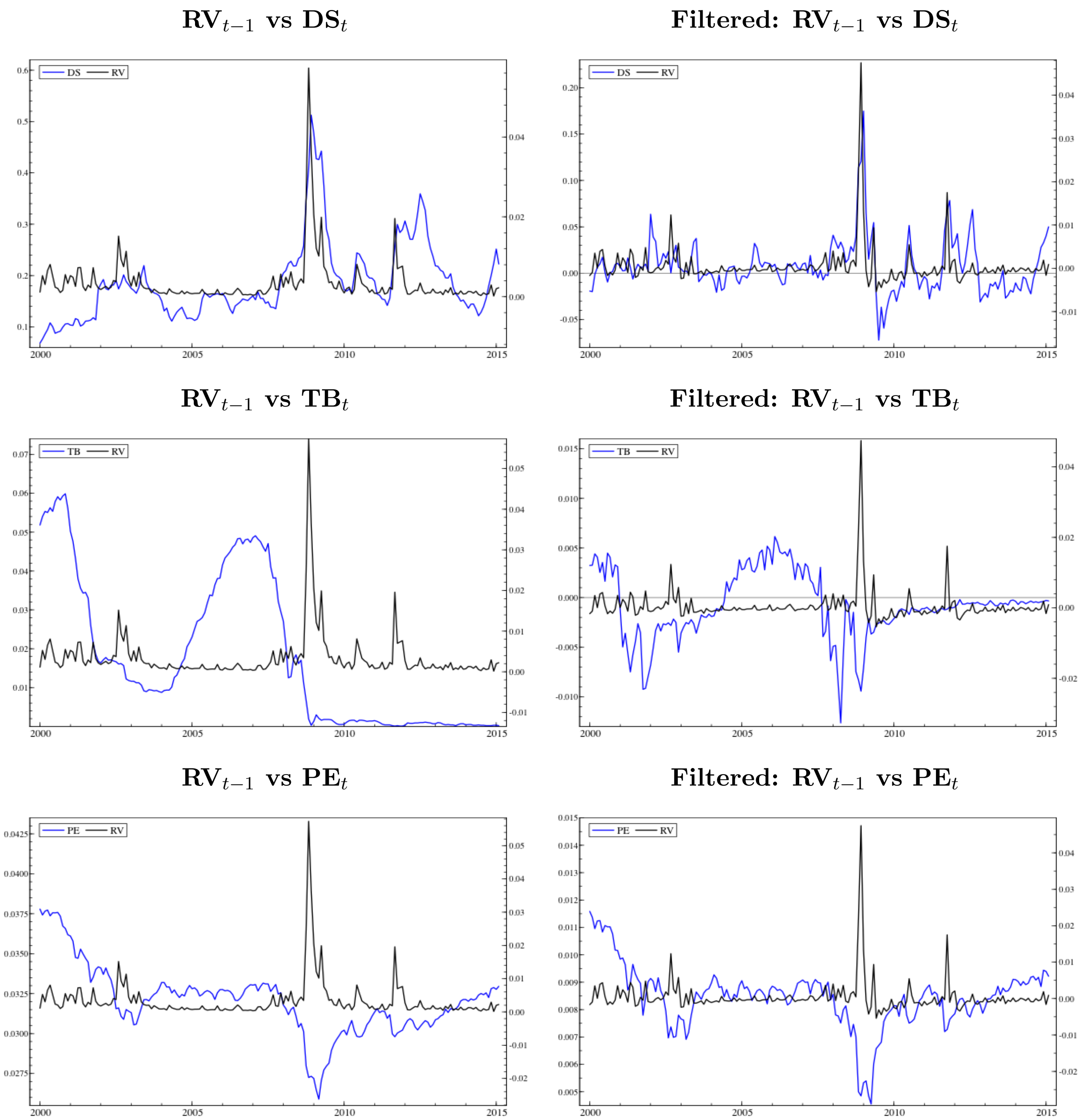

Figure 4: Plotted series. The left panels depicts the realized variance (RV) based on daily data (black) against the state variables (blue) along for the sample period January 2000 through February 2015 ($n = 182$). The right panels show the corresponding series after fractional filtering. In each plot, the left-hand scale is for the state variables (divided by 100, as for the parameter estimates), the right-hand scale for the RV estimates.

Local Spectrum Estimates and Tests: Reverse Causality									
	DS_t, $\nu =$			TB_t, $\nu =$			PE_t, $\nu =$		
Panel A	0.21	0.25	0.30	0.21	0.25	0.30	0.21	0.25	0.30
RV_{t-1}	2.4285	2.4190	2.4093	-0.1672	-0.1539	-0.1676	-0.0616	-0.0590	-0.0711
Wald	120.27	119.48	118.67	7.5211	6.3851	7.5553	27.384	25.210	36.091
$\mathbb{P}$-Wald	1.0000	1.0000	1.0000	0.9939	0.9885	0.9940	1.0000	1.0000	1.0000
	DS_t, $\nu =$			TB_t, $\nu =$			PE_t, $\nu =$		
Panel B	0.21	0.25	0.30	0.21	0.25	0.30	0.21	0.25	0.30
RV_{t-1}	3.0160	2.9736	2.9391	-0.1685	-0.1537	-0.1377	-0.0558	-0.0707	-0.0606
Wald	66.043	64.513	63.268	26.892	22.564	18.259	13.578	21.451	15.904
$\mathbb{P}$-Wald	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.9999
	DS_t, $\nu =$			TB_t, $\nu =$			PE_t, $\nu =$		
Panel C	0.21	0.25	0.30	0.21	0.25	0.30	0.21	0.25	0.30
RV_{t-1}	0.5019	0.4657	0.4735	-0.1415	-0.0859	-0.1021	-0.0445	-0.0384	-0.0749
Wald	7.9791	6.8780	7.1069	2.4612	0.9098	1.2827	7.5614	5.6453	20.767
$\mathbb{P}$-Wald	0.9953	0.9913	0.9923	0.8833	0.6598	0.7426	0.9940	0.9825	1.0000

Table 7: Local spectrum estimates and tests. We report coefficient estimates from the local spectrum procedure to predictability testing as well as corresponding Wald test statistics and $\mathbb{P}$-values. Specifically, Panel A reports results from the full sample where realized variance is estimated using daily log-returns, Panel B using a subsample from February 1990 through February 2015 ($n = 301$) where high-frequency data is utilized, and Panel C shows results from a subsample analysis based on RV data constructed using daily log-returns that ends in December 2007 ($n = 574$). The LCM procedure is implemented using bandwidths determined by $\kappa = 0.799$ and $\kappa_G = 0.9$ as well as the trimming parameters $\nu = \{0.21, 0.25, 0.30\}$ and $\nu_G = 0.25$. The series are fractionally filtered using ARFIMA estimates of the fractional integration orders, which are consistent at rate $n^{-1/2}$. The selection of the ARMA polynomials and properties of the ARFIMA filters are discussed in the Section 5-6 and Figure 3. The fractional filter uses ten observations for initialization. All parameter estimates are scaled with 100.

References

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), 'Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility', *The Review of Economics and Statistics* **89**(4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Andrews, D. W. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**(3), 817–858.

Arteche, J. (2004), 'Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models', *Journal of Econometrics* **119**, 131–154.

Baillie, R., Bollerslev, T. & Mikkelsen, H. (1996), 'Fractionally integrated generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **74**, 3–30.

Baillie, R. T. & Bollerslev, T. (1994), 'Cointegration, fractional cointegration, and exchange rate dynamics', *Journal of Finance* **49**, 737–745.

Bandi, F. M. & Perron, B. (2006), 'Long memory and the relation between implied and realized volatility', *Journal of Financial Econometrics* **4**(4), 636–670.

Bansal, R., Kiku, D., Shaliastovich, I. & Yaron, A. (2014), 'Volatility, the macroeconomy and asset prices', *The Journal of Finance* **LXIX**, 2471–2511.

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society Series B* **64**, 253–280.

Bollerslev, T., Osterrieder, D., Sizova, N. & Tauchen, G. (2013), 'Risk and return: Long-run relationships, fractional cointegration, and return predictability', *Journal of Financial Economics* **108**, 409–424.

Brillinger, D. R. (1981), *Time Series. Data Analysis and Theory*, Siam: Classics in Applied Mathematics.

Campbell, J. Y., Giglio, S., Polk, C. & Turley, R. (2017), 'An intertemporal CAPM with stochastic volatility', *Journal of Financial Economics* **forthcoming**.

Campbell, J. Y. & Viceira, L. M. (2002), *Strategic Asset Allocation: Portfolio Choice for Long-term Investors*, Oxford University Press, Oxford, United Kingdom.

Campbell, J. Y. & Vuolteenaho, T. (2004), 'Bad beta, good beta', *American Economic Review* **94**, 1249–1275.

Campbell, J. Y. & Yogo, M. (2006), 'Efficient tests of stock return predictability', *Journal of Financial Economics* **81**, 27–60.

Cavanagh, C., Elliott, G. & Stock, J. (1995), 'Inference in models with nearly integrated regressors', *Econometric Theory* **11**, 1131?1147.

Cheung, Y. E. & Lai, K. S. (1993), 'A fractional cointegration analysis of purchasing power parity', *Journal of Business and Economic Statistics* **11**, 103–122.

Christensen, B. J. & Nielsen, M. O. (2006), 'Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting', *Journal of Econometrics* **133**, 343–371.

Christensen, B. J. & Varneskov, R. T. (2017), 'Medium band least squares estimation of fractional cointegration in the presence of low-frequency contamination', *Journal of Econometrics* **197**, 218–244.

Christiansen, C., Schmeling, M. & Schrimpf, A. (2012), 'A comprehensive look at financial volatility prediction by economic variables', *Journal of Applied Econometrics* **27**, 956–977.

Chung, C.-F. (2002), 'Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes', *Econometric Theory* **18**, 51–78.

Comte, F. & Renault, E. (1998), 'Long memory in continuous-time stochastic volatility models', *Mathematical Finance* **8**, 291–323.

Conrad, C. & Loch, K. (2014), 'Anticipating long-term stock market volatility', *Journal of Applied Econometrics* **30**, 1090–1114.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**, 174–196.

Davidson, J. (2002), *Stochastic Limit Theory*, Oxford University Press.

Deng, A. (2014), 'Understanding spurious regressions in financial economics', *Journal of Financial Econometrics* **12**, 122–150.

Deo, R. S. & Hurvich, C. M. (2001), 'On the log-periodogram regression estimator of the memory parameter in long memory stochastic volatility models', *Econometric Theory* **17**, 686–710.

Dew-Becker, I., Giglio, S., Le, A. & Rodriguez, M. (2017), 'The price of variance risk', *Journal of Financial Economics* **132**, 225–250.

Duecker, M. & Startz, R. (1998), 'Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates', *The Review of Economics and Statistics* **83**, 420–426.

Elliott, G., Müller, U. & Watson, M. (2015), 'Nearly optimal tests when a nuisance parameter is present under the null hypothesis', *Econometrica* **83**, 771–811.

Ferson, W. E., Sarkissian, S. & Simin, T. (2003), 'Spurious regressions in financial economics', *Journal of Finance* **58**, 1393–1414.

Ferson, W. E., Sarkissian, S. & Simin, T. (2009), Spurious regression and data mining in conditional asset pricing models, *in* C. F. Lee & A. C. Lee, eds, 'Handbook of Quantitative Finance', Springer.

Frederiksen, P. H., Nielsen, F. S. & Nielsen, M. O. (2012), 'Local polynomial whittle estimation of perturbed fractional processes', *Journal of Econometrics* **167**, 426–447.

Granger, C. V. J. & Newbold, P. (1974), 'Spurious regression in econometrics', *Journal of Econometrics* **2**, 111–120.

Granger, C. W. J. & Morris, M. J. (1976), 'Time series modelling and interpretation', *Journal of the Royal Statistical Society. Series A.* **139**, 246–257.

Haldrup, N. & Nielsen, M. O. (2007), 'Estimation of fractional integration in the presence of data noise', *Computational Statistics & Data Analysis* **51**, 3100–3114.

Hall, P. & Heyde, C. C. (1980), *Martingale Limit Theory and Its Appplication*, Boston: Academic Press.

Hobjin, B., Franses, P. H. & Ooms, M. (2004), 'Generalizations of the KPSS-test for stationarity', *Statistica Neerlandica* **58**, 483–502.

Hong, Y. (1996), 'Testing for independence between two covariance stationary time series', *Biometrika* **83**, 615–625.

Hualde, J. & Robinson, P. M. (2011), 'Gaussian pseudo-maximum likelihood estimation of fractional time series models', *Annals of Statistics* **39**, 3152–3181.

Jansson, M. & Moreira, M. J. (2006), 'Optimal inference in regression models with nearly integrated regressors', *Econometrica* **74**, 681–714.

Johansen, S. & Nielsen, M. O. (2012), 'Likelihood inference for a fractionally cointegrated vector autoregressive model', *Econometrica* **80**, 2667–2732.

Kostakis, A., Magdalinos, T. & Stamatogiannis, M. P. (2015), 'Robust econometric inference for stock return predictability', *Review of Financial Studies* **28**, 1506–1553.

Künsch, H. (1987), 'Statistical aspects of self-similar processes.', pp. 67–74. In: Prohorov, Y., Sazarov, E. (Eds.), Proceedings of the First World Congress of the Bernoulli Society. Vol 1. VNU Science Press, Utrecht.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. & Shin, Y. (1992), 'Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?', *Journal of Econometrics* **54**, 159–178.

Lobato, I. (1997), 'Consistency of averaged cross-periodogram in long memory series', *Journal of Time Series Analysis* **18**, 137–155.

Lobato, I. (1999), 'A semiparametric two-step estimator in a multivariate long memory model', *Journal of Econometrics* **90**, 129–155.

Magdalinos, T. & Phillips, P. C. B. (2009), Econometric inference in the vicinity of unity. CoFie Working Paper (7), Singapore Management University.

Maynard, A., Smallwood, A. & Wohar, M. E. (2013), 'Long memory regressors and predictive testing: A two-stage rebalancing approach', *Econometric Reviews* **32**, 318–360.

Mittnik, S., Robinzonov, N. & Spindler, M. (2015), 'Stock market volatility: Identifying major drivers and the nature of their impact', *Journal of Banking and Finance* **58**, 1–14.

Müller, U. & Watson, M. (2016), 'Measuring uncertainty about long-run predictions', *Review of Economic Studies* **83**, 1711–1740.

Müller, U. & Watson, M. (2017), 'Long-run covariability', *Econometrica* **forthcoming**.

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

Newey, W. K. & West, K. D. (1994), ‘Automatic lag selection in covariance matrix estimation’, *Review of Economic Studies* **64**, 631–654.

Ng, S. & Perron, P. (2001), ‘Lag length selection and the construction of unit root tests with good size and power’, *Econometrica* **6**, 1519–1554.

Nielsen, M. O. (2005), ‘Semiparametric estimation in time-series regression with long-range dependence’, *Journal of Time Series Analysis* **26**, 279–304.

Nielsen, M. O. (2015), ‘Asymptotics for the conditional-sum-of-squares estimator in mutivariate fractional time series models’, *Journal of Time Series Analysis* **36**, 154–188.

Nonejad, N. (2017), ‘Forecasting aggregate stock market volatility using financial and macroeconomic predictors: Which models forecast best, when and why?’, *Journal of Empirical Finance* **42**, 131–154.

Pastor, L. & Stambaugh, R. F. (2009), ‘Predictive systems: Living with imperfect predictors’, *Journal of Finance* **64**, 1583–1628.

Paye, B. S. (2012), ‘Déjà vol’: Predictive regressions for aggregate stock market volatility using macroeconomic variables’, *Journal of Financial Economics* **106**, 527–546.

Perron, P. & Qu, Z. (2007), ‘A simple modification to improve the finite sample properties of Ng and Perron’s unit root tests’, *Economics Letters* **94**, 12–19.

Peseran, M. H. & Timmermann, A. (1995), ‘Predictability of stock returns: Robustness and economic significance’, *Journal of Finance* **50**, 1201–1228.

Phillips, P. C. B. (1986), ‘Understanding spurious regressions in econometrics’, *Journal of Econometrics* **33**, 311–340.

Phillips, P. C. B. (2014), ‘On confidence intervals for autoregressive roots and predictive regression’, *Econometrica* **82**, 1177–1195.

Phillips, P. C. B. & Lee, J. H. (2013), ‘Predictive regression under various degrees of persistence and robust long-horizon regression’, *Journal of Econometrics* **177**, 250–264.

Phillips, P. C. B. & Lee, J. H. (2016), ‘Robust econometric inference with mixed integrated and mildly explosive regressors’, *Journal of Econometrics* **192**, 433–450.

Phillips, P. C. & Hansen, B. E. (1990), ‘Statistical inference in instrumental variables regression with i(1) variables’, *Review of Economic Studies* **57**, 99–125.

Phillips, P. C. & Shimotsu, K. (2004), ‘Local whittle estimation in nonstationary and unit root cases’, *The Annals of Statistics* **32**, 656–692.

Robinson, P. M. (1995), ‘Gaussian semiparametric estimation of long range dependence’, *The Annals of Statistics* **23**, 1630–1661.

Robinson, P. M. (2008), ‘Multiple local Whittle estimation in stationary systems’, *Annals of Statistics* **36**, 2508–2530.

Robinson, P. M. & Hidalgo, F. J. (1997), ‘Time series regression with long-range dependence’, *The Annals of Statistics* **25**, 77–104.

Robinson, P. M. & Marinucci, D. (2001), 'Narrow-band analysis of nonstationary processes', *The Annals of Statistics* **29**, 947–986.

Robinson, P. M. & Marinucci, D. (2003), 'Semiparametric frequency domain analysis of fractional cointegration'. In: Robinson, P.M. (Ed.), Time Series with Long Memory. Oxford University Press, Oxford, pp. 334-373.

Schwert, G. W. (1989), 'Why does stock market volatility change over time', *Journal of Finance* **44**, 1115–1153.

Shao, X. (2009), 'A generalized portmanteau test for independence between two stationary time series', *Econometric Theory* **25**, 195–210.

Shiller, R. (2000), *Irrational Exuberance*, Princeton University Press, United States.

Shimotsu, K. (2007), 'Gaussian semiparametric estimation of multivariate fractionally integrated processes', *Journal of Econometrics* **137**, 277–310.

Shimotsu, K. (2010), 'Exact local whittle estimation of fractional integration with unkown mean and time trend', *Econometric Theory* **26**, 501–540.

Shimotsu, K. & Phillips, P. C. (2005), 'Exact local whittle estimation of fractional integration', *The Annals of Statistics* **32**, 656–692.

Shimotsu, K. & Phillips, P. C. B. (2006), 'Local whittle estimation of fractional integration and some of its variants', *Journal of Econometrics* **130**, 209–233.

Sizova, N. (2013), 'Long-horizon return regressions with historical volatility and other long-memory variables', *Journal of Business & Economic Statistics* **31**, 546–559.

Stambaugh, R. F. (1999), 'Predictive regressions', *Journal of Financial Economics* **54**, 783–820.

Torous, W., Valkanov, R. & Yan, S. (2005), 'On predicting stock returns with nearly integrated explanatory variables', *Journal of Business* **77**, 937–966.

Tsay, W.-J. & Chung, C.-F. (2000), 'The spurious regression of fractionally integrated processes', *Journal of Econometrics* **96**, 155–182.

Valkanov, R. (2003), 'Long-horizon regressions: Theoretical results and applications', *Journal of Financial Economics* **68**, 201–232.

Varneskov, R. T. (2016), 'Flat-top realized kernel estimation of quadratic covariation with nonsynchronous and noisy asset prices', *Journal of Business and Economic Statistics* **31**(1), 1–22.

Varneskov, R. T. (2017), 'Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels', *Econometric Theory* **33**(6), 1457–1501.

Welch, I. & Goyal, A. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *Review of Financial Studies* **21**, 1455–1508.

A Assumptions

Instead of stating the assumptions in terms of $\boldsymbol{v}_t$, we follow the frequency domain frameworks in, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Christensen & Varneskov (2017), by imposing the regularity conditions in terms of $\boldsymbol{q}_t = (\boldsymbol{u}_t', \eta_t)'$ when deriving the central limit theory for the proposed local spectrum inference and testing procedures.

Assumption D1. *The vector process $\boldsymbol{q}_t$, $t = 1, \ldots,$ is covariance stationary with spectral density matrix satisfying $\boldsymbol{f}_{qq}(\lambda) \sim \boldsymbol{G}_{qq}$ as $\lambda \to 0^+$, where the upper left $k \times k$ submatrix, $\boldsymbol{G}_{uu}$, has full rank, and the $(k+1)$th element of the diagonal, $G_{\eta\eta}$, is strictly greater than zero. Moreover, there exists a $\varpi \in (0, 2]$ such that $|\boldsymbol{f}_{qq}(\lambda) - \boldsymbol{G}_{qq}| = O(\lambda^{\varpi})$ as $\lambda \to 0^+$. Finally, let $\boldsymbol{G}_{qq}(i, k+1)$ be the $(i, k+1)$th element of $\boldsymbol{G}_{qq}$, which has $\boldsymbol{G}_{qq}(i, k+1) = \boldsymbol{G}_{qq}(k+1, i) = 0$ for all $i = 1, \ldots, k$.*

Assumption D2. *$\boldsymbol{q}_t$ is a linear process, $\boldsymbol{q}_t = \sum_{j=0}^{\infty} \boldsymbol{A}_j \boldsymbol{\epsilon}_{t-j}$, with square summable coefficients $\sum_{j=0}^{\infty} \|\boldsymbol{A}_j\|^2 < \infty$, the innovations satisfy, almost surely, $\mathbb{E}[\boldsymbol{\epsilon}_t | \mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' | \mathcal{F}_{t-1}] = \boldsymbol{I}_{k+1}$, and the matrices $\mathbb{E}[\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' | \mathcal{F}_{t-1}]$ and $\mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' \otimes \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' | \mathcal{F}_{t-1}]$ are nonstochastic, finite, and do not depend on t, with $\mathcal{F}_t = \sigma(\boldsymbol{\epsilon}_s, s \leq t)$. There exists a random variable ζ such that $\mathbb{E}[\zeta^2] < \infty$ and for all c and some C, $\mathbb{P}[\|\boldsymbol{q}_t\| > c] \leq C\mathbb{P}[|\zeta| > c]$. Finally, the periodogram of $\boldsymbol{\epsilon}_t$ is denoted by $\boldsymbol{J}(\lambda)$.*

Assumption D3. *For $\boldsymbol{A}(\lambda, i)$, the ith row of $\boldsymbol{A}(\lambda) = \sum_{j=0}^{\infty} \boldsymbol{A}_j e^{\mathrm{i}j\lambda}$, its partical derivative satisfies $\|\partial \boldsymbol{A}(\lambda, i) \partial \lambda\| = O(\lambda^{-1} \|\boldsymbol{A}(\lambda, i)\|)$ as $\lambda \to 0^+$, for $i = 1, \ldots, k+1$.*

These assumptions are standard in the literature. The main departure from the references above is the use of the exact spectral density representation in (9).

Assumption C. *Suppose $\boldsymbol{c}_t = \boldsymbol{c}_t \mathbf{1}_{\{t \geq 1\}}$ is a mean-zero $k \times 1$ vector process satisfying the same conditions as $\boldsymbol{u}_t$ in Assumptions D1-D3, but with a co-spectrum $\boldsymbol{f}_{c\eta}(\lambda) \sim \boldsymbol{G}_{c\eta}$ as $\lambda \to 0^+$ that allows the constant vector $\boldsymbol{G}_{c\eta}$ to have non-zero entries. Moreover, let $\boldsymbol{u}_t \perp\!\!\!\perp \boldsymbol{c}_s$ for all $t, s \geq 1$.*

Assumption C allows us to treat $\boldsymbol{c}_t$ similarly to $\boldsymbol{u}_t$ in the proofs with one key difference, its local co-spectral coherence with the regression errors η_t is non-trivial.

B Proofs

This section contains the proofs of the main asymptotic results in the paper as well as some technical results in Section B.4. Before proceeding, however, let us introduce some notation. For a generic vector $\boldsymbol{V}$, let $\boldsymbol{V}(i)$ index the ith element, and, similarly, for a matrix $\boldsymbol{M}$, let $\boldsymbol{M}(i, q)$ denote its (i, q)th element. Moreover, denote by $K \in (0, \infty)$ a generic constant, which may take different values from line to line or from (in)equality to (in)equality. Finally, we remark that sometimes the (stochastic) orders refer to scalars, sometimes to vectors and matrices. We refrain from making distinctions.

B.1 Proof of Theorem 1

First, write $\hat{\boldsymbol{v}}_t = \hat{\boldsymbol{D}}(L)\boldsymbol{D}(L)^{-1}\boldsymbol{D}(L)\boldsymbol{z}_t\mathbf{1}_{\{t\geq 1\}}$, and define the terms $\hat{\boldsymbol{A}}(L) \equiv \hat{\boldsymbol{D}}(L)\boldsymbol{D}(L)^{-1}$ as well as $\boldsymbol{a}_t \equiv \boldsymbol{D}(L)\boldsymbol{z}_t\mathbf{1}_{\{t\geq 1\}}$, noticing that by definition $\boldsymbol{a}_t = \boldsymbol{v}_t + \boldsymbol{D}(L)\boldsymbol{\mu}\mathbf{1}_{\{t\geq 1\}}$, with $\boldsymbol{\mu}$ corresponding to the means, or initial values, of $\boldsymbol{z}_t$. Moreover, let $\tilde{\boldsymbol{\mu}}_t \equiv \boldsymbol{D}(L)\boldsymbol{\mu}\mathbf{1}_{\{t\geq 1\}}$ and denote by $\tilde{\mu}_t^{(e)}$ the first element of the vector and by $\tilde{\boldsymbol{\mu}}_t^{(u)}$ the remaining $k \times 1$ vector. Next, decompose the trimmed discretely averaged periodogram of the fractionally filtered sequence $\hat{\boldsymbol{v}}_t$,

$$\begin{aligned}\hat{\boldsymbol{F}}_{\hat{v}\hat{v}}(\ell, m) - \hat{\boldsymbol{F}}_{vv}(1, m) = \Big(\hat{\boldsymbol{F}}_{vv}(\ell, m) - \hat{\boldsymbol{F}}_{vv}(1, m)\Big) + \Big(\hat{\boldsymbol{F}}_{aa}(\ell, m) - \hat{\boldsymbol{F}}_{vv}(\ell, m)\Big) \\ + \Big(\hat{\boldsymbol{F}}_{\hat{v}\hat{v}}(\ell, m) - \hat{\boldsymbol{F}}_{aa}(\ell, m)\Big) \equiv \boldsymbol{\mathcal{E}}_1 + \boldsymbol{\mathcal{E}}_2 + \boldsymbol{\mathcal{E}}_3. \qquad \text{(B.1)}\end{aligned}$$

This decomposition is crucial for showing that the first-stage filtering errors and mean-slippage only have an asymptotically negligible impact on the second stage MBLS estimate when trimming of frequencies in the vicinity of the origin. Specifically, since we can write by addition and subtraction,

$$\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) - \boldsymbol{\mathcal{B}} = \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell, m)^{-1}\hat{\boldsymbol{F}}_{u\eta}(1, m) - \boldsymbol{\mathcal{C}}_1 + \boldsymbol{\mathcal{C}}_2 + \boldsymbol{\mathcal{C}}_3 \qquad \text{(B.2)}$$

where the three error terms, $\boldsymbol{\mathcal{C}}_1$, $\boldsymbol{\mathcal{C}}_2$, and $\boldsymbol{\mathcal{C}}_3$, are defined as

$$\begin{aligned}&\boldsymbol{\mathcal{C}}_1 \equiv \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell, m)^{-1}\hat{\boldsymbol{F}}^{(u)}_{\hat{u}\tilde{\mu}}(\ell, m)\boldsymbol{\mathcal{B}}, \qquad \boldsymbol{\mathcal{C}}_2 \equiv \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell, m)^{-1}\hat{\boldsymbol{F}}^{(e)}_{\hat{u}\tilde{\mu}}(\ell, m), \\ &\boldsymbol{\mathcal{C}}_3 \equiv \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell, m)^{-1}\Big(\hat{\boldsymbol{F}}_{\hat{u}\eta}(\ell, m) - \hat{\boldsymbol{F}}_{u\eta}(\ell, m) + \boldsymbol{\mathcal{E}}_4\Big), \qquad \boldsymbol{\mathcal{E}}_4 \equiv \hat{\boldsymbol{F}}_{u\eta}(\ell, m) - \hat{\boldsymbol{F}}_{u\eta}(1, m),\end{aligned}$$

with the super superscripts indicating $\tilde{\boldsymbol{\mu}}_t^{(u)}$ and $\tilde{\mu}_t^{(e)}$, respectively, the decomposition (B.1) allows us to establish asymptotic bounds on $\boldsymbol{\mathcal{C}}_1$, $\boldsymbol{\mathcal{C}}_2$, and $\boldsymbol{\mathcal{C}}_3$. The proof is now concluded by the following three auxiliary lemmas: Lemma B.1 establishes generic bounds for $\boldsymbol{\mathcal{E}}_1$, $\boldsymbol{\mathcal{E}}_2$, $\boldsymbol{\mathcal{E}}_3$ and $\boldsymbol{\mathcal{E}}_4$. Lemma B.2 uses these bounds to show $\sqrt{m}(\boldsymbol{\mathcal{C}}_1 + \boldsymbol{\mathcal{C}}_2 + \boldsymbol{\mathcal{C}}_3) = o_p(1)$ and $\lambda_m^{-1}\hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell, m) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}$. Finally, Lemma B.3 establishes central limit theory for $\sqrt{m}\lambda_m^{-1}\hat{\boldsymbol{F}}_{u\eta}(1, m)$, i.e., the discretely averaged co-periodogram of $\boldsymbol{u}_{t-1}$ and η_t. Hence, by using these lemmas in conjunction with the continuous mapping theorem and Slutsky's theorem, this provides the stated central limit theory for $\sqrt{m}(\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) - \boldsymbol{\mathcal{B}})$.

Finally, we need to provide conditions on $\varpi \in (0, 2]$ such that the second rate restriction in Assumption T is mutually consistent with all values $0 \leq d_i < 2$, $i = 1, \ldots, k+1$. The second restriction implies $\nu < (\varpi + \kappa/2)/(\varpi + 1)$. Moreover, the worst bound on the trimming rate ν in the fourth restriction is obtained for $\underline{d} = 0$: $\nu > (1 - \kappa/2)/2$. These assumptions are, thus, always mutually consistent when $\max(0, (1 - 3/2\kappa)/(1 + \kappa/2)) < \varpi$ is imposed, as stated in the theorem. Hence, a solution is always guaranteed to exist since $\max(0, (1 - 3/2\kappa)/(1 + \kappa/2)) < 2$. □

Lemma B.1 (Asymptotic Bounds)**.** *Under the conditions for Theorem 1, then, for some arbitrarily small $\epsilon > 0$, the following asymptotic bounds hold:*

(a) $\lambda_m^{-1}\boldsymbol{\mathcal{E}}_1 = O_p\left(m^{-1}\ell\right)$.

(b) $\sqrt{m}\lambda_m^{-1}\boldsymbol{\mathcal{E}}_2 \leq O_p((m/n)^{2\underline{d}}nm^{-1/2}\ell^{-2}) + O_p((m/n)^{\underline{d}-1/2}m^{\epsilon}\ell^{-(1+\epsilon)})$.

(c) Let $\bar{f}(\ell,m,n) \equiv \frac{m}{n} \vee \frac{m^{\epsilon}}{n^{1/2}\ell^{1+\epsilon}} \vee \frac{1}{\ell^2}$ for some arbitrarily small $\epsilon > 0$, then

$$\sqrt{m}\lambda_m^{-1}\boldsymbol{\mathcal{E}}_3 \leq O_p\left(\frac{\ln(n)^2\sqrt{n}}{m_d\sqrt{m}}\bar{f}(\ell,m,n)\right) + O_p\left(\frac{\ln(n)\sqrt{n}}{\sqrt{m_d}}\sqrt{\bar{f}(\ell,m,n)}\right).$$

(d) $\sqrt{m}\lambda_m^{-1}\boldsymbol{\mathcal{E}}_4 \leq O_p\left(\ell^{1+\varpi}/(m^{1/2}n^{\varpi})\right)$.

Proof. **For (a).** First, by Assumptions D1-D3, we may apply the same arguments as in Christensen & Varneskov (2017, Equations (B.3)-(B.7)) to show that when $\ell, n \to \infty$, $\ell/n \to 0$,

$$-\boldsymbol{\mathcal{E}}_1 = \frac{2\pi}{n}\sum_{j=1}^{\ell-1}\Re\left(\boldsymbol{I}_{vv}(\lambda_j)\right), \quad \frac{2\pi}{n}\sum_{j=1}^{\ell-1}\Re\left(\boldsymbol{I}_{uu}(\lambda_j)\right) \sim \boldsymbol{G}_{uu}\lambda_\ell, \qquad \text{as} \quad \lambda_\ell \to 0^+, \tag{B.3}$$

uniformly in probability. Now, by invoking the properties of the matrix $\boldsymbol{G}_{qq}$ in Assumption D1 and since the parameter vector $\boldsymbol{\mathcal{B}}$ in (10) is constant, this readily establishes **(a)**.

For (b). First, make the decomposition,

$$\boldsymbol{\mathcal{E}}_2 = \hat{\boldsymbol{F}}_{\tilde{\mu}\tilde{\mu}}(\ell,m) + \hat{\boldsymbol{F}}_{v\tilde{\mu}}(\ell,m) + \hat{\boldsymbol{F}}_{\tilde{\mu}v}(\ell,m), \tag{B.4}$$

utilizing that $\tilde{\boldsymbol{a}}_t = \boldsymbol{v}_t + \tilde{\boldsymbol{\mu}}_t$. Now, for the first term in (B.4), use the bound for the periodogram of a fractionally differenced constant from Shimotsu (2010, Lemma B.2), see also Lemma B.6(a) below, to deduce the following stochastic order for the (i,i)th diagonal element, $i = 1, \ldots, k+1$,

$$\hat{\boldsymbol{F}}_{\tilde{\mu}\tilde{\mu}}(\ell,m,i,i) = \frac{2\pi}{n}\sum_{j=\ell}^{m}\Re\left(\boldsymbol{I}_{\tilde{\mu}\tilde{\mu}}(\lambda_j,i,i)\right) = \frac{2\pi}{n}\sum_{j=\ell}^{m}O_p\left(\frac{n^{1-2d_i}}{j^{2-2d_i}}\right) \leq \left(\frac{m}{n}\right)^{2d_i}\sum_{j=\ell}^{m}O_p\left(j^{-2}\right), \tag{B.5}$$

where, for the partial sum term, $\mathcal{S}(\ell) = \sum_{j=\ell}^{m}O_p\left(j^{-2}\right)$, we may invoke Varneskov (2017, Lemma C.4) to show $|\mathcal{S}(\ell)| \leq O_p(\ell^{-2})$, since the power series j^{-q} readily has $q > 1$. Combining these results with the Cauchy-Schwarz inequality, this readily establishes $\hat{\boldsymbol{F}}_{\tilde{\mu}\tilde{\mu}}(\ell,m) \leq O_p((m/n)^{2\underline{d}}\ell^{-2})$. For the second term in the decomposition (B.4), use the Cauchy-Schwarz inequality and Assumption D1 in conjunction with the same arguments as for (B.5) to show

$$\begin{aligned}\hat{\boldsymbol{F}}_{v\tilde{\mu}}(\ell,m,i,i) &= \frac{2\pi}{n}\sum_{j=\ell}^{m}\Re\left(\boldsymbol{I}_{v\tilde{\mu}}(\lambda_j,i,i)\right) = \frac{2\pi}{n}\sum_{j=\ell}^{m}O_p(1)\times O_p\left(\frac{n^{1/2-d_i}}{j^{1-d_i}}\right)\\ &= \frac{2\pi}{\sqrt{n}}\sum_{j=\ell}^{m}O_p\left(\frac{j^{d_i+\epsilon}}{n^{d_i}}\frac{1}{j^{1+\epsilon}}\right) \leq O_p\left(\frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{d_i}\frac{m^{\epsilon}}{\ell^{1+\epsilon}}\right),\end{aligned} \tag{B.6}$$

uniformly, for some arbitrarily small $\epsilon > 0$. Hence, by applying the Cauchy-Schwarz inequality, it follows that $\hat{\boldsymbol{F}}_{v\tilde{\mu}}(\ell,m) + \hat{\boldsymbol{F}}_{\tilde{\mu}v}(\ell,m) \leq O_p(n^{-1/2}(m/n)^{\underline{d}}m^{\epsilon}\ell^{-(1+\epsilon)})$, establishing **(b)**.

For (c). First, write $\hat{\boldsymbol{v}}_t = \hat{\boldsymbol{A}}(L)\boldsymbol{a}_t$, whose ith element is given by $\hat{v}_t(i) = (1-L)^{\hat{\theta}_i} a_t(i)$ with power defined as $\hat{\theta}_i = \hat{d}_i - d_i$, and where $\hat{\theta}_i = O_p(m_d^{-1/2})$ by Assumption F. Now, using a Taylor expansion of the fractional filter $(1-L)^{\hat{\theta}_i}$ around $\hat{\theta}_i = 0$ and the mean-value theorem, we have

$$(1-L)^{\hat{\theta}_i} = 1 + \hat{\theta}_i \ln(1-L) + \frac{\hat{\theta}_i^2}{2} \ln(1-L)^2 (1-L)^{\bar{\theta}_i} \tag{B.7}$$

for some $\bar{\theta}_i \in [\hat{\theta}_i, 0]$ and with $\bar{\theta}_i = O_p(m_d^{-1/2})$. Hence, by defining $a_{t-1}^{(1)}(i) \equiv \ln(1-L)a_t(i)$ and the second order term $a_{t-2}^{(2)}(i, \bar{\theta}_i) \equiv \ln(1-L)^2(1-L)^{\bar{\theta}_i} a_t(i)$, we can make the decomposition,

$$\hat{v}_t(i) = a_t(i) + \hat{\theta}_i a_{t-1}^{(1)}(i) + \frac{\hat{\theta}_i^2}{2} a_{t-2}^{(2)}(i, \bar{\theta}_i) \equiv a_t(i) + \bar{a}_{t-1}^{(1)}(i) + \bar{a}_{t-2}^{(2)}(i), \tag{B.8}$$

implying that we may further decompose the (i,i)th element of $\boldsymbol{\mathcal{E}}_3$ as

$$\begin{aligned} \boldsymbol{\mathcal{E}}_3(i,i) &= \hat{\boldsymbol{F}}_{\hat{v}\hat{v}}(\ell, m, i, i) - \hat{\boldsymbol{F}}_{aa}(\ell, m, i, i) \qquad \text{(B.9)} \\ &= \hat{\boldsymbol{F}}_{\bar{a}\bar{a}}^{(1,1)}(\ell, m, i, i) + \hat{\boldsymbol{F}}_{\bar{a}\bar{a}}^{(2,2)}(\ell, m, i, i) + 2\hat{\boldsymbol{F}}_{\bar{a}\bar{a}}^{(1,2)}(\ell, m, i, i) + 2\hat{\boldsymbol{F}}_{a\bar{a}}^{(1)}(\ell, m, i, i) + 2\hat{\boldsymbol{F}}_{a\bar{a}}^{(2)}(\ell, m, i, i) \end{aligned}$$

for $i = 1, \ldots, k+1$, with the first and second term in the decomposition being the trimmed discretely average periodograms of $\bar{a}_{t-1}^{(1)}(i)$ and $\bar{a}_{t-2}^{(2)}(i)$, respectively, the third term is their co-periodogram, and the fourth and fifth terms are their respective co-periodograms with $a_t(i)$.

Now, for the first term of $\boldsymbol{\mathcal{E}}_3(i,i)$, we have

$$\begin{aligned} \hat{\boldsymbol{F}}_{\bar{a}\bar{a}}^{(1,1)}(\ell, m, i, i) &= \frac{2\pi\hat{\theta}_i^2}{n} \sum_{j=\ell}^{m} \Re\left(\boldsymbol{I}_{aa}^{(1,1)}(\lambda_j, i, i)\right) \\ &= O_p\left(\frac{m}{n}\frac{\ln(n)^2}{m_d}\right) + \frac{\ln(n)^2}{n^{1/2}m_d} \sum_{j=\ell}^{m} O_p\left(\frac{j^\epsilon}{j^{1+\epsilon}}\right) + \frac{\ln(n)^2}{m_d} \sum_{j=\ell}^{m} O_p\left(\frac{1}{j^2}\right) \\ &\leq O_p\left(\frac{m}{n}\frac{\ln(n)^2}{m_d}\right) + O_p\left(\frac{\ln(n)^2}{n^{1/2}m_d}\frac{m^\epsilon}{\ell^{1+\epsilon}}\right) + O_p\left(\frac{\ln(n)^2}{m_d}\frac{1}{\ell^2}\right), \end{aligned} \tag{B.10}$$

for some arbitrarily small $\epsilon > 0$, using Assumption F and Lemma B.7(a) in conjunction with Varneskov

(2017, Lemma C.4), as in (B.5) and (B.6). For the second term of $\boldsymbol{\mathcal{E}}_3(i,i)$,

$$\begin{aligned}
\hat{\boldsymbol{F}}^{(2,2)}_{\bar{a}\bar{a}}(\ell,m,i,i) &= \frac{\pi\theta_i^4}{2n}\sum_{j=\ell}^{m}\Re\left(\boldsymbol{I}^{(2,2)}_{aa}(\lambda_j,i,i)\right)\\
&= O_p\left(\frac{m}{n}\frac{\ln(n)^4}{m_d^2}\right) + \frac{\ln(n)^5}{nm_d^2}\sum_{j=\ell}^{m}O_p\left(\frac{j^{1/2+\epsilon}}{j^{1+\epsilon}}\right) + \frac{\ln(n)^4}{n^{1/2}m_d^2}\sum_{j=\ell}^{m}O_p\left(\frac{j^{\epsilon}}{j^{1+\epsilon}}\right)\\
&+ \frac{\ln(n)^5}{n^{1/2}m_d^2}\sum_{j=\ell}^{m}O_p\left(\frac{1}{j^{3/2}}\right) + \frac{\ln(n)^6}{nm_d^2}\sum_{j=\ell}^{m}O_p\left(\frac{j^{\epsilon}}{j^{1+\epsilon}}\right) + \frac{\ln(n)^4}{m_d^2}\sum_{j=\ell}^{m}O_p\left(\frac{1}{j^2}\right)\\
&\leq O_p\left(\frac{m}{n}\frac{\ln(n)^4}{m_d^2}\right) + O_p\left(\frac{\ln(n)^5}{nm_d^2}\frac{m^{1/2+\epsilon}}{\ell^{1+\epsilon}}\right) + O_p\left(\frac{\ln(n)^4}{n^{1/2}m_d^2}\frac{m^{\epsilon}}{\ell^{1+\epsilon}}\right) + O_p\left(\frac{\ln(n)^4}{m_d^2}\frac{1}{\ell^2}\right) \qquad \text{(B.11)}
\end{aligned}$$

using similar arguments, $\ln(n)^6/n \ll \ln(n)^4/n^{1/2}$, and $\ln(n)/\ell^{3/2} \ll m^{\epsilon}/\ell^{1+\epsilon}$. Now, since Assumptions F and T readily imply $\ln(n)^p \ll m_d$ for some finite $p \in \mathbb{N}$, $m \ll n$, then (B.10) and (B.11) provide the bound $\hat{\boldsymbol{F}}^{(2,2)}_{\bar{a}\bar{a}} = o_p\big(\hat{\boldsymbol{F}}^{(1,1)}_{\bar{a}\bar{a}}(\ell,m,i,i)\big)$. Next, define

$$\bar{f}(\ell,m,n) \equiv \frac{m}{n} \vee \frac{m^{\epsilon}}{n^{1/2}\ell^{1+\epsilon}} \vee \frac{1}{\ell^2},$$

then we may write $\hat{\boldsymbol{F}}^{(1,1)}_{\bar{a}\bar{a}}(\ell,m,i,i) \leq O_p(\ln(n)^2/m_d\bar{f}(\ell,m,n))$. Hence, by **(a)**, **(b)**, (B.10) and (B.11) in conjunction with the Cauchy-Schwarz inequality and the continuous mapping theorem, we get the following bounds for the cross-product terms of $\boldsymbol{\mathcal{E}}_3(i,i)$,

$$\begin{aligned}
\hat{\boldsymbol{F}}^{(1,2)}_{\bar{a}\bar{a}}(\ell,m,i,i) &\leq \sqrt{\hat{\boldsymbol{F}}^{(1,1)}_{\bar{a}\bar{a}}(\ell,m,i,i)}\sqrt{\hat{\boldsymbol{F}}^{(2,2)}_{\bar{a}\bar{a}}(\ell,m,i,i)} = o_p\left(\hat{\boldsymbol{F}}^{(1,1)}_{\bar{a}\bar{a}}(\ell,m,i,i)\right),\\
\hat{\boldsymbol{F}}^{(1)}_{a\bar{a}}(\ell,m,i,i) &\leq \sqrt{\hat{\boldsymbol{F}}_{aa}(\ell,m,i,i)}\sqrt{\hat{\boldsymbol{F}}^{(1,1)}_{\bar{a}\bar{a}}(\ell,m,i,i)} \leq O_p\left(\frac{\ln(n)}{m_d^{1/2}}\frac{m^{1/2}}{n^{1/2}}\sqrt{\bar{f}(\ell,m,n)}\right),\\
\hat{\boldsymbol{F}}^{(2)}_{a\bar{a}}(\ell,m,i,i) &\leq \sqrt{\hat{\boldsymbol{F}}_{aa}(\ell,m,i,i)}\sqrt{\hat{\boldsymbol{F}}^{(2,2)}_{\bar{a}\bar{a}}(\ell,m,i,i)} = o_p\left(\hat{\boldsymbol{F}}^{(1)}_{a\bar{a}}(\ell,m,i,i)\right).
\end{aligned}$$

As the stochastic bounds for $\boldsymbol{\mathcal{E}}_3(i,i)$ are independent of $i = 1,\ldots,k+1$, we use the same arguments, the Cauchy-Schwarz inequality and the continuous mapping theorem to establish **(c)**.

For (d). First, similarly to **(a)**, write

$$-\boldsymbol{\mathcal{E}}_4 = \frac{2\pi}{n}\sum_{j=1}^{\ell-1}\Re\left(\boldsymbol{I}_{u\eta}(\lambda_j)\right), \qquad \text{(B.12)}$$

where we let $\boldsymbol{I}_{u\eta}(\lambda_j) \equiv \boldsymbol{I}^{(t,1)}_{u\eta}(\lambda_j)$ denote the co-periodogram of $\boldsymbol{u}_{t-1}$ and η_t. Now, let us similarly write $\boldsymbol{I}_{u\eta}^{(t,0)}(\lambda_j)$ for the corresponding periodogram of $\boldsymbol{u}_t$ and η_t. Then, by Lemma B.10,

$$-\sqrt{m}\lambda_m^{-1}\boldsymbol{\mathcal{E}}_4 = \frac{2\pi\sqrt{m}}{n\lambda_m}\sum_{j=1}^{\ell-1}\Re\left(\boldsymbol{I}_{u\eta}^{(t,0)}(\lambda_j)\right) + O_p\left(\frac{\ell\sqrt{m}}{m\sqrt{n}}\right) \equiv \sqrt{m}\lambda_m^{-1}\hat{\boldsymbol{F}}_{u\eta}^{(t,0)}(1,\ell-1) + o_p(1), \quad \text{(B.13)}$$

since $\ell/m \to 0$ and $m/n \to 0$ as $n \to \infty$, with $\hat{\boldsymbol{F}}_{u\eta}^{(t,0)}(1,\ell-1)$ being the TDAC of $\boldsymbol{I}_{u\eta}^{(t,0)}(\lambda_j)$. Now, to establish asymptotic bound for, $\sqrt{m}\lambda_m^{-1}\hat{\boldsymbol{F}}_{u\eta}^{(t,0)}(1,\ell-1)$, we rely on the Cramér-Wold Theorem, cf. Davidson (2002, Theorem 25.5). Specifically, for an arbitrary $k \times 1$ vector $\boldsymbol{\psi}$, write

$$\sqrt{m}\lambda_m^{-1}\boldsymbol{\psi}'\hat{\boldsymbol{F}}_{u\eta}^{(t,0)}(1,\ell-1) = \sum_{i=1}^{k}\frac{\psi_i 2\pi\sqrt{m}}{n\lambda_m}\sum_{j=1}^{\ell-1}\Re\left(\boldsymbol{I}_{u\eta}^{(t,0)}(\lambda_j,i) - \boldsymbol{A}(\lambda,i)\boldsymbol{J}(\lambda_j)\bar{\boldsymbol{A}}(\lambda,k+1)\right) + \sum_{i=1}^{k}\frac{\psi_i 2\pi\sqrt{m}}{n\lambda_m}\sum_{j=1}^{\ell-1}\Re\left(\boldsymbol{A}(\lambda,i)\boldsymbol{J}(\lambda_j)\bar{\boldsymbol{A}}(\lambda,k+1)\right) \equiv \boldsymbol{\mathcal{E}}_{41} + \boldsymbol{\mathcal{E}}_{42}.$$

For the first term, $\boldsymbol{\mathcal{E}}_{41}$ we may use summation by parts and Lobato (1999, C.2) to show

$$\boldsymbol{\mathcal{E}}_{41} \leq O_p\left(\sum_{i=1}^{k}\psi_i\frac{\sqrt{m}}{n\lambda_m}\left[\ell^{1/3}\ln(\ell)^{2/3} + \ln(\ell) + \frac{\ell^{1/2}}{n^{1/4}}\right]\right) = O_p\left(\frac{\ell^{1/3}\ln(\ell)^{2/3}}{m^{1/2}} + \frac{\ln(\ell)}{m^{1/2}} + \frac{\ell^{1/2}}{m^{1/2}n^{1/4}}\right),$$

showing $\boldsymbol{\mathcal{E}}_{41} \xrightarrow{\mathbb{P}} 0$ since $\ell/m \to 0$ as $n \to \infty$. Next, for $\boldsymbol{\mathcal{E}}_{42}$, we may use Assumptions D1-D3 to invoke the following asymptotic bounds

$$\boldsymbol{\mathcal{E}}_{42} \leq \sup_{i=1,\dots,k}\frac{K\sqrt{m}}{\lambda_m n}\sum_{j=1}^{\ell-1}|\boldsymbol{f}_{qq}(\lambda_j,i,k+1)| = O\left(\frac{\lambda_\ell^{\varpi}\ell\sqrt{m}}{\lambda_m n}\right) = O\left(\frac{\ell^{1+\varpi}}{m^{1/2}n^{\varpi}}\right), \quad \text{(B.14)}$$

utilizing that $\boldsymbol{G}_{qq}(i,k+1) = 0$ for all $i = 1,\dots,k$, thereby concluding the proof. □

Lemma B.2 (Central Limit Theory Errors)**.** *Under the conditions for Theorem 1, then*

(a) $\lambda_m^{-1}\hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}$.

(b) $\sqrt{m}\,(\boldsymbol{C}_1 + \boldsymbol{C}_2 + \boldsymbol{C}_3) = o_p(1)$.

Proof. Before proceeding to the proof, note that $\ell/m \to 0$ and $\ell^{1+\varpi}/(n^{\varpi}\sqrt{m}) \to 0$ in Assumption T implies $\lambda_m^{-1}\boldsymbol{\mathcal{E}}_1 = o_p(1)$ and $\sqrt{m}\lambda_m^{-1}\boldsymbol{\mathcal{E}}_4 \leq o_p(1)$. Moreover, we can write $\sqrt{m}\lambda_m^{-1}\boldsymbol{\mathcal{E}}_2 \leq \boldsymbol{\mathcal{E}}_{21} + \boldsymbol{\mathcal{E}}_{22}$, corresponding to the two asymptotic bounds in Lemma B.1(b). The last condition in Assumption T, then, implies $\boldsymbol{\mathcal{E}}_{21} = o_p(1)$ and $\boldsymbol{\mathcal{E}}_{21}^2 = o_p(1)$ such that $\sqrt{m}\lambda_m^{-1}\boldsymbol{\mathcal{E}}_2 \leq o_p(1)$ follows by the continuous mapping theorem and the Cauchy-Schwarz inequality. Similarly, write $\sqrt{m}\lambda_m^{-1}\boldsymbol{\mathcal{E}}_3 \leq \boldsymbol{\mathcal{E}}_{31} + \boldsymbol{\mathcal{E}}_{32}$, where

$$\boldsymbol{\mathcal{E}}_{31} = O_p\left(\frac{\ln(n)^2}{m_d} \vee \frac{\ln(n)^2 m^{\epsilon}}{m_d m^{1/2}\ell^{1+\epsilon}} \vee \frac{\ln(n)^2 n^{1/2}}{m_d m^{1/2}\ell^2}\right) = o_p(1),$$

trivially by Assumption T, and, similarly,

$$\boldsymbol{\mathcal{E}}_{32} = O_p\left(\frac{\ln(n)m^{1/2}}{m_d^{1/2}} \vee \frac{\ln(n)n^{1/2}}{m_d^{1/2}}\sqrt{\frac{m^{\epsilon}}{n^{1/2}\ell^{1+\epsilon}}} \vee \frac{\ln(n)n^{1/2}}{m_d^{1/2}\ell}\right) = o_p(1),$$

by $\kappa < \varrho$ and $(n/m_d)^{1/2}/\ell \to 0$ in Assumption T. Hence, all bounds in Lemma B.1 are $o_p(1)$. Now, proceeding to the proof. **For (a)**, the result follows by Lemmas B.1(a)-(c) in conjunction with the convergence result in Christensen & Varneskov (2017, (B.4)), see also Robinson & Marinucci (2003) and Lobato (1997). **For (b)**, $(\lambda_m^{-1}\hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m))^{-1} = O_p(1)$ follows by **(a)**, $\boldsymbol{G}_{uu}$ being full rank by Assumption D1 and the continuous mapping theorem. Moreover, since $\sqrt{m}\lambda_m^{-1}\hat{\boldsymbol{F}}^{(u)}_{\hat{u}\tilde{\mu}}(\ell,m) = o_p(1)$ by Lemmas B.1(b)-(c), we readily have that $\sqrt{m}\boldsymbol{\mathcal{C}}_1 = o_p(1)$ by Slutsky's theorem. The corresponding result for $\boldsymbol{\mathcal{C}}_2$ follows similarly. Finally, for $\boldsymbol{\mathcal{C}}_3$, we have

$$\boldsymbol{\mathcal{C}}_3 = \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m)^{-1}\left(\hat{\boldsymbol{F}}_{\hat{u}\eta}(\ell,m) - \hat{\boldsymbol{F}}_{u\eta}(\ell,m)\right) + \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m)^{-1}\left(\hat{\boldsymbol{F}}_{u\eta}(\ell,m) - \hat{\boldsymbol{F}}_{u\eta}(1,m)\right). \tag{B.15}$$

Hence, since $(\lambda_m^{-1}\hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell,m))^{-1} = O_p(1)$, as for $\boldsymbol{\mathcal{C}}_1$, as well as

$$\sqrt{m}\lambda_m^{-1}\left(\hat{\boldsymbol{F}}_{\hat{u}\eta}(\ell,m) - \hat{\boldsymbol{F}}_{u\eta}(\ell,m)\right) = o_p(1) \quad \text{and} \quad \sqrt{m}\lambda_m^{-1}\left(\hat{\boldsymbol{F}}_{u\eta}(\ell,m) - \hat{\boldsymbol{F}}_{u\eta}(1,m)\right) = o_p(1)$$

by Lemmas B.1(b)-(c) and Lemma B.1(d), $\sqrt{m}\boldsymbol{\mathcal{C}}_3 = o_p(1)$ follows by Slutsky's theorem. □

Lemma B.3 (Central Limit Theory)**.** *Under the conditions for Theorem 1, then*

$$\sqrt{m}\lambda_m^{-1}\boldsymbol{F}_{u\eta}(1,m) \xrightarrow{\mathbb{D}} N\left(0, G_{\eta\eta}\boldsymbol{G}_{uu}/2\right).$$

Proof. Before proceeding, let $\boldsymbol{I}_{u\eta}(\lambda_j) \equiv \boldsymbol{I}^{(t,1)}_{u\eta}(\lambda_j)$ and $\boldsymbol{F}_{u\eta}(1,m) \equiv \boldsymbol{F}^{(t,1)}_{u\eta}(1,m)$ denote the periodogram and TDAC for $\boldsymbol{u}_{t-1}$ and η_t, respectively. Moreover, denote by $\boldsymbol{I}^{(t,0)}_{u\eta}(\lambda_j)$ and $\boldsymbol{F}^{(t,0)}_{u\eta}(1,m)$ the corresponding quantities for $\boldsymbol{u}_t$ and η_t. Then, using Lemma B.10, we have

$$\sqrt{m}\lambda_m^{-1}\left(\boldsymbol{F}^{(t,1)}_{u\eta}(1,m) - \boldsymbol{F}^{(t,0)}_{u\eta}(1,m)\right) = \frac{2\pi\sqrt{m}}{n\lambda_m}\sum_{j=1}^{m}\Re\left(\boldsymbol{I}^{(t,1)}_{u\eta}(\lambda_j) - \boldsymbol{I}^{(t,0)}_{u\eta}(\lambda_j)\right) = O_p\left(\frac{m^{1/2}}{n^{1/2}}\right),$$

and, as a result, we may continue by establishing the central limit theory for $\sqrt{m}\lambda_m^{-1}\boldsymbol{F}^{(t,0)}_{u\eta}(1,m)$. To this end, we will apply the Cramér-Wold Theorem, cf. Davidson (2002, Theorem 25.5), for some arbitrary $k \times 1$ vector $\boldsymbol{\psi}$. First use Assumptions D1-D3 to decompose $\sqrt{m}\lambda_m^{-1}\boldsymbol{\psi}'\boldsymbol{F}^{(t,0)}_{u\eta}(1,m)$ similarly to the proof of Christensen & Nielsen (2006, Theorem 2),

$$\sqrt{m}\lambda_m^{-1}\boldsymbol{\psi}'\boldsymbol{F}^{(t,0)}_{u\eta}(1,m) = \sum_{i=1}^{k}\psi_i\sqrt{m}\lambda_m^{-1}\frac{2\pi}{n}\sum_{j=1}^{m}\Re\left(\boldsymbol{I}^{(t,0)}_{u\eta}(\lambda_j)\right) = \sum_{i=1}^{4}\mathcal{U}_i \tag{B.16}$$

where the four terms on the right-hand-side are defined as

$$\mathcal{U}_1 \equiv \sum_{i=1}^{k} \psi_i \sqrt{m} \lambda_m^{-1} \frac{2\pi}{n} \sum_{j=1}^{m} \Re\left(\boldsymbol{I}_{u\eta}^{(t,0)}(\lambda_j, i) - \boldsymbol{A}(\lambda_j, i) \boldsymbol{J}(\lambda_j) \bar{\boldsymbol{A}}(\lambda_j, k+1) \right),$$

$$\mathcal{U}_2 \equiv \sum_{i=1}^{k} \psi_i \sqrt{m} \lambda_m^{-1} \frac{1}{n} \sum_{j=1}^{m} \Re\left(\boldsymbol{A}(\lambda_j, i) \bar{\boldsymbol{A}}(\lambda_j, k+1) \right),$$

$$\mathcal{U}_3 \equiv \sum_{i=1}^{k} \psi_i \sqrt{m} \lambda_m^{-1} \frac{1}{n} \sum_{j=1}^{m} \Re\left(\boldsymbol{A}(\lambda_j, i) \left(\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' - \boldsymbol{I}_{k+1} \right) \bar{\boldsymbol{A}}(\lambda_j, k+1) \right),$$

$$\mathcal{U}_4 \equiv \sum_{i=1}^{k} \psi_i \sqrt{m} \lambda_m^{-1} \frac{1}{n} \sum_{j=1}^{m} \Re\left(\boldsymbol{A}(\lambda_j, i) \left(\frac{1}{n} \sum_{t=1}^{n} \sum_{s=1, s\neq t}^{n} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s' e^{\mathrm{i}(t-s)\lambda_j} \right) \bar{\boldsymbol{A}}(\lambda_j, k+1) \right).$$

We will now show that $\mathcal{U}_1$, $\mathcal{U}_2$ and $\mathcal{U}_3$ are asymptotically negligible, before establishing central limit theory for $\mathcal{U}_4$. For $\mathcal{U}_1$, we use summation by parts and Lobato (1999, C.2) to show

$$\mathcal{U}_1 \leq O_p\left(\sum_{i=1}^{k} \psi_i \frac{\sqrt{m}}{\lambda_m n} \left[m^{1/3} \ln(m)^{2/3} + \ln(m) + \frac{m^{1/2}}{n^{1/4}} \right] \right) = O_p\left(\frac{\ln(m)^{2/3}}{m^{1/6}} + \frac{\ln(m)}{m^{1/2}} + \frac{1}{n^{1/4}} \right),$$

and, hence, $\mathcal{U}_1 \xrightarrow{\mathbb{P}} 0$. Next, for $\mathcal{U}_2$, we may use Assumptions D1-D3 to invoke the bounds

$$\mathcal{U}_2 \leq \sup_{i=1,\dots,k} \frac{K\sqrt{m}}{\lambda_m n} \sum_{j=1}^{m} |\boldsymbol{f}_{qq}(\lambda_j, i, k+1)| = O\left(\frac{\lambda_m^{\varpi} m \sqrt{m}}{\lambda_m n} \right) = O\left(\frac{m^{1+2\varpi}}{n^{2\varpi}} \right),$$

utilizing that $\boldsymbol{G}_{qq}(i, k+1) = 0$ for all $i = 1, \dots, k$. This shows that $\mathcal{U}_2 \to 0$ by Assumption T. For the third term, $\mathcal{U}_3$, we have $\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' - \boldsymbol{I}_{k+1}$ being a martingale sequence with respect to the filtration $\mathcal{F}_{t-1}$, implying that the convergence result $n^{-1} \sum_{t=1}^{n} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' - \boldsymbol{I}_{k+1} = o_p(1)$ readily follows by Assumption D2. Hence, by Lebesque's dominated convergence theorem, we have $\mathcal{U}_3 \leq o_p(\mathcal{U}_2) \xrightarrow{\mathbb{P}} 0$.

Next, the final term in the decomposition (B.16) may be rewritten as

$$\mathcal{U}_4 = \sum_{t=1}^{n} \boldsymbol{\epsilon}_t' \sum_{s=1, s\neq t}^{n} \sum_{i=1}^{k} \psi_i \frac{\sqrt{m}}{\lambda_m n^2} \sum_{j=1}^{m} \Re\left(\boldsymbol{A}(\lambda_j, i)' e^{\mathrm{i}(t-s)\lambda_j} \bar{\boldsymbol{A}}(\lambda_j, k+1)' \right) \boldsymbol{\epsilon}_s = \sum_{t=1}^{n} \boldsymbol{\epsilon}_t' \sum_{s=1}^{t-1} \boldsymbol{C}_{t-s}^{n} \boldsymbol{\epsilon}_s$$

where the sequence of $(k+1) \times (k+1)$ coefficient matrices $\boldsymbol{C}_t^n$ is defined by

$$\boldsymbol{C}_t^n \equiv \frac{1}{2\pi n \sqrt{m}} \sum_{j=1}^{m} \boldsymbol{\chi}_j \cos(t\lambda_j), \quad \boldsymbol{\chi}_j \equiv \sum_{i=1}^{k} \psi_i \Re\left(\boldsymbol{A}(\lambda_j, i)' \bar{\boldsymbol{A}}(\lambda_j, k+1)' + \boldsymbol{A}(\lambda_j, k+1)' \bar{\boldsymbol{A}}(\lambda_j, i)' \right).$$

Now, since the components in the sum $\mathcal{M}_t \equiv \boldsymbol{\epsilon}_t' \sum_{s=1}^{t-1} \boldsymbol{C}_{t-s}^{n} \boldsymbol{\epsilon}_s$ are martingale difference sequences with respect to the filtration $\mathcal{F}_{t-1}$, we may establish the final central limit theory by showing

$$\sum_{t=1}^{n} \mathbb{E}[\mathcal{M}_t^2|\mathcal{F}_{t-1}] - \sum_{i=1}^{k}\sum_{p=1}^{k} \psi_i\psi_p G_{\eta\eta}\boldsymbol{G}_{uu}(i,p)/2 \xrightarrow{\mathbb{P}} 0, \quad \sum_{t=1}^{n} \mathbb{E}[\mathcal{M}_t^4] \to 0, \tag{B.17}$$

see, e.g., Hall & Heyde (1980, Chapter 3). From (B.17), there is no difference between the remaining arguments and those for Christensen & Nielsen (2006, Theorem 2), see their pp. 366-369 where we, in their notation, set $d_i = d_k = d_e = 0$. Hence, we invoke their arguments in conjunction with the continuous mapping theorem and Slutsky's theorem to establish the final result. □

B.2 Proof of Theorem 2

To establish the distribution result in Theorem 2, it suffices to show consistency of the asymptotic covariance estimators $\widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G, m_G)$ and $\widehat{\boldsymbol{G}}_{\hat{\eta}\hat{\eta}}(\ell_G, m_G)$ for $\boldsymbol{G}_{uu}$ and $G_{\eta\eta}$, respectively, since, in this case, the result follows by applying Theorem 1, the continuous mapping theorem and Slutsky's theorem. As consistency of the two is provided by the following lemma, this concludes the proof. □

Lemma B.4. *Under the conditions of Theorem 2,*

(a) $\widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G, m_G) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}$,

(b) $\widehat{\boldsymbol{G}}_{\hat{\eta}\hat{\eta}}(\ell_G, m_G) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{\eta\eta}$.

Proof. **For (a).** Similarly to the proof of Theorem 1, make a decomposition

$$\begin{aligned}\widehat{\boldsymbol{G}}_{\hat{v}\hat{v}}(\ell_G, m_G) &= \widehat{\boldsymbol{G}}_{vv}(\ell_G, m_G) + \Big(\widehat{\boldsymbol{G}}_{aa}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{vv}(\ell_G, m_G)\Big) + \Big(\widehat{\boldsymbol{G}}_{\hat{v}\hat{v}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{aa}(\ell_G, m_G)\Big) \\ &\equiv \boldsymbol{\mathcal{G}}_1 + \boldsymbol{\mathcal{G}}_2 + \boldsymbol{\mathcal{G}}_3,\end{aligned}$$

noticing, again, that $\hat{\boldsymbol{u}}_t$ comprises the last k elements of the vector $\hat{\boldsymbol{v}}_t$. Now, since the convergence result $\widehat{\boldsymbol{G}}_{uu}(\ell_G, m_G) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}$ follows by arguments similar to those for Christensen & Varneskov (2017, Lemma 6), it suffices to show that $\boldsymbol{\mathcal{G}}_2 = o_p(1)$ and $\boldsymbol{\mathcal{G}}_3 = o_p(1)$. To this end, and similarly to (B.4), let us first decompose the (i,i)th element of the matrix $\boldsymbol{\mathcal{G}}_2$ for $i = 1, \ldots, k+1$ as

$$\begin{aligned}\boldsymbol{\mathcal{G}}_2(i,i) &= \frac{1}{m_G - \ell_G + 1}\sum_{j=\ell_G}^{m_G} \left(\Re\left(\boldsymbol{I}_{\tilde{\mu}\tilde{\mu}}(\lambda_j, i, i)\right) + 2\Re\left(\boldsymbol{I}_{v\tilde{\mu}}(\lambda_j, i, i)\right)\right) \\ &= \frac{m_G}{m_G - \ell_G + 1}\frac{1}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{n^{1-2d_i}}{j^{2-2d_i}}\right) + \frac{m_G}{m_G - \ell_G + 1}\frac{1}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{n^{1/2-d_i}}{j^{1-d_i}}\right) \\ &\leq O_p\left(\frac{n}{m_G\ell_G^2}\left(\frac{m_G}{n}\right)^{2d_i}\right) + O_p\left(\frac{n^{1/2}}{m_G}\left(\frac{m_G}{n}\right)^{d_i}\frac{m_G^{\epsilon}}{\ell_G^{1+\epsilon}}\right)\end{aligned}$$

for some arbitrarily small $\epsilon > 0$, using the same arguments as for (B.5) and (B.6). Hence, by using these bounds for all $i = 1, \ldots, k+1$ diagonal elements in conjunction with the Cauchy-Schwarz inequality and Assumption T-G, we readily have $\boldsymbol{\mathcal{G}}_2 = o_p(1)$. Next, expand $\boldsymbol{\mathcal{G}}_3$ similarly to (B.9),

$$\boldsymbol{\mathcal{G}}_3(i,i) = \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \Big(\Re\left(\boldsymbol{I}_{\bar{a}\bar{a}}^{(1,1)}(\lambda_j, i, i)\right) + \Re\left(\boldsymbol{I}_{\bar{a}\bar{a}}^{(2,2)}(\lambda_j, i, i)\right) + 2\Re\left(\boldsymbol{I}_{\bar{a}\bar{a}}^{(1,2)}(\lambda_j, i, i)\right)$$
$$+ 2\Re\left(\boldsymbol{I}_{a\bar{a}}^{(1)}(\lambda_j, i, i)\right) + 2\Re\left(\boldsymbol{I}_{a\bar{a}}^{(2)}(\lambda_j, i, i)\right)\Big) \equiv \sum_{p=1}^{5} \boldsymbol{\mathcal{G}}_{3p}(i,i).$$

Now, for the first term, $\boldsymbol{\mathcal{G}}_{31}(i,i)$, we may use the same arguments as for (B.10) to show

$$\boldsymbol{\mathcal{G}}_{31}(i,i) \leq O_p\left(\frac{\ln(n)^2}{m_d}\right) + O_p\left(\frac{n^{1/2}\ln(n)^2}{m_G m_d}\frac{m_G^{\epsilon}}{\ell_G^{1+\epsilon}}\right) + O_p\left(\frac{n\ln(n)^2}{m_G m_d \ell^2}\right), \tag{B.18}$$

for some arbitrarily small $\epsilon > 0$. Hence, by defining

$$\bar{f}_G(\ell_G, m_G, n) = 1 \vee \frac{n^{1/2} m_G^{\epsilon}}{m_G \ell_G^{1+\epsilon}} \vee \frac{n}{m_G \ell_G^2}, \quad \text{with} \quad \bar{f}_G(\ell_G, m_G, n) \to 1,$$

as $n \to \infty$ by Assumptions T and T-G, we may write $\boldsymbol{\mathcal{G}}_{31}(i,i) \leq O_p\big(\ln(n)^2 m_d^{-1} \bar{f}_G(\ell_G, m_G, n)\big)$. Next, for the second term, $\boldsymbol{\mathcal{G}}_{32}(i,i)$, and by the same arguments as for (B.11), we have

$$\boldsymbol{\mathcal{G}}_{32}(i,i) \leq O_p\left(\frac{\ln(n)^4}{m_d^2}\right) + O_p\left(\frac{\ln(n)^5}{m_G m_d^2}\frac{m_G^{1/2+\epsilon}}{\ell_G^{1+\epsilon}}\right) + O_p\left(\frac{n^{1/2}\ln(n)^4}{m_G m_d^2}\frac{m_G^{\epsilon}}{\ell_G^{1+\epsilon}}\right) + O_p\left(\frac{n\ln(n)^4}{m_G m_d^2 \ell_G^2}\right), \tag{B.19}$$

using also that $\ln(n)^6 \ll \ln(n)^4 n^{1/2}$ and $\ln(n)/\ell_G^{3/2} \ll m_G^{\epsilon}/\ell_G^{1+\epsilon}$. Since Assumptions F and T readily imply $\ln(n)^p \ll m_d$ for some finite $p \in \mathbb{N}$, then we may combine the bounds in (B.18) and (B.19) with Assumption T-G to show $\boldsymbol{\mathcal{G}}_{32}(i,i) = o_p\big(\boldsymbol{\mathcal{G}}_{31}(i,i)\big)$. Hence, by the results for $\boldsymbol{\mathcal{G}}_1$, (B.18), (B.19) in conjunction with the Cauchy-Schwarz inequality and the continuous mapping theorem, we get the following bounds for the cross-product terms of $\boldsymbol{\mathcal{G}}_3(i,i)$,

$$\boldsymbol{\mathcal{G}}_{33}(i,i)/2 \leq \sqrt{\boldsymbol{\mathcal{G}}_{31}(i,i)}\sqrt{\boldsymbol{\mathcal{G}}_{32}(i,i)} = o_p\big(\boldsymbol{\mathcal{G}}_{31}(i,i)\big)$$
$$\boldsymbol{\mathcal{G}}_{34}(i,i)/2 \leq \sqrt{\boldsymbol{\mathcal{G}}_{1}(i,i)}\sqrt{\boldsymbol{\mathcal{G}}_{31}(i,i)} = O_p\left(\frac{\ln(n)}{m_d^{1/2}}\sqrt{\bar{f}_G(\ell_G, m_G, n)}\right), \quad \boldsymbol{\mathcal{G}}_{35}(i,i)/2 \leq o_p\big(\boldsymbol{\mathcal{G}}_{34}(i,i)\big).$$

Since the bounds for $\boldsymbol{\mathcal{G}}_3(i,i)$ are independent of $i = 1, \ldots, k+1$, we may use the Cauchy-Schwarz inequality and the continuous mapping theorem to show $\boldsymbol{\mathcal{G}}_3 = o_p(1)$, concluding the proof of **(a)**.

For (b). First, write $\hat{\eta}_t = \hat{\varsigma}_t + \hat{\epsilon}_t$, where $\hat{\varsigma}_t = \hat{e}_t - \boldsymbol{\mathcal{B}}'\hat{\boldsymbol{u}}_{t-1}$ and $\hat{\epsilon}_t = \big(\boldsymbol{\mathcal{B}} - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\big)'\hat{\boldsymbol{u}}_{t-1}$, and make the decomposition

$$\widehat{\boldsymbol{G}}_{\hat{\eta}\hat{\eta}}(\ell_G, m_G) = \widehat{\boldsymbol{G}}_{\hat{\varsigma}\hat{\varsigma}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\hat{\epsilon}\hat{\epsilon}}(\ell_G, m_G) + 2\widehat{\boldsymbol{G}}_{\hat{\varsigma}\hat{\epsilon}}(\ell_G, m_G). \tag{B.20}$$

Now, $\widehat{\boldsymbol{G}}_{\hat{\varsigma}\hat{\varsigma}}(\ell_G, m_G) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{\eta\eta}$ readily follows by (10) and **(a)**. Moreover, since $\widehat{\boldsymbol{G}}_{\hat{\epsilon}\hat{\epsilon}}(\ell_G, m_G) = O_p(m^{-1})$ by applying Theorem 1 to $\big(\boldsymbol{\mathcal{B}} - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\big)$ and **(a)**, and $\widehat{\boldsymbol{G}}_{\hat{\varsigma}\hat{\epsilon}}(\ell_G, m_G) \leq O_p(m^{-1/2})$ by the Cauchy-Schwarz inequality and the first two results, this concludes the proof. $\square$

B.3 Proof of Theorem 3

First, for **(a)**, write $\hat{\boldsymbol{F}}^c_{\hat{u}\hat{u}}(\ell, m) - \hat{\boldsymbol{F}}_{\hat{u}\hat{u}}(\ell, m) = \hat{\boldsymbol{F}}_{\hat{c}\hat{c}}(\ell, m) + \hat{\boldsymbol{F}}_{\hat{u}\hat{c}}(\ell, m) + \hat{\boldsymbol{F}}_{\hat{c}\hat{u}}(\ell, m)$. Hence, it suffices to establish bounds for $\hat{\boldsymbol{F}}_{\hat{c}\hat{c}}(\ell, m, i, i)$ and $\hat{\boldsymbol{F}}_{\hat{u}\hat{c}}(\ell, m, i, i)$ since the bounds for the off-diagonal terms of the error matrices will, then, follow by the Cauchy-Schwarz inequality. Hence, by Lemma B.9(b),

$$\hat{\boldsymbol{F}}_{\hat{c}\hat{c}}(\ell, m, i, i) = \frac{2\pi}{n} \sum_{j=\ell}^{m} O_p(\lambda_j^{2d_i}) \leq \frac{2\pi m^{1+2d_i}}{n^{1+2d_i}} \frac{2\pi}{n} \sum_{j=\ell}^{m} O_p\left(\left(\frac{j}{m}\right)^{2d_i} \frac{1}{j^{1+\epsilon}}\right)$$
$$\leq O_p\left(\left(\frac{m}{n}\right)^{1+2d_i} \frac{1}{\ell^{1+\epsilon}}\right), \tag{B.21}$$

for some arbitrarily small $\epsilon > 0$, using Varneskov (2017, Lemma C.4) for the last inequality. Similarly for the cross-product term, the bounds in Lemma B.9(b) allows us to write

$$\hat{\boldsymbol{F}}_{\hat{u}\hat{c}}(\ell, m, i, i) = \frac{2\pi}{n} \sum_{j=\ell}^{m} O_p(\lambda_j^{d_i}) + \frac{2\pi}{n} \sum_{j=\ell}^{m} O_p\left(\frac{\lambda_j^{2d_i} n^{1/2}}{j}\right) + \frac{2\pi}{n} \sum_{j=\ell}^{m} O_p\left(\frac{\lambda_j^{d_i} \ln(n) n^{1/2}}{m_d^{1/2} j}\right)$$
$$\leq O_p\left(\frac{m^{1+d_i}}{n^{1+d_i}} \frac{1}{\ell^{1+\epsilon}}\right) + O_p\left(\frac{m^{2d_i+\epsilon}}{n^{1/2+2d_i}} \frac{1}{\ell^{1+\epsilon}}\right) + O_p\left(\frac{\ln(n) m^{d_i+\epsilon}}{m_d^{1/2} n^{1/2+d_i}} \frac{1}{\ell^{1+\epsilon}}\right), \tag{B.22}$$

using the same arguments as above. However, since we can write the scaled bound (B.22) as

$$\lambda_m^{-1} \hat{\boldsymbol{F}}_{\hat{u}\hat{c}}(\ell, m, i, i) \leq O_p\left(\left(\frac{m}{n}\right)^{d_i} \frac{1}{\ell^{1+\epsilon}} \left(1 + \left(\frac{m}{n}\right)^{d_i} \frac{n^{1/2}}{m} + \frac{\ln(n) n^{1/2}}{m_d^{1/2} m^{1-\epsilon}}\right)\right),$$

the final result follows since $n^{1/2}/m \to 0$ and $\hat{\boldsymbol{F}}_{\hat{c}\hat{c}}(\ell, m, i, i) \leq o_p(\hat{\boldsymbol{F}}_{\hat{u}\hat{c}}(\ell, m, i, i))$.

For **(b)**, we have $\hat{\boldsymbol{F}}^c_{\hat{u}\hat{e}}(\ell, m) - \hat{\boldsymbol{F}}_{\hat{u}\hat{e}}(\ell, m) = \hat{\boldsymbol{F}}_{\hat{c}\hat{e}}(\ell, m)$. Now, by Assumption C, since $\hat{\boldsymbol{c}}_t$ and $\hat{e}_t$ have a non-trivial spectrum in the vicinity of the origin and satisfy the remaining conditions of Assumption D1-D3, we have $\hat{\boldsymbol{F}}_{\hat{c}\hat{e}}(\ell, m) = O_p(\hat{\boldsymbol{F}}_{\hat{u}\hat{c}}(\ell, m, i, i))$. Hence, the result follows by **(a)**.

For **(c)**, write $\widehat{\boldsymbol{G}}^c_{\hat{u}\hat{u}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G, m_G) = \widehat{\boldsymbol{G}}_{\hat{c}\hat{c}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\hat{c}\hat{u}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\hat{u}\hat{c}}(\ell_G, m_G)$, similarly to the decomposition in **(a)**. As in **(a)**, we may invoke Lemma B.9(b) to show

$$\widehat{\boldsymbol{G}}_{\hat{c}\hat{c}}(\ell_G, m_G, i, i) = \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{2d_i}) \leq \frac{K m_G^{2d_i}}{n^{2d_i}} \sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{2d_i} \frac{1}{j^{1+\epsilon}}\right)$$
$$\leq O_p\left(\left(\frac{m_G}{n}\right)^{2d_i} \frac{1}{\ell_G^{1+\epsilon}}\right), \tag{B.23}$$

using $m_G/(m_G - \ell_G + 1) \leq K$ and Varneskov (2017, Lemma C.4). Similarly, Lemma B.9(b) may be used to bound the cross-product term as follows

$$\widehat{\boldsymbol{G}}_{\hat{u}\hat{c}}(\ell_G, m_G, i, i) \leq \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{d_i}) + \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{2d_i} n^{1/2}}{j}\right) + \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{d_i}\ln(n) n^{1/2}}{m_d^{1/2} j}\right)$$
$$\leq O_p\left(\left(\frac{m_G}{n}\right)^{d_i}\frac{1}{\ell_G^{1+\epsilon}}\right) + O_p\left(\left(\frac{m_G}{n}\right)^{2d_i}\frac{n^{1/2}}{m_G^{1-\epsilon}\ell_G^{1+\epsilon}}\right) + O_p\left(\left(\frac{m_G}{n}\right)^{d_i}\frac{n^{1/2}\ln(n)}{m_G m_d^{1/2}}\frac{1}{\ell_G^{1+\epsilon}}\right). \quad \text{(B.24)}$$

Hence, by $n^{1/2}/m_G \to 0$ and the same arguments as in **(a)**, $\widehat{\boldsymbol{G}}_{\hat{u}\hat{c}}(\ell_G, m_G, i, i) \leq O_p((m_G/n)^{d_i}/\ell_G^{1+\epsilon})$ as well as $\widehat{\boldsymbol{G}}_{\hat{c}\hat{c}}(\ell_G, m_G, i, i) \leq o_p(\widehat{\boldsymbol{G}}_{\hat{u}\hat{c}}(\ell_G, m_G, i, i))$, providing the result.

For **(d)**, make the decomposition $\hat{\eta}_t^c = \hat{e}_t - \widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m)'\hat{\boldsymbol{u}}_t^c = \hat{\eta}_t - \hat{\tau}_t^{(1)} - \hat{\tau}_t^{(2)}$, where $\hat{\eta}_t$ is defined as in (17) and we let $\hat{\tau}_t^{(1)} = (\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m))'\hat{\boldsymbol{u}}_t^c$ and $\hat{\tau}_t^{(2)} = \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)'\hat{\boldsymbol{c}}_t$. Hence, write

$$\widehat{\boldsymbol{G}}^c_{\hat{\eta}\hat{\eta}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\hat{\eta}\hat{\eta}}(\ell_G, m_G) = \widehat{\boldsymbol{G}}^{(1,1)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}^{(2,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G)$$
$$- 2\widehat{\boldsymbol{G}}^{(1)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G) - 2\widehat{\boldsymbol{G}}^{(2)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G) + 2\widehat{\boldsymbol{G}}^{(1,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G),$$

where the first two terms are the (trimmed) long-run variance estimates for $\hat{\tau}_t^{(1)}$ and $\hat{\tau}_t^{(2)}$, respectively, and the last three terms are covariances between $\hat{\tau}_t^{(1)}$, $\hat{\tau}_t^{(2)}$ and $\hat{\eta}_t$. First, for $\widehat{\boldsymbol{G}}^{(1,1)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G)$, write

$$\widehat{\boldsymbol{G}}^{(1,1)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) = (\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m))'\Big(\widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\hat{c}\hat{c}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\hat{c}\hat{u}}(\ell_G, m_G)$$
$$+ \widehat{\boldsymbol{G}}_{\hat{u}\hat{c}}(\ell_G, m_G)\Big)(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)). \quad \text{(B.25)}$$

Then, by **(a)** and **(b)**, $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m) \leq O_p((m/n)^{\underline{d}}/\ell^{1+\epsilon})$. Moreover, $\widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G, m_G) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}$ follows by Lemma B.4(a). Hence, we may use (B.23), (B.24) and $n^{1/2}/m_G \to 0$ to show

$$\widehat{\boldsymbol{G}}^{(1,1)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) \leq O_p\left(\left(\frac{m}{n}\right)^{2\underline{d}}\frac{1}{\ell^{2(1+\epsilon)}}\right) \times \left(1 + \left(\frac{m_G}{n}\right)^{\underline{d}}\frac{1}{\ell_G^{1+\epsilon}}\right). \quad \text{(B.26)}$$

Next, for the second term $\widehat{\boldsymbol{G}}^{(2,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G)$, we have $\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) \xrightarrow{\mathbb{P}} \boldsymbol{\mathcal{B}}$ by Theorem 1, which, together with (B.23), delivers the bound $\widehat{\boldsymbol{G}}^{(2,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) \leq O_p((m_G/n)^{2\underline{d}}/\ell_G^{1+\epsilon})$.

For the covariance term $\widehat{\boldsymbol{G}}^{(2)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G)$, use the decomposition from Lemma B.4(b), $\hat{\eta}_t = \hat{\varsigma}_t + \hat{\epsilon}_t$ where, once again, $\hat{\varsigma}_t = \hat{e}_t - \boldsymbol{\mathcal{B}}'\hat{\boldsymbol{u}}_{t-1}$ and $\hat{\epsilon}_t = \big(\boldsymbol{\mathcal{B}} - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\big)'\hat{\boldsymbol{u}}_{t-1}$. Here, we may readily invoke Theorem 1 to show $\boldsymbol{\mathcal{B}} - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m) = O_p(m^{-1/2})$, such that by (B.24) and the same arguments used for **(b)**,

$$\widehat{\boldsymbol{G}}^{(2)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G) \leq O_p\left(\left(\frac{m_G}{n}\right)^{\underline{d}}\frac{1}{\ell_G^{1+\epsilon}}\right) \times \left(1 + m^{-1/2}\right). \quad \text{(B.27)}$$

Next, for $\widehat{\boldsymbol{G}}^{(1)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G)$, we write $\hat{\tau}_t^{(1)} = \hat{\tau}_t^{(1,1)} + \hat{\tau}_t^{(1,2)}$ with $\hat{\tau}_t^{(1,1)} = (\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m))'\hat{\boldsymbol{u}}_t$ and, similarly, $\hat{\tau}_t^{(1,2)} = (\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m))'\hat{\boldsymbol{c}}_t$, and use these with decomposition of $\hat{\eta}_t$ to expand

$$\widehat{\boldsymbol{G}}^{(1)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G) = \widehat{\boldsymbol{G}}^{(1,1)}_{\hat{\varsigma}\hat{\tau}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}^{(1,2)}_{\hat{\varsigma}\hat{\tau}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}^{(1,1)}_{\hat{\epsilon}\hat{\tau}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}^{(1,2)}_{\hat{\epsilon}\hat{\tau}}(\ell_G, m_G), \tag{B.28}$$

explicating the covariance terms between $\hat{\varsigma}_t$, $\hat{\epsilon}_t$, $\hat{\tau}_t^{(1,1)}$ and $\hat{\tau}_t^{(1,2)}$. By the same arguments as above,

$$\begin{aligned}
\widehat{\boldsymbol{G}}^{(1,1)}_{\hat{\varsigma}\hat{\tau}}(\ell_G, m_G) &= O_p\left(\widehat{\boldsymbol{G}}_{\hat{\varsigma}\hat{\varsigma}}(\ell_G, m_G)\right) \times O_p\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right), \\
\widehat{\boldsymbol{G}}^{(1,2)}_{\hat{\varsigma}\hat{\tau}}(\ell_G, m_G) &= O_p\left(\widehat{\boldsymbol{G}}_{\hat{u}\hat{c}}(\ell_G, m_G)\right) \times O_p\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right), \\
\widehat{\boldsymbol{G}}^{(1,1)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G) &= O_p\left(\widehat{\boldsymbol{G}}_{\hat{u}\hat{u}}(\ell_G, m_G)\right) \times O_p\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right) \times O_p\left(\boldsymbol{\mathcal{B}} - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right), \\
\widehat{\boldsymbol{G}}^{(1,2)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G) &= O_p\left(\widehat{\boldsymbol{G}}_{\hat{u}\hat{c}}(\ell_G, m_G)\right) \times O_p\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right) \times O_p\left(\boldsymbol{\mathcal{B}} - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right).
\end{aligned}$$

Hence, by using bounds for $\boldsymbol{\mathcal{B}} - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)$, $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)$, (B.24) and Lemma B.4, we have

$$\widehat{\boldsymbol{G}}^{(1)}_{\hat{\eta}\hat{\tau}}(\ell_G, m_G) \leq O_p\left(\left(\frac{m}{n}\right)^{\underline{d}} \frac{1}{\ell^{1+\epsilon}}\right). \tag{B.29}$$

For the last term, $\widehat{\boldsymbol{G}}^{(1,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G)$, we, again, apply the decomposition $\hat{\tau}_t^{(1)} = \hat{\tau}_t^{(1,1)} + \hat{\tau}_t^{(1,2)}$ to write

$$\widehat{\boldsymbol{G}}^{(1,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) = \widehat{\boldsymbol{G}}^{(1,2,1)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}^{(1,2,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G), \tag{B.30}$$

where, using the same arguments as above,

$$\begin{aligned}
\widehat{\boldsymbol{G}}^{(1,2,1)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) &= O_p\left(\widehat{\boldsymbol{G}}_{\hat{u}\hat{c}}(\ell_G, m_G)\right) \times O_p\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right), \\
\widehat{\boldsymbol{G}}^{(1,2,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) &= O_p\left(\widehat{\boldsymbol{G}}_{\hat{c}\hat{c}}(\ell_G, m_G)\right) \times O_p\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right),
\end{aligned}$$

which, if additionally invoking (B.23), delivers the asymptotic bound

$$\widehat{\boldsymbol{G}}^{(1,2)}_{\hat{\tau}\hat{\tau}}(\ell_G, m_G) \leq O_p\left(\left(\frac{m_G}{n}\right)^{\underline{d}} \frac{1}{\ell_G^{1+\epsilon}}\right) \times O_p\left(\left(\frac{m}{n}\right)^{\underline{d}} \frac{1}{\ell^{1+\epsilon}}\right). \tag{B.31}$$

Hence, by collecting results, we have

$$\widehat{\boldsymbol{G}}^{c}_{\hat{\eta}\hat{\eta}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\hat{\eta}\hat{\eta}}(\ell_G, m_G) \leq O_p\left(\left(\frac{m_G}{n}\right)^{\underline{d}} \frac{1}{\ell_G^{1+\epsilon}}\right) + O_p\left(\left(\frac{m}{n}\right)^{\underline{d}} \frac{1}{\ell^{1+\epsilon}}\right), \tag{B.32}$$

providing the final result, thereby concluding the proof. □

B.4 Technical Lemmas and Definitions

Definition 1 (Fractional Filters)**.** *Let $\theta \in \mathbb{R}$ be a fixed scalar, then the following definitions are used to derive the (higher-order) periodogram bounds below:*

(a) $D_n(e^{\mathrm{i}\lambda};\theta) = \sum_{k=0}^{n} \frac{(-\theta)_k}{k!} e^{\mathrm{i}k\lambda}$ *where* $(\theta)_k = \frac{\Gamma(\theta+k)}{\Gamma(\theta)} = (\theta)(\theta+1)\cdots(\theta+k-1)$,

(b) $D_n(e^{-\mathrm{i}\lambda}L;\theta) = \sum_{p=0}^{n-1} \tilde{\theta}_{\lambda p} e^{-\mathrm{i}p\lambda} L^p$, *where* $\tilde{\theta}_{\lambda p} = \sum_{k=p+1}^{n} \frac{(-\theta)_k}{k!} e^{\mathrm{i}k\lambda}$,

(c) $J_n(e^{\mathrm{i}\lambda}) = \sum_{k=1}^{n} e^{\mathrm{i}k\lambda}/k$,

(d) $\tilde{J}_{n\lambda_j}(e^{-\mathrm{i}\lambda_j}L) = \sum_{p=0}^{n-1} \tilde{j}_{\lambda_j p} e^{-\mathrm{i}p\lambda_j} L^p$ *where* $\tilde{j}_{\lambda_j p} = \sum_{k=p+1}^{n} e^{\mathrm{i}k\lambda_j}/k$.

Lemma B.5 (Shimotsu & Phillips (2005, Lemma 5.9))**.** *Let the conditions for Theorem 1 hold. Moreover, let u_t be the ith element of $\boldsymbol{q}_t$, define $\zeta_t = (1-L)^{\theta} u_t \mathbf{1}_{\{t\geq 1\}}$ and let $\theta \in [-1+\epsilon, C]$ where $\epsilon > 0$ is arbitrarily small as well as $C \in (1,\infty)$, then*

(a) $-w_{\ln(1-L)\zeta}(\lambda_j) = J_n(e^{\mathrm{i}\lambda_j}) D_n(e^{\mathrm{i}\lambda_j};\theta) w_u(\lambda_j) + n^{-1/2} V_n(\lambda_j;\theta)$,

(b) $-w_{\ln(1-L)u}(\lambda_j) = J_n(e^{\mathrm{i}\lambda_j}) w_u(\lambda_j) - (2\pi n)^{-1/2} \tilde{J}_{n\lambda_j}(e^{-\mathrm{i}\lambda_j}L) \boldsymbol{A}(0,i)\boldsymbol{\epsilon}_n + r_n(\lambda_j)$,

(c) $w_{\ln(1-L)^2\zeta}(\lambda_j) = J_n(e^{\mathrm{i}\lambda_j})^2 D_n(e^{\mathrm{i}\lambda_j};\theta) w_u(\lambda_j) + n^{-1/2} \Psi_n(\lambda_j;\theta)$,

where, uniformly in $j = 1,\dots,m$, with $\frac{m}{n} + \frac{1}{n} = o(1)$,

$$\mathbb{E}[\sup_{\theta} |n^{\theta-1/2} j^{1/2-\theta} V_n(\lambda_j;\theta)|^2] = O(\ln(n)^4), \qquad \mathbb{E}[|j^{1/2} r_n(\lambda_j)|^2] = o(1) + O(j^{-1}),$$
$$\mathbb{E}[\sup_{\theta} |n^{\theta-1/2} j^{1/2-\theta} \Psi_n(\lambda_j;\theta)|^2] = O(\ln(n)^6).$$

Lemma B.6 (Shimotsu (2010, Lemma B.2))**.** *Let $\kappa_t = \mathbf{1}_{\{t\geq 1\}}$ and $C \in (1,\infty)$. The following results, then, holds uniformly in $j = 1,\dots,m$ with $\frac{m}{n} + \frac{1}{n} = o(1)$ and in θ:*

(a) $w_{\Delta^{\theta}\kappa}(\lambda_j) = O(n^{1/2-\theta} j^{\theta-1})$ *when* $\theta \geq 0$,

(b) $-w_{\ln(1-L)\Delta^{\theta}\kappa} = J_n(e^{\mathrm{i}\lambda_j}) w_{\Delta^{\theta}\kappa}(\lambda_j) + O\big(n^{1/2-\theta} \ln(n) \big(j^{-1} \mathbf{1}_{\{\theta\in[-C,1]\}} + \mathbf{1}_{\{\theta\in[1,2]\}}\big)\big)$,

(c) $w_{\ln(1-L)^2\Delta^{\theta}\kappa} = J_n(e^{\mathrm{i}\lambda_j})^2 w_{\Delta^{\theta}\kappa}(\lambda_j) + O\big(n^{1/2-\theta} \ln(n)^2 \big(j^{-1} \mathbf{1}_{\{\theta\in[-C,1]\}} + \mathbf{1}_{\{\theta\in[1,2]\}}\big)\big)$.

Lemma B.7 (Higher-order Periodogram Bounds)**.** *Let the conditions for Theorem 1 hold. Moreover, define $\theta_n > 0$ such that $\theta_n \asymp n^{-\rho}$, $\rho \in (0,1)$ and $\max_{i=1,\dots,k+1} |\hat{\theta}_i| < \theta_n$ with probability approaching 1. Then, the periodograms of $a_{t-1}^{(1)}(i)$ and $a_{t-2}^{(2)}(i,\bar{\theta}_i)$, the two terms in the Taylor expansion* (B.8)*, satisfy the following stochastic bounds for $i = 1,\dots,k+1$ and $j = 1,\dots,m$, with $\frac{m}{n} + \frac{1}{n} = o(1)$,*

(a) $\boldsymbol{I}_{aa}^{(1,1)}(\lambda_j,i,i) = O_p\left(\ln(n)^2 \big(1 + n^{1/2}/j + n/j^2\big)\right)$,

(b) $\boldsymbol{I}_{aa}^{(2,2)}(\lambda_j,i,i) \leq O_p\left(\ln(n)^4 \left(1 + \ln(n) j^{-1/2} + n^{1/2} j^{-1}\right)^2\right)$.

Proof. Before proceeding to the proofs, note that for $j = 1, \ldots, m$ with $\frac{m}{n} + \frac{1}{n} = o(1)$, we have

$$J_n(e^{\mathrm{i}\lambda_j}) = O(\ln(n)) \quad \text{and} \quad \lambda_j^{-\theta} D_n(e^{\mathrm{i}\lambda}; \theta) = e^{-(\pi/2)\theta \mathrm{i}} + O(\lambda_j) + O(j^{-1-\theta}), \tag{B.33}$$

uniformly for $\theta \in (-1, C]$ and $C \in (1, \infty)$, by Shimotsu & Phillips (2005, Lemmas 5.2 and 5.8). These bounds will be used throughout, sometimes without explicit reference.

For (a). First, use the decomposition of $a_t(i)$ to write

$$a_{t-1}^{(1)}(i) = \ln(1-L)v_t(i)\mathbf{1}_{\{t \geq 1\}} + \ln(1-L)\tilde{\mu}_t(i)\mathbf{1}_{\{t \geq 1\}} \equiv \tilde{v}_{t-1}^{(1)}(i) + \tilde{\mu}_{t-1}^{(1)}(i). \tag{B.34}$$

and, as a result, the discrete Fourier transform as $\boldsymbol{w}_a^{(1)}(\lambda_j, i) = \boldsymbol{w}_{\tilde{v}}^{(1)}(\lambda_j, i) + \boldsymbol{w}_{\tilde{\mu}}^{(1)}(\lambda_j, i)$. Now, by combining (B.33), Assumptions D1-D3, and Lemma B.5(b), we have

$$\boldsymbol{w}_{\tilde{v}}^{(1)}(\lambda_j, i) = O_p(\ln(n)) + o_p(j^{-1/2}) + O_p(j^{-1}) + (2\pi n)^{-1/2} \tilde{J}_{n\lambda_j}(e^{-\mathrm{i}\lambda_j}L)\boldsymbol{A}(0, i)\boldsymbol{\epsilon}_n,$$

and, moreover, since $\mathbb{E}[|\tilde{J}_{n\lambda_j}(e^{-\mathrm{i}\lambda_j}L)\boldsymbol{A}(0, i)\boldsymbol{\epsilon}_n|^2] = O(n/j)$ follows by applying the results in Shimotsu & Phillips (2005, (77) and (89)), this readily shows that $\boldsymbol{w}_{\tilde{v}}^{(1)}(\lambda_j, i) = O_p(\ln(n))$. Hence, by using this bound in conjunction with Lemmas B.6(a)-(b) for $\boldsymbol{w}_{\tilde{\mu}}^{(1)}(\lambda_j, i)$, it follows that

$$\boldsymbol{w}_a^{(1)}(\lambda_j, i) = O_p\left(\ln(n)\big(1 + n^{1/2}/j\big)\right), \tag{B.35}$$

which, together with the Cauchy-Schwarz inequality, establishes **(a)** for elements $i = 2, \ldots, k+1$, that is, when $v_t(i) = u_t(i-1)$. Moreover, by using the relation (10) with Assumption D2 and the same arguments, we get an identical bound (B.35) for $i = 1$, that is, with $v_t(i) = e_t$. As this will also hold true for the remaining arguments given below, we only establish results for $i = 2, \ldots, k+1$.

For (b). First, decompose $a_{t-2}^{(2)}(i, \bar{\theta}_i)$ similarly to (B.34),

$$\begin{aligned} a_{t-2}^{(2)}(i, \bar{\theta}_i) &= \ln(1-L)^2(1-L)^{\bar{\theta}_i} v_t(i)\mathbf{1}_{\{t \geq 1\}} + \ln(1-L)^2(1-L)^{\bar{\theta}_i}\tilde{\mu}_t(i)\mathbf{1}_{\{t \geq 1\}} \\ &\equiv \tilde{v}_{t-2}^{(2)}(i) + \tilde{\mu}_{t-2}^{(2)}(i), \quad \text{and} \quad \boldsymbol{w}_a^{(2)}(\lambda_j, i) = \boldsymbol{w}_{\tilde{v}}^{(2)}(\lambda_j, i) + \boldsymbol{w}_{\tilde{\mu}}^{(2)}(\lambda_j, i). \end{aligned} \tag{B.36}$$

Now, using $|\bar{\theta}_i| \leq |\hat{\theta}_i| < \theta_n$ for all $i = 1, \ldots, k+1$, Lemma B.5(c) and (B.33), it follows that

$$\boldsymbol{w}_{\tilde{v}}^{(2)}(\lambda_j, i) \leq O_p\left(\ln(n)^2 \lambda_j^{-\theta_n}\right) + O_p\left(\ln(n)^3 \lambda_j^{-\theta_n} j^{-1/2}\right). \tag{B.37}$$

Next, to get a simpler asymptotic approximation, we make a second-order Taylor expansion of the power function of Fourier frequencies $\lambda_j^{-\theta_n}$ around $\theta_n = 0$,

$$\lambda_j^{-\theta_n} = 1 + \theta_n \ln(1/\lambda_j) + \frac{\theta_n^2}{2}\ln(1/\lambda_j)^2 \lambda_j^{-\bar{\theta}_n}, \tag{B.38}$$

for some $\bar{\theta}_n \in [0, \theta_n]$, using the mean-value theorem. Hence, since $\theta_n \asymp n^{-\rho}$, $\rho \in (0, 1)$, this readily shows that $\lambda_j^{-\theta_n} = 1+o(1)$ and, as a result, $\boldsymbol{w}_{\tilde{v}}^{(2)}(\lambda_j, i) \leq O_p\left(\ln(n)^2(1+\ln(n)j^{-1/2})\right)$. Moreover, using Lemma B.6(c) and the fact that $\theta_n \to 0$ as $n \to \infty$, this delivers the bounds

$$\boldsymbol{w}_{\tilde{\mu}}^{(2)}(\lambda_j, i) \leq K\ln(n)^2\left(\frac{n^{1/2}}{j}\right)\left((n/j)^{\theta_n} + (n)^{\theta_n}\right) \leq K\ln(n)^2\left(\frac{n^{1/2}}{j}\right)n^{\theta_n}. \tag{B.39}$$

Now, by combining (B.38) and (B.39) with the bound for $\boldsymbol{w}_{\tilde{u}}^{(2)}(\lambda_j, i)$, this shows that

$$\boldsymbol{w}_{a}^{(2)}(\lambda_j, i) \leq O_p\left(\ln(n)^2\left(1+\ln(n)j^{-1/2} + n^{1/2}j^{-1}\right)\right), \tag{B.40}$$

which, together with the Cauchy-Schwarz inequality, establishes **(b)**. □

The following lemma collects various bounds from Shimotsu & Phillips (2005) to state a bound for the discrete Fourier transform of $\zeta_t = (1-L)^{\theta}c_t\mathbf{1}_{\{t\geq 1\}}$, where c_t is an element of $\boldsymbol{c}_t$, that satisfies Assumption C, and $\theta \in [-1+\epsilon, C]$ with $\epsilon > 0$ being arbitrarily small and $C \in (1,\infty)$.

Lemma B.8 (Discrete Fourier Transform Bound for ζ_t)**.** *Suppose the conditions of Assumption C holds and $\theta \in [-1+\epsilon, C]$. Then, for $j = 1,\ldots,m$ with $\frac{m}{n} + \frac{1}{n} = o(1)$,*

$$\boldsymbol{w}_\zeta(\lambda_j) = O_p(\lambda_j^{\theta}) + O_p(\lambda_j^{\theta}\ln(n)j^{-1/2}) + O_p(\lambda_j^{-1}n^{-\theta-1})\mathbf{1}_{\{\theta\in[-1+\epsilon,-1/2]\}} + O_p(n^{-\theta-1})_{\{\theta\in[1/2,C]\}}.$$

Proof. The proof proceeds by providing separate arguments and asymptotic bounds for different ranges the parameter space $\Theta = \{\theta \in [-1+\epsilon, C]\}$. First, for $\Theta_1 = \{\theta \in [-1/2, 1/2]\}$, use Shimotsu & Phillips (2005, Lemma 5.1) make the decomposition,

$$\boldsymbol{w}_\zeta(\lambda_j) = D_n(e^{\mathrm{i}\lambda_j};\theta)\boldsymbol{w}_c(\lambda_j) - (2\pi n)^{-1/2}\tilde{C}_n(\lambda_j,\theta), \quad \tilde{C}_n(\lambda,\theta) = \tilde{D}_{n\lambda}(e^{-\mathrm{i}\lambda}L;\theta)c_n, \tag{B.41}$$

where $D_n(e^{\mathrm{i}\lambda_j};\theta)$ and $\tilde{D}_{n\lambda}(e^{-\mathrm{i}\lambda}L;\theta)$ are given in Defintion 1. Hence, by Assumption C in conjunction with an application of (B.33) and Shimotsu & Phillips (2005, Lemma 5.3) to the terms (B.41), we have, uniformly in $j = 1,\ldots,m$ with $\frac{m}{n} + \frac{1}{n} = o(1)$, that $\boldsymbol{w}_\zeta(\lambda_j) = O_p(\lambda_j^{\theta}) + O_p(\lambda_j^{\theta}\ln(n)j^{-1/2})$. Next, for $\Theta_2 = \{\theta \in [-1+\epsilon, -1/2]\}$, use the decomposition in Shimotsu & Phillips (2005, (45)),

$$\lambda_j^{-\theta}\boldsymbol{w}_\zeta(\lambda_j) = \bar{D}_n(\lambda_j;\theta)\boldsymbol{w}_c(\lambda_j) - \bar{C}_n(\lambda_j,\theta) + \lambda_j^{-\theta}(2\pi n)^{-1/2}e^{\mathrm{i}\lambda_j}(1-e^{\mathrm{i}\lambda_j})^{-1}\zeta_n, \tag{B.42}$$

noting that ζ_n is $I(-\theta)$, and where $\bar{D}_n(\lambda_j;\theta)$ and $\bar{C}_n(\lambda_j,\theta)$ are defined as

$$\begin{aligned}\bar{D}_n(\lambda_j;\theta) &= \lambda_j^{-\theta}(1-e^{\mathrm{i}\lambda_j})^{-1}D_n(e^{\mathrm{i}\lambda_j};\theta+1),\\ \bar{C}_n(\lambda_j,\theta) &= \lambda_j^{-\theta}(1-e^{\mathrm{i}\lambda_j})^{-1}(2\pi n)^{-1/2}\tilde{C}_n(\lambda_j,\theta+1).\end{aligned}$$

Now, by Shimotsu & Phillips (2005, (31) and (39)), $\bar{D}_n(\lambda_j;\theta) = e^{-(\pi/2)\theta\mathrm{i}} + O(\lambda_j) + O(j^{-1/2})$ as well as $\bar{C}_n = O_p(\ln(n)j^{-1/2})$, respectively, uniformly in θ. For the last term in (B.42), since it follows that the terms $e^{\mathrm{i}\lambda_j} = O(1)$ and $(1 - e^{\mathrm{i}\lambda_j})^{-1} = O(nj^{-1})$ as well as $\zeta_n = O_p(n^{-\theta-1/2})$ by, e.g., Phillips & Shimotsu (2004, Lemma A.5), we have $\boldsymbol{w}_\zeta(\lambda_j) = O_p(\lambda_j^\theta) + O_p(\lambda_j^\theta \ln(n) j^{-1/2}) + O_p(\lambda_j^{-1} n^{-\theta-1})$. For the subspace $\Theta_3 = \{\theta \in [1/2, 3/2]\}$, the decomposition in Shimotsu & Phillips (2005, (30)) and the same arguments give the bound $\boldsymbol{w}_\zeta(\lambda_j) = O_p(\lambda_j^\theta) + O_p(\lambda_j^\theta \ln(n) j^{-1/2}) + O_p(n^{-\theta-1})$. As these arguments and steps may be repeated for subsequent subspaces $\Theta_{3+i} = \{\theta \in [1/2 + i, 3/2 + i]\}$, $i = 1, \ldots, K$ for some finite integer K, this gives the result for some finite $C \in (1, \infty)$, concluding the proof. □

Lemma B.9 (Approximate DFT bound for filtered series)**.** *Suppose that the regularity conditions of Theorem 3 hold. Then, the following bounds hold uniformly in* $0 < d_i < 2$, $i = 1, \ldots, k+1$.

(a) for $j = 1, \ldots, m$ *with* $\frac{m}{n} + \frac{1}{n} = o(1)$, *it follows*

$$\boldsymbol{w}_{\hat{c}}(\lambda_j, i) = O_p\left(\lambda_j^{d_i}\left(1 + \frac{\ln(n)}{j^{1/2}} + \frac{\ln(n)}{m_d^{1/2}} + \frac{\ln(n)^2}{j^{1/2} m_d^{1/2}} + \frac{\ln(n)^2}{m_d} + \frac{\ln(n)^3}{j^{1/2} m_d}\right)\right),$$
$$\boldsymbol{w}_{\hat{v}}(\lambda_j, i) = O_p(1) + O_p\left(\frac{n^{1/2-d_i}}{j^{1-d_i}}\right) + O_p\left(\frac{\ln(n)}{m_d^{1/2}}\left(1 + \frac{n^{1/2}}{j} + \frac{\ln(n)}{m_d^{1/2}}\left(1 + \frac{\ln(n)}{j^{1/2}} + \frac{n^{1/2}}{j}\right)\right)\right).$$

(b) for $j = \ell, \ldots, m$ *with* $\ell \asymp n^\nu$, $m \asymp n^\kappa$, $m_d \asymp n^\varrho$, $0 < \nu < \kappa \leq \varrho \leq 1$ *and* $1/n \to 0$,

$$\boldsymbol{w}_{\hat{c}}(\lambda_j, i) = O_p(\lambda_j^{d_i}), \quad \boldsymbol{w}_{\hat{v}}(\lambda_j, i) = O_p(1) + O_p\left(\frac{n^{1/2-d_i}}{j^{1-d_i}}\right) + O_p\left(\frac{\ln(n) n^{1/2}}{m_d^{1/2} j}\right).$$

Proof. First, for **(a)**, write similarly to the proof of Lemma B.1, $\hat{c}_t(i) = (1-L)^{\hat{\theta}_i} \zeta_t(i)$ where we define the series $\zeta_t(i) = (1-L)^{d_i} c_t$ and let $\hat{\theta}_i = \hat{d}_i - d_i = O_p(m_d^{-1/2})$ by Assumption F. Next, by applying the Taylor expansion of $(1-L)^{\hat{\theta}_i}$ in (B.7), we may readily decompose $\boldsymbol{w}_{\hat{c}}(\lambda_j, i)$ as

$$\boldsymbol{w}_{\hat{c}}(\lambda_j, i) = \boldsymbol{w}_\zeta(\lambda_j, i) + \hat{\theta}_i \boldsymbol{w}_\zeta^{(1)}(\lambda_j, i) + \frac{\hat{\theta}_i^2}{2} \boldsymbol{w}_\zeta^{(2)}(\lambda_j, i), \tag{B.43}$$

with $\boldsymbol{w}_\zeta^{(1)}(\lambda_j, i)$ and $\boldsymbol{w}_\zeta^{(2)}(\lambda_j, i)$ being the discrete Fourier transforms of $\zeta_{t-1}^{(1)}(i) \equiv \ln(1-L)\zeta_t(i)$ and $\zeta_{t-2}^{(2)}(i, \bar{\theta}_i) \equiv \ln(1-L)^2 (1-L)^{\bar{\theta}_i} \zeta_t(i)$, respectively, for some $\bar{\theta}_i \in [\hat{\theta}_i, 0]$. Now, the bound for $\boldsymbol{w}_{\hat{c}}(\lambda_j, i)$ follows by applying Lemmas B.5(a), B.5(c) and B.8 in conjunction with (B.33) and the same arguments used to establish Lemma B.7 ((B.38)-(B.39)). Next, the bound for $\boldsymbol{w}_{\hat{v}}(\lambda_j, i)$ follows by the decomposition in (B.8), Lemma B.6(a) as well as the bounds (B.35) and (B.40). **(b)** follows by **(a)**. □

Lemma B.10 (Lead-lag Co-periodogram Bound)**.** *Suppose the conditions of Theorem 1 hold. Moreover, denote by* $\boldsymbol{I}_{u\eta}^{(t,1)}(\lambda_j)$ *the co-periodogram of* $\boldsymbol{u}_{t-1}$ *and* η_t, *and by* $\boldsymbol{I}_{u\eta}^{(t,0)}(\lambda_j)$ *the corresponding periodogram of* $\boldsymbol{u}_t$ *and* η_t, *then*

$$\boldsymbol{I}_{u\eta}^{(t,1)}(\lambda_j) - \boldsymbol{I}_{u\eta}^{(t,0)}(\lambda_j) = O_p(n^{-1/2}).$$

Proof. First, expand $\boldsymbol{I}_{u\eta}^{(t,1)}(\lambda_j)$ and write by addition and subtraction,

$$\begin{aligned}\frac{1}{2\pi n}\sum_{t=2}^{n}\sum_{s=2}^{n}\boldsymbol{u}_{t-1}\eta_s e^{\mathrm{i}\lambda_j(t-1-s)} &= \frac{1}{2\pi n}\sum_{t=2}^{n-1}\sum_{s=2}^{n-1}\boldsymbol{u}_t\eta_s e^{\mathrm{i}\lambda_j(t-s)} + \frac{1}{2\pi n}\sum_{s=2}^{n}\boldsymbol{u}_1\eta_s e^{\mathrm{i}\lambda_j(1-s)} \\ &\quad + \frac{1}{2\pi n}\sum_{t=1}^{n-1}\boldsymbol{u}_t\eta_n e^{\mathrm{i}\lambda_j(t-n)} \\ &= \boldsymbol{I}_{u\eta}^{(t,0)}(\lambda_j) - \frac{1}{2\pi n}\sum_{s=1}^{n}\boldsymbol{u}_n\eta_s e^{\mathrm{i}\lambda_j(n-s)} - \frac{1}{2\pi n}\sum_{t=1}^{n-1}\boldsymbol{u}_t\eta_1 e^{\mathrm{i}\lambda_j(t-1)}.\end{aligned}$$

Hence, since we have for $j = 1, \ldots, m$ and $\frac{m}{n} + \frac{1}{n} = o(1)$, that

$$\frac{1}{2\pi n}\sum_{s=1}^{n}\boldsymbol{u}_n\eta_s e^{\mathrm{i}\lambda_j(n-s)} = \left(\frac{\boldsymbol{u}_n e^{\mathrm{i}\lambda_j n}}{\sqrt{2\pi n}}\right)\bar{\boldsymbol{w}}_\eta(\lambda_j) = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{B.44}$$

$$\frac{1}{2\pi n}\sum_{t=1}^{n-1}\boldsymbol{u}_t\eta_1 e^{\mathrm{i}\lambda_j(t-1)} = \left(\frac{\eta_1 e^{-\mathrm{i}\lambda_j}}{\sqrt{2\pi n}}\right)\boldsymbol{w}_u(\lambda_j) = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{B.45}$$

with the last equalities following by Assumptions D1-D3, Euler's formula and boundedness of the sine an cosine functions, this concludes the proof. □

C Data Construction

We first account for the construction of the high-frequency (HF) return based realized volatility measure. Here, in particular, we decompose the daily return as $r_{t,i} = r_{t,i,\tau} + r_{t,i-\tau,1-\tau}$, where $r_{t,i,\tau}$ measures the open-to-close return for the ith trading day in month t of length τ, that is, the return over the interval $[i-\tau, i]$. Similarly, we denote by $r_{t,i-\tau,1-\tau}$ the overnight return from the preceding trading day, that is, over $[i-1, i-\tau]$. We then utilize HF data over the trading day to estimate the quadratic variation of $r_{t,i,\tau}$, denoted by $[r,r]_{t,i,\tau}$, with arbitrarily high precision as the number of intra-daily observations increases. Let $\bar{V}_{t,i,\tau}$ denote a generic estimator of $[r,r]_{t,i,\tau}$, then our HF estimator of the monthly realized return variance is defined as

$$\bar{V}_t = \sum_{i=1}^{n_t}\left(\bar{V}_{t,i,\tau} + r_{t,i-\tau,1-\tau}^2\right), \tag{C.1}$$

that is, with a correction to account for the variation outside the trading window. We use 30-second observations of S&P 500 futures contracts, traded on the Chicago Mercantile Exchange (CME), for the computation of (C.1) for the period up and including 2010, and we use one-minute observations obtained from TickData on the same ticker after 2010. Since the recorded high-frequency prices are prone to be contaminated by, e.g., bid-ask bounce and asymmetric information effects (or other market microstructure effects), it is pertinent to use an estimator that accounts for an array of market frictions. For this purpose, we use the flat-top realized kernel of Varneskov (2016, 2017), which is robust to general forms of market microstructure noise and has been shown to estimate the quadratic variation with optimal asymptotic and good finite sample properties.

The monthly S&P 500, DS and TB data are obtained from the website of the Federal Reserve Bank of St. Louis, while the PE data are from Professor Robert Shiller's website, see Shiller (2000) for details on its construction.

D IVX Estimation and Testing

The IVX methodology is motivated by the system:

$$y_t = \mu + \boldsymbol{A}'\boldsymbol{x}_{t-1} + e_t, \quad \boldsymbol{x}_t = \boldsymbol{R}_n\boldsymbol{x}_{t-1} + \boldsymbol{u}_t, \quad \boldsymbol{R}_n = \boldsymbol{I}_k + \frac{\boldsymbol{C}}{n^{\alpha}}, \tag{D.1}$$

for some $\alpha \geq 0$, where $\boldsymbol{A}$ is a $k \times 1$ coefficient vector and $\boldsymbol{C} = \text{diag}[c_1, \ldots, c_k]$, with $c_i \leq 0$ for all $i = 1, \ldots, k$. Whereas the innovations of the system $\boldsymbol{v}_t = (e_t, \boldsymbol{u}_t')'$ are assumed to obey conditions similar to those in Assumptions D1-D3, see Kostakis et al. (2015, Assumption INNOV) and Phillips & Lee (2016, Assumption 2.1), the persistence of (D.1) is, in this setting, governed by the autoregressive matrix $\boldsymbol{R}_n$, in analogy with the fractional filter $\boldsymbol{D}(L)$ determining the persistence of (3). In particular, the regressors in (D.1) are said to be integrated if $c_i = 0$ or $\alpha > 1$; local-to-unity if $c_i < 0$ and $\alpha = 1$; mildly stationary if $c_i < 0$ and $\alpha \in (0, 1)$; stationary if $c_i < 0$ and $\alpha = 0$; locally explosive if $c_i > 0$ and $\alpha = 1$; and mildly explosive of $c_i > 0$ and $\alpha \in (0, 1)$. In this setting, and as outlined in the introduction, standard inference for $\boldsymbol{A}$ will generally be distorted and/or suffer from first-order predictive biases if the regressors are no longer stationary. The IVX methodology represents one potential remedy for such problems and is valid across autoregressive persistence regimes.

The key step when developing feasible IVX inference on $\boldsymbol{A}$ is the construction of self-generated instruments using an artificial persistence matrix,

$$\tilde{\boldsymbol{z}}_t = \boldsymbol{R}_{nz}\tilde{\boldsymbol{z}}_t + \Delta\boldsymbol{x}_t, \quad \boldsymbol{R}_{nz} = \boldsymbol{I}_k + \frac{\boldsymbol{C}_z}{n^{\beta}}, \quad \beta \in (0, 1), \quad \boldsymbol{C}_z < \boldsymbol{0}, \tag{D.2}$$

serving to reduce the asymptotic order of the regressors to the mildly integrated case, leaving a residual term that turns out to be asymptotically negligible. In conjunction with the following steps, this reduction of persistence allows for nuisance parameter free inference. Step 1, the variables y_t and $\boldsymbol{x}_t$ are demeaned and stacked into an $n \times 1$ vector $\boldsymbol{Y}$ and an $n \times k$ matrix $\boldsymbol{X}$, respectively. Similarly, write $\tilde{\boldsymbol{Z}} = (\tilde{\boldsymbol{z}}_0', \ldots, \tilde{\boldsymbol{z}}_{n-1}')'$. Step 2, obtain OLS residuals from the regressions in (D.1), denoted by $\hat{e}_t$ and $\hat{\boldsymbol{u}}_t$, and estimate their covariances,

$$\widehat{\Sigma}_{ee} = \frac{1}{n}\sum_{t=1}^{n}\hat{e}_t^2, \quad \widehat{\boldsymbol{\Sigma}}_{eu} = \frac{1}{n}\sum_{t=1}^{n}\hat{e}_t\hat{\boldsymbol{u}}_t', \quad \widehat{\boldsymbol{\Sigma}}_{uu} = \frac{1}{n}\sum_{t=1}^{n}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t'. \tag{D.3}$$

To accommodate potential autocorrelation in the OLS residuals, these estimates are complemented with kernel estimates of the lead-lag structure using a bandwidth M_n and Bartlett weights,

$$\hat{\Lambda}_{ee} = \frac{1}{n}\sum_{h=1}^{M_n}\left(1 - \frac{h}{M_n+1}\right)\sum_{t=h+1}^{n}\hat{e}_t\hat{e}_{t-h}, \quad \hat{\Omega}_{ee} = \widehat{\Sigma}_{ee} + 2\hat{\Lambda}_{ee}, \tag{D.4}$$

$$\hat{\boldsymbol{\Lambda}}_{ue} = \frac{1}{n}\sum_{h=1}^{M_n}\left(1 - \frac{h}{M_n+1}\right)\sum_{t=h+1}^{n}\hat{\boldsymbol{u}}_t\hat{e}_{t-h}, \quad \hat{\boldsymbol{\Omega}}_{eu} = \widehat{\boldsymbol{\Sigma}}_{eu} + \hat{\boldsymbol{\Lambda}}'_{ue}, \tag{D.5}$$

$$\hat{\boldsymbol{\Lambda}}_{uu} = \frac{1}{n}\sum_{h=1}^{M_n}\left(1 - \frac{h}{M_n+1}\right)\sum_{t=h+1}^{n}\hat{\boldsymbol{u}}_t\hat{u}_{t-h}, \quad \hat{\boldsymbol{\Omega}}_{uu} = \widehat{\boldsymbol{\Sigma}}_{uu} + \hat{\boldsymbol{\Lambda}}_{uu} + \hat{\boldsymbol{\Lambda}}'_{uu}. \tag{D.6}$$

These estimates may, then, be used to construct a bias-corrected IVX estimator,

$$\tilde{\boldsymbol{A}} = \left(\tilde{\boldsymbol{Z}}'\boldsymbol{X}\right)^{-1}\left(\tilde{\boldsymbol{Z}}'\boldsymbol{X} - n\hat{\boldsymbol{\Omega}}'_{eu}\right), \tag{D.7}$$

following Phillips & Lee (2016, (2.11)). Whereas the latter, together with Magdalinos & Phillips (2009) and Kostakis et al. (2015), show that the limit theory for $\tilde{\boldsymbol{A}}$ is pivotal, it differs across persistence regimes. However, its associated Wald statistic is always χ^2, similarly to Wald testing for the LCM procedure in Theorem 2. Hence, to carry out consistent Wald testing, we let the residuals of the IVX regression be denoted by $\tilde{e}_t$ and construct HAC covariance estimates,

$$\widetilde{\boldsymbol{\Sigma}}_{ez} = \frac{1}{n}\sum_{t=1}^{n}(\tilde{\boldsymbol{z}}_t\tilde{e}_t)(\tilde{\boldsymbol{z}}_t\tilde{e}_t)', \quad \tilde{\boldsymbol{\Lambda}}_{ze} = \frac{1}{n}\sum_{h=1}^{M_n}\left(1 - \frac{h}{M_n+1}\right)\sum_{t=h+1}^{n}(\tilde{\boldsymbol{z}}_t\tilde{e}_t)(\tilde{\boldsymbol{z}}_{t-h}\tilde{e}_{t-h})', \tag{D.8}$$

and $\tilde{\boldsymbol{\Omega}}_{zu} = \widetilde{\boldsymbol{\Sigma}}_{ez} + \hat{\boldsymbol{\Lambda}}_{ze} + \hat{\boldsymbol{\Lambda}}'_{ze}$. Using these, the IVX-based Wald statistic is constructed via the following asymptotic covariance estimator,

$$\widetilde{\mathbf{AVAR}} = (\tilde{\boldsymbol{Z}}'\boldsymbol{X})^{-1}\mathbb{M}(\tilde{\boldsymbol{Z}}'\boldsymbol{X}), \quad \mathbb{M} = n\tilde{\boldsymbol{\Omega}}_{zu} - n\bar{\boldsymbol{z}}\bar{\boldsymbol{z}}'\hat{\Omega}_{FM}, \tag{D.9}$$

with $\bar{\boldsymbol{z}}$ being the average of $\tilde{\boldsymbol{z}}_t$ and $\hat{\Omega}_{FM} = \hat{\Omega}_{ee} - \hat{\boldsymbol{\Omega}}_{eu}\hat{\boldsymbol{\Omega}}_{uu}^{-1}\hat{\boldsymbol{\Omega}}'_{eu}$ is a modified long-run covariance estimator, in the spirit of Phillips & Hansen (1990). Note that the asymptotic covariance estimator differs slightly from the corresponding in, e.g., Kostakis et al. (2015, (19)) Specifically, we use $\tilde{\boldsymbol{\Omega}}_{zu}$ in place of $(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})\widehat{\Sigma}_{ee}$ and $\hat{\Omega}_{ee}$ in place of $\widehat{\Sigma}_{ee}$. Both changes are motivated by our application to realized volatility forecasting, where, as Table 1 clearly demonstrates, volatility is very persistent, in contrast to the application in Kostakis et al. (2015), who consider predictions of stock returns. Hence, a long-run variance estimator is needed for robust inference in our setting. This issue is acknowledged as well as discussed on Kostakis et al. (2015, p. 1516). Finally, note that the selection of tuning parameters in this setting, $\boldsymbol{C}_z$ and β, are follow the latter, see Table 6 for details.

Research Papers 2018

CREATES
Center for Research in Econometric Analysis of Time Series

2017-31:	Martin M. Andreasen, Jens H.E. Christensen and Glenn D. Rudebusch: Term Structure Analysis with Big Data
2017-32:	Timo Teräsvirta: Nonlinear models in macroeconometrics
2017-33fa:	Isabel Casas, Eva Ferreira and Susan Orbe: Time-varying coefficient estimation in SURE models. Application to portfolio management
2017-34:	Hossein Asgharian, Charlotte Christiansen, Ai Jun Hou and Weining Wang: Long- and Short-Run Components of Factor Betas: Implications for Equity Pricing
2017-35:	Juan Carlos Parra-Alvarez, Olaf Posch and Mu-Chun Wang: Identification and estimation of heterogeneous agent models: A likelihood approach
2017-36:	Andrés González, Timo Teräsvirta, Dick van Dijk and Yukai Yang: Panel Smooth Transition Regression Models
2017-37:	Søren Johansen and Morten Ørregaard Nielsen: Testing the CVAR in the fractional CVAR model
2017-38:	Nektarios Aslanidis and Charlotte Christiansen: Flight to Safety from European Stock Markets
2017-39:	Tommaso Proietti, Niels Haldrup and Oskar Knapik: Spikes and memory in (Nord Pool) electricity price spot prices
2018-01:	Emilio Zanetti Chini: Forecaster's utility and forecasts coherence
2018-02:	Torben G. Andersen, Nicola Fusari and Viktor Todorov: The Pricing of Tail Risk and the Equity Premium: Evidence from International Option Markets
2018-03:	Torben G. Andersen, Nicola Fusari, Viktor Todorov and Rasmus T. Varneskov: Unified Inference for Nonlinear Factor Models from Panels with Fixed and Large Time Span
2018-04:	Torben G. Andersen, Nicola Fusari, Viktor Todorov and Rasmus T. Varneskov: Option Panels in Pure-Jump Settings
2018-05:	Torben G. Andersen, Martin Thyrsgaard and Viktor Todorov: Time-Varying Periodicity in Intraday Volatility
2018-06:	Niels Haldrup and Carsten P. T. Rosenskjold: A Parametric Factor Model of the Term Structure of Mortality
2018-07:	Torben G. Andersen, Nicola Fusari and Viktor Todorov: The Risk Premia Embedded in Index Options
2018-08:	Torben G. Andersen, Nicola Fusari and Viktor Todorov: Short-Term Market Risks Implied by Weekly Options
2018-09:	Torben G. Andersen and Rasmus T. Varneskov: Consistent Inference for Predictive Regressions in Persistent VAR Economies