CONSISTENT INFERENCE FOR PREDICTIVE REGRESSIONS IN PERSISTENT
ECONOMIC SYSTEMS

Torben G. Andersen
Rasmus T. Varneskov

Consistent Inference for Predictive Regressions in Persistent Economic Systems
Torben G. Andersen and Rasmus T. Varneskov
NBER Working Paper No. 28568
March 2021
JEL No. G12,G17

ABSTRACT

We study standard predictive regressions in economic systems governed by persistent vector autoregressive dynamics for the state variables. In particular, all – or a subset – of the variables may be fractionally integrated, which induces a spurious regression problem. We propose a new inference and testing procedure – the Local speCtruM (LCM) approach – for joint significance of the regressors, that is robust against the variables having different integration orders and remains valid regardless of whether predictors are significant and if they induce cointegration. Specifically, the LCM procedure is based on fractional filtering and band-spectrum regression using a suitably selected set of frequency ordinates. Contrary to existing procedures, we establish a uniform Gaussian limit theory and a standard 2-distributed test statistic. Using LCM inference and testing techniques, we explore predictive regressions for the realized return variation. Standard least squares inference indicates that popular financial and macroeconomic variables convey valuable information about future return volatility. In contrast, we find no significant evidence using our robust LCM procedure. If anything, our tests support a reverse chain of causality: rising financial volatility predates adverse innovations to macroeconomic variables. Simulations illustrate the relevance of the theoretical arguments for finite-sample inference.

Torben G. Andersen
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208
and NBER
t-andersen@kellogg.northwestern.edu

Rasmus T. Varneskov
Copenhagen Business School
Department of Finance
Solberg Plads 3
2000 Frederiksberg
Denmark
rtv.fi@cbs.dk

1 Introduction and Literature Review

An important function of econometric analysis is the provision of forecasts for economic and financial variables. A commonly adopted approach is predictive regressions, in which the variable of interest is regressed on a set of lagged predictors, generating coefficient estimates that constitute the basis for out-of-sample forecasts. However, this procedure has been subject to controversy due to the inherent nature of economic data. Many relevant series and associated predictors are highly persistent, inducing an array of inferential problems. In particular, economic time series often display slowly decaying autocorrelations and even outright non-stationarities. In such scenarios, a number of pitfalls arise. Most prominently, in the spurious regression case, where one persistent unit root, or $I(1)$, process is projected onto another independent $I(1)$ process, standard significance tests display large size distortions; see, e.g., Granger & Newbold (1974) and Phillips (1986). Another canonical example is the prediction of a noisy, and possibly persistent, dependent variable using a highly persistent regressor. This setting is motivated by, among others, predictability of stock returns or return volatility and associated empirical issues such as unstable predictive relations, e.g., Peseran & Timmermann (1995) and Welch & Goyal (2008), and predictive biases, e.g., Stambaugh (1999).

This paper considers inference and testing for predictive regressions, in which all the variables may be highly persistent. Formally, they 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 vector autoregressions (VARs), if all variables have $d = 0$ or $d = 1$, respectively, but also accommodates many intermediate cases, thus comprising a very flexible setting. The variables are (asymptotically) stationary if $0 \le d < 1/2$, and non-stationary if $d \ge 1/2$.

Our analysis builds on prior contributions studying different aspects of the above scenarios. For spurious regressions, the theory in Phillips (1986) is extended by Tsay & Chung (2000), demonstrating how many basic insights carry over to regressions with independent, fractionally integrated variables. Ng & Perron (1997) show that spurious inference problems also appear in systems with unit root and near unit root variables. In addition, 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) demonstrate that this, more generally, is infeasible, when the regressor displays local-to-unity, unit root or explosive persistence. In summary, 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 alleviate spurious inference problems and 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 parametric inference, and Robinson & Marinucci (2003) and

Christensen & Nielsen (2006) for the semiparametric case. Moreover, robust inference procedures have been proposed for the case where the persistent regressor follow local-to-unity dynamics; e.g., using Bonferroni corrections as in Cavanagh, Elliott & Stock (1995) and Campbell & Yogo (2006), a conditional likelihood approach in Jansson & Moreira (2006), and nearly optimal tests in Elliott, Müller & Watson (2015). Phillips & Lee (2013) and Phillips (2014) discuss issues associated with these approaches, e.g., the lack of power of the Bonferroni procedure, if the regressor is stationary rather than local-to-unity, and the lack of extendability to multivariate regressions. As an alternative, they develop the IVX methodology, which applies generally to persistent autoregressive processes (stationary, local-to-unity, unity and explosive) as well as to multivariate testing problems; with extensions and applications pursued by, among others, Kostakis, Magdalinos & Stamatogiannis (2015). Finally, Sizova (2013) considers predictive return regressions using stationary fractionally integrated variables, including historical volatility, but from a different asymptotic perspective.

Although inference in systems with persistent variables has received considerable attention, no results apply 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. In the stationary-persistent variable prediction literature, including the IVX methodology, the regressand is only weakly dependent, not fractionally integrated. Finally, the setting with diverse values of d puts us squarely outside the local-to-unity framework.

Consequently, in this paper, we seek to fill a critical 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. Importantly, our methodology is robust to stationary and non-stationary variables. Moreover, if the regressors are significant, we simultaneously accommodate predictive relations that may or may not be cointegrated. Such uniformity across persistent predictive regimes is highly desirable, allowing us to retain statical power, while letting the user, a priori, remain agnostic about the stationarity of the system as well as whether it exhibits predictability.

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 $d \geq 1$), first-stage filtering errors, and other potential sources of bias. We show that the LCM inference is asymptotically Gaussian and the associated test for joint significance of the regressors is χ^2-distributed, and thus readily implementable. These results apply in spite of the original variables consisting of a, possibly highly diverse, mixture of stationary and non-stationary variables *and* regardless of whether the (potentially) significant predictors induce cointegration. Importantly, our estimator achieves a semi-parametric convergence rate in the absence of cointegration and properties analogous to "super consistency" in its presence. 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 is in contrast to Phillips & Lee (2013), who find the corresponding bias to be of first order and uncorrectable in the local-to-unity case. Along the same lines, we show that LCM can accommodate regressors that have been pre-estimated as fractional "cointegrating residuals". Finally, we provide a feasible inference procedure that includes a new trimmed exact local Whittle (TELW) estimator of the fractional cointegration strength.

To illustrate the workings of the LCM approach, we briefly discuss two polar cases. In the first scenario, we have a balanced system with the integration order of the regressand, $d > 0$, matching the integration order among the regressors. In this scenario, inference is generally non-standard, including the case of unit roots generating non-standard convergence rates or spurious regression relations. Without exploiting a priori knowledge of whether the system is balanced or not, the LCM approach affords a simple procedure to obtain standard inference, independent of the persistence in the system, albeit at some cost in efficiency. The second case involves unknown degrees of integration across variables and possibly an unbalanced system. The LCM approach enables us to convert the system to an approximately balanced and stationary setting, and to directly test for predictability of the regressand through (weak) dependence on the regressors.

The methodological contributions closest to ours include Shao (2009), Maynard, Smallwood & Wohar (2013), Christensen & Varneskov (2017), and Müller & Watson (2018). 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.[1] Müller & Watson (2017, 2018) study the different problem of drawing inference and generating predictions about the very long run. For that purpose, they estimate a bivariate spectral density using low-frequency trigonometric averages of observations, thus extracting information about the "long-run" coherence between two series from a fixed number of frequencies in the vicinity of the origin. The asymptotic properties of this low-frequency methodology depends on the persistence of the system, with a potentially very slow convergence rate, e.g., when d is close to $1/2$, and it generates confidence intervals that are, mainly, computed by numerical means. In contrast, the LCM procedure captures information about the coherence from a wider range of the spectrum and directly accommodates multivariate systems, allowing for more general statements about predictability, the rate of convergence is invariant to d, and the procedure enjoys properties analogous to "super consistency," if the predictive relation is cointegrating. Moreover, the LCM approach achieves closed-form Gaussian inference regardless of the, possibly different, integration orders of the variables and whether they exhibit cointegration, when significant. That is, contrary to existing methodologies, the LCM inference and tests remain uniformly valid across a range of empirically relevant predic-

---

[1]Generally, the first-stage filtering error impacts both the rate of convergence and the asymptotic distribution theory.

tion scenarios.[2] 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. Finally, none of these methodologies accommodate endogenous regressors or regressors that have been obtained as residuals from a pre-estimated fractional co-integrating relation, in contrast to the LCM procedure.

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 (2018). 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). Even though some of these studies apply 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 explicitly account for the persistence in the state variables or the long memory in the realized market variance. The latter feature is a stylized fact of the financial econometrics literature.[3] 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 volatility forecasting is highlighted by, among others, Campbell et al. (2018). We first document that all these variables may be characterized as fractionally integrated processes; realized variance as a persistent stationary process, the state variables as non-stationary ones. Second, we confirm, in line with the extant literature, that all three state variables, seemingly, are 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), showing 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

---

[2]Despite achieving a memory-dependent and potentially very slow rate-of-convergence as well as providing less general inference within fractionally (co)integrated systems, it is important to note that the Müller & Watson (2017, 2018) procedure also applies to local-level and local-to-unity models, thus adding generality along that dimension.

[3]See, for example, Baillie, Bollerslev & Mikkelsen (1996), Comte & Renault (1998), Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001, 2003), Christensen & Nielsen (2006), Andersen, Bollerslev & Diebold (2007), Corsi (2009), Bollerslev, Osterrieder, Sizova & Tauchen (2013) and Varneskov & Perron (2018).

4

in finite samples. Fourth, when testing the predictive ability of the state variables using the robust LCM procedure, we fail to find significant predictive power. Finally, we test for a reverse predictive relation, i.e., 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. This suggest that standard least squares techniques may generate spurious findings regarding realized variance predictions, while the LCM inference procedure may help uncover new predictive relations.

The paper proceeds as follows. Section 2 provides the setup, defines our notion of regression balance, and reviews problems associated with standard predictive regressions. The local spectrum (LCM) approach is introduced in Section 3. Section 4 discusses robustness to the endogeneity bias and feasibility of including fractionally cointegrating residuals among the regressors. Section 5 describes the data and establishes baseline OLS evidence. Section 6 contains a simulation study documenting the finite sample performance of the LCM procedure. Section 7 explores realized variance prediction using LCM and considers reverse predictive relations. Finally, Section 8 concludes, while proofs and additional details on the data and simulation results are relegated to a Supplementary Appendix.

2 Predictive Regressions with Persistent Variables

This section introduces a regression framework in which all variables may exhibit long memory of different orders. The setting is used to study predictive regressions within persistent economic systems. We further discuss the possibility of drawing spurious inference and issues with incompatibility of the regression models under the null and alternative hypotheses. These considerations motivate our notion of balancedness and, subsequently, the design of the LCM testing procedure in Section 3.

2.1 Predictive Regression Assumptions

We observe a $(k+1) \times 1$ vector $\boldsymbol{z}_t = (y_t, \boldsymbol{x}'_{t-1})'$, which is assumed to obey,

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

where $\boldsymbol{\mu}$ is a $(k+1) \times 1$ vector of nonrandom unknown finite numbers, either the means or initial values of the variables, $\boldsymbol{v}_t = (e_t, \boldsymbol{u}'_{t-1})'$ is a weakly dependent vector process, and $\boldsymbol{D}(L) = \mathrm{diag}[(1 - L)^{d_1}, \ldots, (1 - L)^{d_{k+1}}]$, with $(1 - L)^d$ being a generic fractional filter,

$$(1 - L)^d \;=\; \sum_{i=0}^{\infty} \frac{\Gamma(i - d)}{\Gamma(i + 1)\Gamma(-d)}\, L^i, \tag{2}$$

and $\Gamma(\cdot)$ is the gamma function. In this setting, we define and study the predictive relation between the variables y_t and $\boldsymbol{x}_{t-1}$ through their weakly dependent components. Specifically, we assume,

$$e_t = \boldsymbol{\mathcal{B}}'\boldsymbol{u}_{t-1} + \eta_t^{(b)}, \qquad t = 1, \ldots, n, \tag{3}$$

5

where $\eta_t^{(b)} = (1 - L)^b \eta_t$ for some constant $b \geq 0$ and $\eta_t \sim I(0)$. Importantly, however, while the predictive relation between y_t and $\boldsymbol{x}_{t-1}$ is characterized by equation (3), this is tantamount to a balanced model for the persistent, observable variables,

$$y_t = a + \boldsymbol{\mathcal{B}}' \boldsymbol{\mathcal{Q}}(L)\boldsymbol{x}_{t-1} + v_t, \qquad t = 1, \dots, n, \tag{4}$$

where $\boldsymbol{\mathcal{Q}}(L) = \boldsymbol{D}_x(L)(1 - L)^{-d_1}$, with $\boldsymbol{D}_x(L)$ being the $k \times k$ lower-right submatrix of $\boldsymbol{D}(L)$, $a = \mu_y - \boldsymbol{\mathcal{B}}' \boldsymbol{\mathcal{Q}}(L)\boldsymbol{\mu}_x$ for $\boldsymbol{\mu} = (\mu_y, \boldsymbol{\mu}_x')'$ and $v_t = (1 - L)^{-d_1} \eta_t^{(b)} = (1 - L)^{b-d_1} \eta_t$.

The observation equation (1) in conjunction with the functional forms (3)-(4) encompass most multivariate fractionally integrated systems in the literature. Moreover, its design offers two particular advantages, as detailed in Section 2.2. First, it ensures a well-defined and *balanced* relation between the observable series y_t and $\boldsymbol{x}_{t-1}$, despite the variables being allowed to exhibit different degrees of fractional integration. Second, it accommodates inference scenarios with stationary and nonstationary variables that may ($b > 0$) or may not ($b = 0$) cointegrate. It is important to note, however, that the balanced relations (3)-(4) are not directly observable as the persistence matrix, $\boldsymbol{D}(L)$, is generally unknown ex-ante. We will show how to overcome this challenge below.

Before detailing the implications of the predictive setting, we impose some formal structure on the components of the model. To this end, we let "$\sim$" signify that the ratio of the left- and right-hand-side tends to one in the limit, element-wise. Then, following the fractional cointegration analyses in, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Christensen & Varneskov (2017), we impose the regularity conditions in terms of $\boldsymbol{q}_t = (\boldsymbol{u}_{t-1}', \eta_t)'$ rather than $\boldsymbol{v}_t$, when deriving the asymptotic properties for the proposed local spectrum procedure.

Assumption D1. *The vector process $\boldsymbol{q}_t$, $t = 1, \dots,$ 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 $i = 1, \dots, k$.*

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 i-th row of $\boldsymbol{A}(\lambda) = \sum_{j=0}^\infty \boldsymbol{A}_j e^{\mathrm{i}j\lambda}$, its partial derivative satisfies $\|\partial \boldsymbol{A}(\lambda, i)/\partial\lambda\| = O(\lambda^{-1}\|\boldsymbol{A}(\lambda, i)\|)$ as $\lambda \to 0^+$, for $i = 1, \dots, k+1$.*

The regularity conditions in Assumptions D1-D3 are standard in the literatures analyzing semi-parametric fractional co-integration as well as the estimation of multivariate fractional models, e.g., Shimotsu (2007) and Nielsen (2015). Specifically, D1 and D3 impose a mild rate of convergence for

the spectral density matrix $\boldsymbol{f}_{qq}(\lambda)$ in the vicinity of the origin, which depends on the smoothness parameter $\varpi \in (0, 2]$. Moreover, D1 requires full rank of $\boldsymbol{u}_{t-1}$ and it being locally exogenous to η_t as $\lambda \to 0^+$, but not global exogeneity. Finally, D2 specifies linearity, martingale and moment conditions for $\boldsymbol{q}_t$, allowing for general multivariate dependence, but rules out time-variation in the conditional covariance between the innovations as well as their third and fourth conditional moments.

Importantly, these conditions allow $\boldsymbol{z}_t$ to obey vector ARFIMA processes (for $\varpi = 2$), thereby nesting VAR dynamics as a special case with $d_1 = \cdots = d_{k+1} = b = 0$, which is commonly used to describe predictive economies in macro finance; see, e.g., Campbell & Vuolteenaho (2004), Bansal et al. (2014), and Campbell et al. (2018). Similarly, an integrated VAR system is recovered by imposing unit root dynamics, $d_1 = \cdots = d_{k+1} = 1$ and $b = 0$. Generally, however, the properties of the variables depend critically on their integration orders. First, if $0 \le d_i < 1/2$, the ith variable is (asymptotically) stationary with long memory, whenever $d_i > 0$. Long memory processes feature hyperbolically decaying autocovariances, contrary to the geometric decay for short memory processes (with $d_i = 0$). Second, the variable is non-stationary if $d_i \ge 1/2$, but has a well-defined mean for $d_i < 1$. Hence, the flexibility of (1) is particularly useful for characterizing the dynamic properties of multivariate systems, whose components are very persistent, yet display different degrees of persistence, which is often the case for applications with multiple financial and macroeconomic variables.

The main departure from the standard fractional (co-)integration settings is the flexible predictive setup encompassed by equations (1) and (3), allowing all variables to exhibit different integration orders *and* accommodating cases where cointegration is not ex-ante imposed on the system. In addition to these features, the fact that the dependent variable may also display strong persistence sets our framework apart from prior studies, which develop robust inference for predictive regressions with local-to-unity regressors. The relevance of these fairly subtle features are discussed in detail next.

2.2 Regression Balance and Inference Considerations

The relation in (4) has several distinct features. First, to be well-defined, despite the observable regressors possibly having different fractional integration orders, the predictive model must be balanced under both,

$$\mathcal{H}_0 : \boldsymbol{\mathcal{B}} = \boldsymbol{0}, \qquad \text{and} \qquad \mathcal{H}_A : \boldsymbol{\mathcal{B}} \ne \boldsymbol{0}. \tag{5}$$

Hence, irrespective of the forecasting prowess of the regressors, the implied dynamics of the prediction model is required to be consistent with the observed dynamics of the dependent variable, $y_t \sim I(d_1)$. To this end, the fractional filter $\boldsymbol{\mathcal{Q}}(L)$ adjusts the persistence of $\boldsymbol{x}_{t-1}$ to ensure regression balance under both $\mathcal{H}_0$ and $\mathcal{H}_A$. If the system is balanced, then $\boldsymbol{\mathcal{Q}}(L) = \boldsymbol{I}_k$, a k-dimensional identity matrix, and the adjustment is trivial. Second, if the regressors contain forecasting power, the model may exhibit cointegration. We entertain three distinct scenarios:

(i) $\boldsymbol{\mathcal{B}} = \boldsymbol{0}$ and $b = 0$: $\boldsymbol{x}_{t-1}$ contains no predictive power for y_t.

(ii) $\boldsymbol{\mathcal{B}} \ne \boldsymbol{0}$ and $b = 0$: $\boldsymbol{x}_{t-1}$ partially spans the persistent component(s) of y_t.

(iii) $\boldsymbol{\mathcal{B}} \neq \boldsymbol{0}$ and $b > 0$: $\quad \boldsymbol{x}_{t-1}$ cointegrate with y_t, spanning its persistent component(s).

We may illustrate the three scenarios through a simple univariate regression setting with integrated series, y_t and x_{t-1}, where y_t contains two separate components, $y_t = y_{1,t} + y_{2,t}$, and both $y_{1,t}$ and $y_{2,t}$ are integrated of order one. In case (i), we obtain a standard spurious regression (4), while the first-differenced representation in equation (3) contains a stationary innovation term, so $b = 0$. For case (ii), let $y_{1,t} = \mathcal{B}'x_{t-1} + u_{1,t}$ and $y_{2,t} = y_{2,t-1} + u_{2,t}$, where $u_{1,t}$ and $u_{2,t}$ are independent i.i.d. innovations. Now, the regressor partially spans the persistent y_t variable, but the error term in equation (4) will be integrated, and the first-differenced system in (3) features a stationary error, $b = 0$. Finally, case (iii) arises from case (ii) with $y_{2,t} = 0$ for all $t \geq 0$. Here, regression (4) is a standard cointegrating relation, the persistent component of y_t is fully spanned by the regressor, and equation (3) involves (over-)differencing of a stationary error term, $b = 1$.

The three competing hypotheses above demonstrate that rigorous testing for predictive ability in persistent systems require estimation and inference procedures, which remain valid across distinct settings, because it generally is unknown, ex ante, whether the predictors are significant and, if they are, whether the system features cointegration. Given the persistence of the variables in system (1) as well as the possibility of cointegration, it is well-known that standard OLS procedures applied to y_t and $\boldsymbol{x}_{t-1}$ produces spurious inference, unless $d_1 = \cdots = d_{k+1} = 0$ and $b = 0$, e.g., Granger & Newbold (1974), Phillips (1986) and Tsay & Chung (2000). Moreover, standard (fractional) cointegration settings, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006), and Johansen & Nielsen (2012) impose balanced memory $d_1 = \cdots = d_{k+1} = d$ *and* $0 < b \leq d$ for validity of their respective inference procedures, thereby assuming the regressors have predictive power under the null hypothesis. Thus, examining and testing for balanced predictability in systems involving persistent variables with, possibly, distinct degrees of memory require the development of new inference techniques.

Third, we note that regression balance may be achieved through a different channel than the stipulated persistence adjustment, $\boldsymbol{\mathcal{Q}}(L)$. Specifically, if a cointegrating relation exists among the regressors, possibly in conjunction with variables external to the economic system, a "cointegrating residual" may be included in lieu of the original regressors. This involves additional complications, however, as it requires a priori identification of the cointegrating relation, since we assume the set of regressors included in the predictive analysis to have full rank, cf. Assumption D1. Moreover, these "residuals" may require pre-estimation of the cointegrating vectors. Consequently, for ease of exposition, we defer formal treatment of this scenario to Section 4.2, while stressing that our LCM inference techniques retain robustness to this case. Finally, to further illustrate the notion of regression balance and associated inference challenges for persistent economic systems, we provide a number of additional examples.[4] Importantly, these scenarios are all encompassed by our predictive setting and may be studied using the inference procedures developed below.

---

[4]In a companion paper, we provide detailed accounts of spurious inference results as well as biases and degeneracies for unbalanced OLS regressions; see Andersen & Varneskov (2019b). Specifically, we adapt results from Tsay & Chung (2000) and Robinson & Marinucci (2003) to describe OLS and apply simulations to illustrate the arguments.

Example 1 (Balanced System, Standard Integration)**.** *Suppose $y_t \sim I(1)$ and $\boldsymbol{x}_{t-1} \sim I(1)$. As discussed above, our framework then embeds: (i) unit root regressions for $b = 0$; (ii) standard cointegrating regressions for $b = 1$. In these settings, OLS generally produces spurious inference, e.g., Granger & Newbold (1974) and Phillips (1986) . However, if the true cointegration structure, and thus predictability, is known ex-ante, inference procedures are available, e.g., Phillips (1991) and Johansen (1996).*

Example 2 (Balanced System, Fractional Integration)**.** *Suppose $y_t \sim I(d)$ and $\boldsymbol{x}_{t-1} \sim I(d)$, $0 \leq d < 2$, then our framework contains: (i) long-memory regressions for $b = 0$; (ii) fractional cointegrating regressions for $0 \leq b \leq d$. In these settings, OLS similarly produces spurious inference, e.g., Tsay & Chung (2000). However, if the true (fractional) cointegration structure, and thus predictability, is known ex-ante, inference procedures are available, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006), Robinson & Hualde (2003) and Johansen & Nielsen (2012).*

Example 3 (Unbalanced System, Standard Integration)**.** *Suppose $y_t \sim I(0)$ and $\boldsymbol{x}_{t-1} \sim I(1)$, then the predictive system is unbalanced, which have the following implications:*

- *There exists no "standard linear prediction model" for which $y_t \sim I(0)$ under both $\mathcal{H}_0$ and $\mathcal{H}_A$; it is incompatible with regression balance.*

- *The OLS coefficient estimate, $\widehat{\boldsymbol{B}}$, decomposes as $\widehat{\boldsymbol{B}} \simeq \frac{2}{n}\boldsymbol{B} + \mathrm{error}(b)$, i.e., it is degenerate with an error that depends on whether there is cointegration between y_t and $\Delta\boldsymbol{x}_{t-1}$.*

As in the balanced case, problems with spurious inference is well-known from the literature on stationary-persistent predictive relations, e.g., Ferson et al. (2003), Deng (2014) and Phillips (2015). However, our framework and associated definition of regression balance stipulate that a well-defined predictive relation may exists between: (i) y_t and $\Delta\boldsymbol{x}_{t-1}$; (ii) y_t and a cointegrating residual involving $\boldsymbol{x}_{t-1}$ and some other external $I(1)$ variables, which may have to be pre-estimated.

Example 4 (Unbalanced System, Fractional Integration)**.** *Suppose $y_t \sim I(d_y)$ and $\boldsymbol{x}_{t-1} \sim I(d_x)$, $0 \leq b \leq d_y < d_x$, then the system is unbalanced, which have the following implications:*

- *There exists no "standard linear prediction model" for which $y_t \sim I(d_y)$ under both $\mathcal{H}_0$ and $\mathcal{H}_A$; it is incompatible with regression balance.*

- *Moreover, the OLS coefficient estimate, $\widehat{\boldsymbol{B}}$, decomposes as $\widehat{\boldsymbol{B}} \simeq \boldsymbol{B} \times \mathcal{S}_n(d_y, d_x) + \mathrm{error}(d_y, d_x, b)$ where $\mathcal{S}_n(d_y, d_x)$ is a, possibly, degenerate function of the sample size, depending on (d_y, d_x), which captures the bias of the point estimate, and the error depends on (d_y, d_x, b).*

As in the balanced case, Tsay & Chung (2000) show that such systems is subject to spurious inference. However, our framework and associated notion of regression balance stipulate that a well-defined predictive relation may exists between: (i) y_t and $\boldsymbol{\mathcal{Q}}(L)\boldsymbol{x}_{t-1}$; (ii) y_t and a fractional cointegrating residual involving $\boldsymbol{x}_{t-1}$ and some other external $I(d_x)$ variables, which may have to be pre-estimated.

Our notion of predictability, specified through balanced regression in the form of equations (3) and (4), stipulates that either $\boldsymbol{\mathcal{B}} = \mathbf{0}$ or $\boldsymbol{\mathcal{B}} \neq \mathbf{0}$. That is, we focus on distinct persistence scenarios for which the memory of each variable can be consistently estimated. Consequently, we rule out local asymptotic alternatives of the form $\boldsymbol{\mathcal{B}} = \boldsymbol{b}/n^{\gamma}$ for some $\gamma > 0$ large enough for the memory of y_t to become unidentifiable, even asymptotically. This requirement is formalized by Assumption F in the next section. Furthermore, we briefly discuss this scenario in the context of the empirical application below. A separate treatment for this case is provided in Andersen & Varneskov (2019a), where we analyze the possibility of asset returns having a "hidden" fractionally integrated component.

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 a spectral density decomposition of the vector $\boldsymbol{z}_t$, before describing its two-step implementation in Sections 3.2 and 3.3. Section 3.4 develops the asymptotic theory. Section 3.5 provides a new estimator for the fractional cointegration strength. Note that Section 4 discusses theoretical extensions such as robustness to the regressor endogeneity bias and regressors that are generated as cointegration residuals. The material in this section, however, suffices for readers only interested in the core LCM procedure and asymptotic theory.

3.1 Motivating Observations and Implications for Inference

The intuition behind the local spectrum inference and testing procedure is readily conveyed by the relation between the *observable* dynamics (1) and the *latent* predictive models (3) and (4). Specifically, whereas we associate predictability with interaction in a balanced relation, the observable system (1) may contain different integration orders, which, together with Assumption D1-D3, implies that the spectral densities of the regressors and dependent variable behave according to,

$$\boldsymbol{f}_{xx}(\lambda) \sim \boldsymbol{\Lambda}_{xx}^{-1} \, \boldsymbol{G}_{uu} \, \overline{\boldsymbol{\Lambda}}_{xx}^{-1}, \qquad \boldsymbol{f}_{yy}(\lambda) \sim \Lambda_{yy}^{-1} \, \boldsymbol{\mathcal{B}}' \, \boldsymbol{G}_{uu} \, \boldsymbol{\mathcal{B}} \, \overline{\Lambda}_{yy}^{-1} + \Lambda_{y-\eta}^{-1} G_{\eta\eta} \overline{\Lambda}_{y-\eta}^{-1}, \qquad (6)$$

respectively, where Λ_{yy} and $\boldsymbol{\Lambda}_{xx}$ are the first element and lower-right $k \times k$ submatrix of $\boldsymbol{\Lambda}$, defined as,

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

and $\Lambda_{y-\eta} = (1 - e^{\mathrm{i}\lambda})^{d_1 - b}$, with $\overline{\boldsymbol{\Lambda}}$ and $\overline{\Lambda}_{y-\eta}$ denoting the corresponding complex conjugates, and $\mathrm{i} = \sqrt{-1}$. Importantly, these spectral densities, and thus the implied properties of the second moments, have different divergence rates in the vicinity of the origin, depending on the fractional integration orders. However, the corresponding spectral densities of the unobservable weakly dependent components $\boldsymbol{u}_{t-1}$ and e_t, $\boldsymbol{f}_{uu}(\lambda) \sim \boldsymbol{G}_{uu}$ and $\boldsymbol{f}_{ee}(\lambda) \sim \boldsymbol{\mathcal{B}}' \, \boldsymbol{G}_{uu} \, \boldsymbol{\mathcal{B}} + \Lambda_{-\eta}^{-1} G_{\eta\eta} \overline{\Lambda}_{-\eta}^{-1}$, with $\Lambda_{-\eta} = (1 - e^{\mathrm{i}\lambda})^{-b}$, are asymptotically bounded and convey similar information about $\boldsymbol{\mathcal{B}}$, i.e., about the presence of balanced predictability in equation (4). This suggests that inference based on $\boldsymbol{u}_{t-1}$ and e_t may circumvent

concerns regarding regression balance, degeneracy of point estimates, and spurious inference.

Hence, we propose a two-step procedure in which the variables initially are stripped of their persistence via *fractional filtering*, that is, we form an estimate of $\boldsymbol{v}_t = (e_t, \boldsymbol{u}'_{t-1})'$. Second, we introduce a consistent frequency domain estimator of $\boldsymbol{\mathcal{B}}$, explicitly accounting for the first-stage estimation and filtering errors. This local spectrum (LCM) procedure is, thus, designed to achieve balancedness in an agnostic way by purging all variables of their fractional integration, before carrying out robust frequency domain estimation and testing for predictive ability, that is void of spurious inference concerns because, effectively, the fractionally filtered economic system is only weakly dependent.

Remark 1. *We rely on an exact spectral density representation around the origin in equation (6), 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 suitable approximation when $d < 1/2$, but it deteriorates as d moves into the non-stationary range, eventually generating inconsistent estimators of the integration order.*

Remark 2. *Frequency domain methods are often used for estimation involving fractionally integrated series. However, it follows from equation (6) that, if an entry has $d_i \neq d_1$, then the coefficient for this explanatory variable cannot be consistently estimated using standard local band-spectrum estimators applied to y_t and $\boldsymbol{x}_{t-1}$, as such estimates are degenerate, asymptotically. Of course, Examples 1-4 illustrate that similar problems plague least squares inference procedures in this context.*

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 to have an estimator, $\widehat{d_i}$ 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 $0 < \varrho \leq 1$. For all $i = 1, \ldots, k+1$ elements of $\boldsymbol{z}_t$, we assume to have an estimator with the property,*

$$\widehat{d_i} - d_i = O_p\left(1/\sqrt{m_d}\right), \quad \text{and we then let,} \quad \widehat{\boldsymbol{D}}(L) = \mathrm{diag}\left[(1 - L)^{\widehat{d_1}}, \ldots, (1 - L)^{\widehat{d_{k+1}}}\right].$$

Assumptions F is very mild, essentially only requiring the existence of an estimator which, under Assumptions D1-D3, is consistent. Of course, we need such consistency to hold for a wide range of d_i. To simplify the subsequent analysis, we impose another mild restriction,

$$0 \leq d_i < 2, \quad \text{for } i = 1, \ldots, k+1, \quad \text{and we then define,} \quad \underline{d} = \min_{i=1,\ldots,k+1} d_i. \tag{7}$$

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.

In addition to condition (7), the cointegration strength parameter, b, needs to be restricted under the alternative hypothesis, $\mathcal{H}_A$. Specifically, we impose the condition,

$$0 \leq b \leq \min(\underline{d}, 1). \tag{8}$$

If the observable system (1) is balanced, the intuition behind restriction (8) is straightforward; our framework accommodates $I(1) - I(0)$ cointegration, fractional cointegration and settings without cointegration, as detailed in Examples 1 and 2. If the variables have different integration orders, the condition is more subtle, yet remains intuitive. In particular, the cointegrating relation must be balanced *and* its residuals, at least, weakly dependent. To see this, suppose $y_t \sim I(0.4)$ and $\boldsymbol{x}_{t-1} \sim I(0.8)$. Then the upper bound $b = 0.4$, in conjunction with the regression balance requirement (4), implies a balanced cointegrating relation with $I(0)$ residuals. Conversely, if $y_t \sim I(0.8)$ and the predictors $\boldsymbol{x}_{t-1} \sim I(0.4)$, then the restriction $b = 0.4$ generates $I(0.4)$ residuals, meaning $\boldsymbol{x}_{t-1}$ cannot account for a higher degree of fractional integration in y_t than is implied by its own persistence.[5]

Given these restrictions and estimates for the system-wide integration orders $\widehat{\boldsymbol{D}}(L)$, the innovations, $\boldsymbol{v}_t$, are estimated by,

$$\widehat{\boldsymbol{v}}_t \equiv (\widehat{e}_t, \widehat{\boldsymbol{u}}_{t-1}')' = \widehat{\boldsymbol{D}}(L)\boldsymbol{z}_t, \tag{9}$$

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, $\widehat{\boldsymbol{D}}(L)\boldsymbol{\mu}$, in a unified manner during the second stage estimation.

We conclude the section with two examples of memory parameter estimators, one semiparametric and one parametric procedure, which are compatible with our framework.

Example 5 (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). The same holds for the trimmed ELW (TELW) estimator, which we introduce in Section 3.5 to estimate b.*

Example 6 (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 $\varrho = 1$.[6]*

---

[5]The upper bound may be changed to $\min(\underline{d}_s, 1)$, where $\underline{d}_s = \{\min_{2=,\ldots,k+1}(d_1 \wedge d_i) | \boldsymbol{\mathcal{B}}(i-1) \neq 0\}$, that is, the restriction in (8) applies only to the significant regressors. However, to ease exposition and avoid exceedingly complicated cross-restrictions on bandwidth and trimming parameters in Section 3.3, we refrain from making this distinction.

[6]Assumptions D1-D3 mirror the corresponding assumptions in Shimotsu & Phillips (2005, Assumptions 1'-3'), but we

3.3 Step 2: Medium Band Least Squares

After having computed $\widehat{\boldsymbol{v}}_t$, we estimate the parameter vector $\boldsymbol{\mathcal{B}}$ in equations (3)-(4) in a second step using a new frequency-domain least squares estimator. To this end, we let,

$$\boldsymbol{w}_h(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \boldsymbol{h}_t \, e^{\mathrm{i}t\lambda_j}, \qquad \boldsymbol{I}_{hk}(\lambda_j) = \boldsymbol{w}_h(\lambda_j)\, \overline{\boldsymbol{w}}_k(\lambda_j), \tag{10}$$

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 $\lambda_j = 2\pi j/n$ indexes the Fourier frequencies. Moreover, we let the real and imaginary decomposition of $\boldsymbol{I}_{hk}(\lambda_j)$ be denoted $\boldsymbol{I}_{hk}(\lambda_j) = \Re(\boldsymbol{I}_{hk}(\lambda_j)) + \mathrm{i}\Im(\boldsymbol{I}_{hk}(\lambda_j))$. Finally, define the trimmed discretely averaged co-periodogram (TDAC) as,

$$\widehat{\boldsymbol{F}}_{hk}(\ell, m) = \frac{2\pi}{n} \sum_{j=\ell}^{m} \Re(\boldsymbol{I}_{hk}(\lambda_j)), \qquad 1 \le \ell \le m \le n, \tag{11}$$

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

$$\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) = \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1} \, \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}(\ell, m), \tag{12}$$

for which $\ell, m \to \infty$ and $\ell/m + m/n \to 0$, as $n \to \infty$.[7] The MBLS estimator has some distinct advantages. First, we avoid being fully parametric about the dynamics of $\boldsymbol{z}_t$, needing only the 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, $\widehat{\boldsymbol{D}}(L)\boldsymbol{\mu}$, occurring at lower frequencies and, in conjunction with the bandwidth, the estimation errors in Assumptions F. Combining these features ensures robust testing for the predictive power of the regressors in equations (3)-(4).

Remark 3. *Importantly, despite Assumptions D1-D3 only parameterizing the low-frequency part of the spectrum (as $\lambda \to 0^+$), we emphasize that the (latent) test for predictability in equation (3) 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.*

Christensen & Varneskov (2017) introduces the generic structure of the MBLS estimator (12) to analyze fractional co-integration among stationary long-memory processes in the presence of structural breaks and other low-frequency contaminants. Despite these similarities, their estimator differs from

---

need to impose slightly stronger differentiability assumptions in D3 to satisfy the (still very) mild regularity conditions of Nielsen (2015, Assumption C). We refer to the latter for details.

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

the one in equation (12) 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 (6), we accommodate non-stationary variables, whereas Christensen & Varneskov (2017) require stationarity for all variables *and* cointegration between y_t and $\boldsymbol{x}_{t-1}$. 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 $d_i \geq 1/2$). Furthermore, direct applicability of their estimator is subject to the issues outlined in Remark 2, 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 $d_i = d \in (0, 1/2)$, $i = 1, \ldots, k+1$, the differences behind the procedures will, as detailed below, have a first-order impact on the distribution theory.[8]

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. The following cross-restrictions are assumed to apply for ℓ, m, m_d and n,*

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

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 impose only local exogeneity in the spectrum (as $\lambda \to 0^+$) between the regressors, $\boldsymbol{u}_{t-1}$, in equation (3), 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 4.1. We stress, however, that our second stage MBLS estimator does not suffer from an asymptotic bias, regardless of the persistence of the VAR system (1). Note also that $\varpi = 2$ holds for the empirically relevant vector ARFIMA process, implying that $\kappa < 4/5$ must be satisfied.

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

$$\nu < (\varpi + \kappa(1/2 + b))/(1 + \varpi + b) \quad \text{and} \quad (1 - \kappa/2 - (2\underline{d} - b)(1 - \kappa))/2 < \nu, \tag{13}$$

---

[8]Drawing an analogy with the differences between the LW and ELW estimators, the Christensen & Varneskov (2017) procedure is reminiscent of the former and ours of the latter. This explains why, as described in the following section, our proofs bear resemblance to those in, e.g., Shimotsu & Phillips (2005) and Shimotsu (2010). Nonetheless, the relation to the MBLS estimator in Christensen & Varneskov (2017) does suggest some inherent robustness to outliers, structural breaks, Markov switching means, certain deterministic trends, etc., which are known to contaminate co-periodograms at frequency ordinates close to the origin, e.g., Perron & Qu (2010). While we do not formally analyze those effects here, they are currently being examined and will be the subject of future work.

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 $(3/5 - (2\underline{d} - b)/5)/2 < \nu < 4/5$. Since the lower bound is strictly decreasing in $2\underline{d} - b \geq 0$, the highest lower bound is attained for $\underline{d} = 0$ and equals $3/10$. The condition $0 < \nu < \kappa < \varrho \leq 1$ as well as the trimming restriction, $(1 - \varrho)/2 + b(1 - \kappa) < \nu$, are needed to eliminate additional errors stemming from the estimation of the 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 set $\kappa \simeq 4/5$, the restriction is determined by the cointegration strength parameter as $b/5 < \nu$, leaving the highest bound for $b = \min(\underline{d}, 1)$. If the estimator is semiparametric, and we select $\kappa < \varrho$ as well as ϱ arbitrarily close to $4/5$, the highest lower bound for the trimming becomes $1/10 + \min(\underline{d}, 1)/5 \leq 3/10 < \nu$. Finally, condition five imposes a mild lower bound for the bandwidth rate, $b/(1/2 + b) \leq \min(\underline{d}, 1)/(1/2 + \min(\underline{d}, 1)) < \kappa$.

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

Theorem 1. *Let Assumptions F, T, and D1-D3 hold. Moreover, suppose that $0 \leq d_i < 2$ for $i = 1, \ldots, k + 1$, $0 \leq b \leq \min(\underline{d}, 1)$ under $\mathcal{H}_A$, and $\max(0, (1 - 3\kappa/2)/(1 + \kappa/2)) < \varpi \leq 2$, then,*

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

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).[9] Moreover, its convergence rate depends on whether $\boldsymbol{\mathcal{B}} \neq \mathbf{0}$ and, if this is the case, whether there is cointegration. Specifically, the rate is $\sqrt{m}$ when cointegration is absent ($b = 0$), in line with well-known results for semiparametric estimators in the frequency domain, e.g., Brillinger (1981, Chapters 7-8), Robinson (1995), and Shimotsu & Phillips (2005). In contrast, the rate is $\sqrt{m}\lambda_m^{-b} \asymp \sqrt{m}(n/m)^b$, when (a subset of) the predictors are significant and $b > 0$. Hence, cointegration "helps" improve the rate of convergence of the MBLS estimator, in analogy with super consistency properties, and lowers its asymptotic variance, as conveyed by the scale $1/(2(1 + 2b))$. The Gaussian distribution theory is remarkable; holding uniformly for prediction scenarios with and without cointegration, across (asymptotically) stationary and non-stationary variables in the observable persistent system (1), as well as for both weak ($b < 1/2$) and strong ($b \geq 1/2$) cointegration settings. In fractional cointegration contexts, assuming predictability under $\mathcal{H}_0$, this uniformity neither applies to the semiparametric NBLS estimator ($\ell = 1$ in equation (3.3)) nor for maximum likelihood inference in the parametric fractionally cointegrated VAR model, see Johansen & Nielsen (2012), where the inference is Gaussian for stationary cases and exhibiting different forms of non-Gaussianity in non-stationary cases.

Interestingly, the asymptotic distribution for our second-stage MBLS estimator differs from the corresponding in Christensen & Varneskov (2017, Theorem 3) by being independent of the integration

---

[9]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 $0 \leq \underline{d} < 2$.

orders of the variables, $\{d_i\}_{i=1}^{k+1}$, by applying for stationary as well as non-stationary variables, and by being valid in the absence of cointegration ($b = 0$). Moreover, the asymptotic variance only depends on the noise-to-signal ratio, $G_{\eta\eta}G_{uu}^{-1}$, and the cointegration strength parameter, b. 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 $G_{\eta\eta}$ as well as the cointegration strength, b. Again, the main challenge is that we observe $\widehat{\boldsymbol{v}}_t$, not $\boldsymbol{v}_t$. Similarly, the residuals η_t are latent and we estimate them as,

$$\widehat{\eta}_t^{(b)} \; = \; \widehat{e}_t \; - \; \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)' \, \widehat{\boldsymbol{u}}_{t-1}, \qquad \widehat{\eta}_t \equiv (1 - L)^{-\widehat{b}} \, \widehat{\eta}_t^{(b)}, \tag{14}$$

where $\widehat{b}$ denotes a consistent estimator of b. Hence, not only are the residuals $\eta_t^{(b)}$ latent, but we also need to undo their (over-)differencing using b to recover η_t, which may then be employed to estimate $G_{\eta\eta}$. For ease of exposition, we handle these obstacles sequentially and assume having a valid cointegration strength estimator, $\widehat{b}$, available before introducing one in the following section.

Assumption B. *Let $m_b \asymp n^\varepsilon$ be a sequence of integers where $0 < \varepsilon \leq 1$, then $\widehat{b} - b = O_p\left(1/\sqrt{m_b}\right)$.*

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

$$\widehat{\boldsymbol{G}}_{hh}(\ell_G, m_G) \; = \; \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \Re\left(\boldsymbol{I}_{hh}(\lambda_j)\right), \tag{15}$$

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 (15) differs importantly by using fractionally filtered series as input and an exact spectrum representation (6), not an approximation valid only in the stationary case, $d_i < 1/2$, $i = 1, \ldots, k+1$. These points mirror the distinction among the MBLS estimators discussed earlier. Now, from equation (15), the asymptotic variance of $\widehat{\boldsymbol{\mathcal{B}}}(\ell, m)$ is estimated as,

$$\widehat{\mathbf{AVAR}} \; = \; \widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}(\ell_G, m_G)^{-1} \, \frac{\widehat{G}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) \, \lambda_m^{2\widehat{b}}}{2(1 + 2\widehat{b}) \, m} \tag{16}$$

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 $h \times k$ selection matrix $\boldsymbol{\mathcal{R}}$ and $h \times 1$ 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 define the sequence of intergers $\underline{m}_n = m_d \wedge m_b \wedge m$, then the following cross-restrictions are imposed on the trimming and bandwidth parameters: $n/(m_G \, \ell_G^2) + n^2/(m_G \, \ell_G^2 \, \underline{m}_n) \to 0$ as $n \to \infty$.*

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

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

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, rendering the procedure easy to implement in practice.

Remark 4. *Utilizing results from Phillips & Shimotsu (2004) and Shimotsu (2010), and furthermore letting $y_t \in I(d_y)$ and $\boldsymbol{x}_{t-1} \in I(d_x)$, our proofs of Theorem 1 illustrate that,*

$$\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) = \left(\frac{2\pi}{n} \sum_{j=\ell}^{m} D_n(e^{i\lambda_j}; d_x)^2 \Re(\boldsymbol{I}_{xx}(\lambda_j))\right)^{-1}$$

$$\times \left(\frac{2\pi}{n} \sum_{j=\ell}^{m} D_n(e^{i\lambda_j}; d_x) D_n(e^{i\lambda_j}; d_y) \Re(\boldsymbol{I}_{xy}(\lambda_j))\right) + \boldsymbol{\mathcal{E}}(\ell, m, m_d, n), \tag{17}$$

with $D_n(e^{i\lambda_j}; d_y)$ and $D_n(e^{i\lambda_j}; d_x)$ being defined in Appendix A.6 and $\boldsymbol{\mathcal{E}}(\ell, m, m_d, n)$ denoting an approximation error that is asymptotically negligible under trimming, locally as $m/n \to 0$. This illustrates that our LCM procedure is related to the frequency domain GLS estimators in Robinson & Hidalgo (1997) and Nielsen (2005) who, in parametric and semi-parametric frameworks, weight periodograms in the numerator and denominator by (the same) functions of the form $\lambda_j^{2d_v}$, d_v being the integration order of the residuals in the original regression (4). Hence, our LCM estimator, as conveyed by equation (17), differs importantly from their respective estimators by utilizing exact differencing, resulting in differential weights of the form $D_n(e^{i\lambda_j}; d)$, in conjunction with trimming.

Remark 5. *LCM estimation and inference in the latent regression (3) is not only useful for determining significance of the vector $\boldsymbol{x}_{t-1}$ in general economic systems, but may also readily be used for prediction of y_t. To see this, let $\pi_i = \Gamma(i - d_1)/(\Gamma(i+1)\Gamma(-d_1))$, then, utilizing equation (1), we obtain,*

$$\mathbb{E}_t(y_{t+1}) = \mu_y + \sum_{i=0}^{\infty} \pi_{i+1} (y_{t-i} - \mu_y) + \mathbb{E}_t(e_{t+1}).$$

Hence, once the mean (or initial value) μ_y and integration order d_1 are determined, e.g., using the estimators in Examples 5 and 6, we only need a forecast from the latent regression, $\mathbb{E}_t(e_{t+1})$, to generate a forecast for the dependent variable. This circumvents the need to develop a direct estimation strategy for the original regression (4), once the two steps of the LCM procedure have been applied.

3.5 Trimmed ELW Estimation of the Fractional Cointegration Strength

A critical ingredient for feasible inference and testing is consistent estimation of the cointegration strength parameter, b. This is challenging, however, as $\widehat{\eta}_t^{(b)}$ is estimated from a set of variables that may be stationary or non-stationary, have non-trivial means or initial values, and the estimator must be valid for an (over-)differenced series where, potentially, $0 \le b \le 1$. To accommodate these features, we introduce the trimmed ELW (TELW) estimator.

Specifically, we define the objective function,

$$\mathfrak{Q}\big(G_{\eta\eta}, \varphi, \widehat{\eta}_t^{(b)}\big) \;=\; \frac{1}{m_b - \ell_b + 1} \sum_{j=\ell_b}^{m_b} \left[\ln\left(G_{\eta\eta}\, \lambda_j^{-2\varphi}\right) + G_{\eta\eta}^{-1}\, I_{\widehat{\eta}\widehat{\eta}}^{(\varphi+b)}(\lambda_j)\right], \tag{18}$$

where $I_{\widehat{\eta}\widehat{\eta}}^{(\varphi+b)}(\lambda_j) = w_{\widehat{\eta}}^{(\varphi+b)}(\lambda_j)\, \overline{w}_{\widehat{\eta}}^{(\varphi+b)}(\lambda_j)$ is the periodogram of $\Delta^\varphi \widehat{\eta}_t^{(b)}$, and $\ell_b = \ell_b(n)$, $m_b = m_b(n)$ are TELW specific trimming and bandwidth sequences, respectively. Hence, φ represents a differencing argument, for which we denote its "true" value by $\varphi_0 = -b$. Importantly, in addition to being based on trimming, equation (18) differs from the objective function in Shimotsu & Phillips (2005) because the series is not directly observable, but rather estimated, and it may be impacted by non-trivial mean components of the observed variables. Similarly, it differs from the setting in Shimotsu (2010) by being based on an estimated series and utilizing trimming to eliminate mean effects.

Next, by concentrating $\mathfrak{Q}\big(G_{\eta\eta}, \varphi, \widehat{\eta}_t^{(b)}\big)$ with respect to $G_{\eta\eta}$, it follows that,

$$\widehat{\varphi} \;=\; \operatorname*{argmin}_{\varphi \in [\Delta_1, \Delta_2]} \mathfrak{R}\big(\varphi, \widehat{\eta}_t^{(b)}\big), \qquad \text{where we let} \quad \widehat{\varphi} = -\widehat{b}, \tag{19}$$

and for which the objective function is given by,

$$\mathfrak{R}\big(\varphi, \widehat{\eta}_t^{(b)}\big) \;=\; \ln \mathfrak{S}\big(\varphi, \widehat{\eta}_t^{(b)}\big) - \frac{2\varphi}{m_b - \ell_b + 1} \sum_{j=\ell_b}^{m_b} \ln(\lambda_j), \quad \widehat{\mathfrak{S}}\big(\varphi, \widehat{\eta}_t^{(b)}\big) = \frac{1}{m_b - \ell_b + 1} \sum_{j=\ell_b}^{m_b} I_{\widehat{\eta}\widehat{\eta}}^{(\varphi+b)}(\lambda_j).$$

Moreover, as pointed out by Shimotsu & Phillips (2005) and Shimotsu (2010), we must restrict the bounds on the admissible values of φ. Specifically, we impose $-3/2 < \Delta_1 < 0 < \Delta_2 < 2$ to not only span the range of $\varphi_0 = -b \in [-1, 0]$, but also for the TELW estimator to remain applicable for the observable variables in equation (1), which, equivalently, may be analyzed as special cases of equation (19), where the variables are void of first-stage estimation errors and have $0 \le d_i < 2$, $i = 1, \ldots, k+1$.

As for the LCM inference and testing procedure, we need to impose restrictions on ℓ_b and m_b.

Assumption T-B. *Let $m_b \asymp n^{\kappa_b}$ and $\ell_b \asymp n^{\nu_b}$ with $0 < \nu_b < \kappa_b \le 1$ and define the constant $0 < \Delta < 1/2$, then the following bandwidth and trimming restrictions holds, as $n \to \infty$,*

(a) $\quad \dfrac{n^{2b}}{\ell_b^{1+2b}m} + \dfrac{m_b^{2b}}{\ell_b^{2(1+\Delta)}m} + \left(\dfrac{n}{m_b}\right)^{1-2(\underline{d}-b)} \ell_b^{-2}\left(\dfrac{m_b}{\ell_b} + \dfrac{m_b}{\ell_b^{2(\underline{d}-b)+2\Delta}}\right) \;\to\; 0.$

(b) *in addition to (a),* $\quad \dfrac{m_b^{1+2\varpi}}{n^{2\varpi}} + \dfrac{n^{2b}\sqrt{m_b}}{\ell_b^{1+2b}m} + \left(\dfrac{n}{m_b}\right)^{1-2(\underline{d}-b)} \dfrac{\sqrt{m_b}}{\ell_b^2} \;\to\; 0.$

The intuition behind the trimming and bandwidth restrictions are similar to MBLS; namely, reflecting a need for trimming to eliminate first-stage estimation and filtering errors. However, the requirements may be slightly stronger, depending on b. To illustrate this, consider the second condition in Assumption T-B(b), which attains its highest lower bound on the trimming rate, when $\underline{d} = b = 1$, $2 - \kappa/2 - (\kappa - \kappa_b)/2 < 3\nu_b$. As a result, if we set $\kappa_b = \kappa = 7/10$, this implies $0.55 < \nu_b$. The potentially stronger trimming conditions are not surprising given the work by Perron & Qu (2010), Qu (2011) and McCloskey & Perron (2013) on estimating the fractional integration order in the presence of structural breaks and other low-frequency contaminants, where corresponding trimming rates for the log-periodogram and LW estimators are equivalently strong, when the true memory parameter is close to zero. In our case, the mean-slippage behaves as low-frequency contamination, generating the need for stronger trimming, when b is close to one, as we have to reverse an (over-)differenced spectral density in the vicinity of the origin. However, in contrast to their estimation procedures for stationary variables, the TELW estimator has its own unique trimming conditions to account for the problem at hand as well as to apply for over-differenced, stationary and non-stationary variables.

Theorem 3. *Let the conditions of Theorem 1 hold, and additionally,*

(a) *Assumption T-B(a) holds, then* $\widehat{b} \xrightarrow{\mathbb{P}} b$.

(b) *Assumption T-B(b) holds, then* $\sqrt{m_b}\,(\widehat{b} - b) \xrightarrow{\mathbb{D}} N(0, 1/4)$.

Theorem 3 demonstrates that TELW inference on b enjoy the asymptotic efficiency of the ELW estimator, despite being applied to an estimated variable. The same comments apply to the memory parameters for the observable system (1). Hence, the TELW estimator not only extends results from Nielsen & Frederiksen (2011) and Christensen & Varneskov (2017) on residual memory parameter estimation for weak cointegration and stationary settings, but also provides mean-correction when estimating the fractional integration orders for observable variables, thus providing an attractive alternative to Shimotsu (2010). Importantly, Theorem 3 allows us to draw feasible inference on, and carry out testing for, predictive ability of the regressors using Theorems 1 and 2, uniformly for scenarios with stationary and non-stationary variables, which may or may not exhibit cointegration under $\mathcal{H}_A$.

Remark 6. *Theorem 3 may readily be used to test for cointegration under $\mathcal{H}_A$ by examining the one-sided hypotheses $\mathcal{J}_0 : b = 0$ against $\mathcal{J}_A : b > 0$. Moreover, the (feasible) limit theory based on Theorems 1-3 may be applied to jointly test for predictability and cointegration, thus providing a powerful framework for examining the three competing hypotheses in Section 2.2.*

4 Theoretical Extensions and their Asymptotic Theory

Assumption D1 requires zero coherence between $\boldsymbol{u}_{t-1}$ and η_t in the vicinity of the origin. Importantly, this does allow them to correlate at medium- and short-run frequencies, i.e., as $\lambda \to c > 0$. In this section, we first relax this condition and show that LCM is robust to even stronger forms of

endogeneity, with the aid of trimming. Next, we provide similar results for scenarios where regressors have been estimated as cointegration residuals prior to the LCM analysis. In order to ease exposition in the next two subsections, we focus on the case $b = 0$, i.e., the *no fractional cointegration setting*, where we, as a result, refrain from estimating b for the feasible inference and testing procedures. However, the corresponding results for the cointegration case, $b > 0$, can readily be accommodated, with minor changes to the requisite trimming and bandwidth conditions. Finally, Section 4.3 reviews the properties of the LCM procedure and draws parallels to related methodologies.

4.1 Endogenous Regressors

The local exogeneity condition in Assumption D1 accommodates a setting reminiscent of Stambaugh (1999), as long as the correlation between the innovation to $\boldsymbol{u}_{t-1}$ and η_t is not too persistent, that is, as long as $\boldsymbol{u}_{t-1}$ 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 is, arguably, more aligned with the spirit of Stambaugh (1999), as well as the imperfect predictor definition in Pastor & Stambaugh (2009).

Suppose we observe $\boldsymbol{x}_{t-1}^c = \boldsymbol{x}_{t-1} + \boldsymbol{c}_{t-1}$ where, as before, $\boldsymbol{x}_{t-1} = \boldsymbol{\mu}_x + \boldsymbol{D}_x(L)^{-1}\boldsymbol{u}_{t-1}$, 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 $\boldsymbol{c}_{t-1}$ be a $k \times 1$ mean-zero error process with η_t co-spectrum $\boldsymbol{f}_{c\eta}(\lambda) \sim \boldsymbol{G}_{c\eta}$, as $\lambda \to 0^+$, where $\boldsymbol{G}_{c\eta}$ can be non-trivial, and the components $\boldsymbol{u}_{t-1}$ and $\boldsymbol{c}_{t-1}$ are independent:

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

In this setting, the predictor, or signal, of interest, $\boldsymbol{x}_{t-1}$, is contaminated with errors, giving rise to endogenous regressor problems, generating 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, for the vector of predictors, $\boldsymbol{u}_{t-1}^c = \boldsymbol{u}_{t-1} + \widetilde{\boldsymbol{c}}_{t-1}$ with $\widetilde{\boldsymbol{c}}_t = \boldsymbol{D}_x(L)\boldsymbol{c}_t$, clearly highlighting that the added challenges to estimation arise through $\boldsymbol{x}_{t-1}^c$. Then, by recalling that $\eta_t^{(b)} = (1-L)^b \eta_t = \eta_t$ in the absence of cointegration, we have the decomposition,

$$\widehat{\boldsymbol{F}}_{ue}^c(\ell, m) = \widehat{\boldsymbol{F}}_{uu}(\ell, m)\boldsymbol{\mathcal{B}} + \widehat{\boldsymbol{F}}_{u\eta}(1, m) + (\widehat{\boldsymbol{F}}_{u\eta}(\ell, m) - \widehat{\boldsymbol{F}}_{u\eta}(1, m)) + \widehat{\boldsymbol{F}}_{\tilde{c}e}(\ell, m). \tag{20}$$

Now, utilizing local exogeneity between $\boldsymbol{u}_{t-1}$ and η_t, we know, from Lemma A.2 and A.3 of the Appendix, that $\sqrt{m}\lambda_m^{-1}\widehat{\boldsymbol{F}}_{u\eta}(1, m)$ satisfies the central limit theory in Theorem 1, i.e., an $O_p(1)$ limit purged of any asymptotic bias, and $\sqrt{m}\lambda_m^{-1}(\widehat{\boldsymbol{F}}_{u\eta}(\ell, m) - \widehat{\boldsymbol{F}}_{u\eta}(1, m)) = o_p(1)$. However, the endogeneity-generated bias term $\sqrt{m}\lambda_m^{-1}\widehat{\boldsymbol{F}}_{\tilde{c}e}(\ell, m)$ 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 (20) 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, $\widehat{\boldsymbol{u}}_{t-1}^c = \widehat{\boldsymbol{u}}_{t-1} + \widehat{\boldsymbol{c}}_{t-1}$ with $\widehat{\boldsymbol{c}}_t = \widehat{\boldsymbol{D}}_x(L)\boldsymbol{c}_t$. Specifically, we next establish asymptotic bounds for $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}^c(\ell,m) - \widehat{\boldsymbol{F}}_{\widetilde{u}\widehat{u}}(\ell,m)$ and $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}^c(\ell,m) - \widehat{\boldsymbol{F}}_{\widetilde{u}\widehat{e}}(\ell,m)$, that is, the additional source of errors for the estimator (12), stemming from having an endogenous component embedded in the regressors of interest. Moreover, to carry out testing using the LCM approach in Theorem 2, without having estimated b, we define $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) = \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}^c(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}^c(\ell,m)$ and $\widehat{\eta}_t^c = \widehat{e}_t - \widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m)'\widehat{\boldsymbol{u}}_{t-1}^c$, and obtain equivalent bounds for the differences $\widehat{G}_{\widehat{\eta}\widehat{\eta}}^c(\ell_G,m_G) - \widehat{G}_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G)$ as well as $\widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}^c(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\widetilde{u}\widehat{u}}(\ell_G,m_G)$, which control the additional errors entering through the variance estimator (16).

Theorem 4. *Suppose the conditions of Theorems 1-2 as well as Assumption C hold with $b = \widehat{b} = 0$. Moreover, suppose $n^{1/2}/m \to 0$, $n^{1/2}/m_G \to 0$ and $\underline{d} > 0$, then, for some arbitrarily small $\epsilon > 0$,*

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

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

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

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

Theorem 4 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). Importantly, this restriction may be relaxed to only require $\min_{2,\dots,k+1} d_i > 0$, that is, we do not need y_t to be fractionally integrated for this identification, and thereby Theorem 4, to hold. Fourth, it is important to emphasize, once again, that the LCM procedure can accommodate regressors, which are $I(0)$. However, these cannot be contaminated by measurement errors of the general form $\boldsymbol{c}_{t-1}$, since, in this case, the signal of interest cannot be separated from the noise. Fifth, 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 4 hold and $\kappa/2 - \underline{d}(1-\kappa) < \nu$. The asymptotic limit results in Theorem 1 and 2 then still apply with $\boldsymbol{x}_{t-1}^c$ in lieu of $\boldsymbol{x}_{t-1}$.*

Corollary 1 demonstrates that, by 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 follow stationary autoregressions, 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 7. *Despite the latent regressor signal being perturbed by noise, $\boldsymbol{x}_t^c = \boldsymbol{x}_t + \boldsymbol{c}_t$, it follows from, e.g., Deo & Hurvich (2001), Arteche (2004) and Frederiksen, Nielsen & Nielsen (2012), that standard semi-parametric estimators of d_i, $i = 2, \ldots, k+1$, 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).*

4.2 Cointegration-Based Regressors

As noted in Section 2.2, one way to achieve balance in regression systems is to utilize economically motivated cointegrating relations, a priori, to reduce the persistence of the set of explanatory variables, which is assumed to be of full rank. However, as such relations lack identification under $\mathcal{H}_0$, we cannot study predictive testing using, e.g., fractionally cointegrated VAR model (Johansen & Nielsen 2012). Hence, we add an additional step to the LCM procedure where the cointegration vector is estimated, and the residuals formed, prior to the fractional filtering. To this end, let $\underline{\boldsymbol{x}}_t \subseteq \boldsymbol{x}_t$, $\boldsymbol{x}_t \in \mathbb{R}^{\underline{k}}$ with $\underline{k} \le k$ be a subvector of $\boldsymbol{x}_t$ with elements $\underline{\boldsymbol{x}}_t = (x_{1,t}, \ldots, x_{\underline{k},t})'$, which are defined from a linear fractional cointegration relation between external auxiliary variables as,

$$x_{i,t} = \varkappa_{i,t} - \boldsymbol{\vartheta}_i{}' \boldsymbol{\chi}_{i,t}, \qquad i = 1, \ldots, \underline{k}, \tag{21}$$

where $\boldsymbol{\chi}_{i,t} \in \mathbb{R}^{\jmath_i}$ with $\jmath_i \ge 1$, and we let $(\varkappa_{i,t}, \boldsymbol{\chi}_{i,t}')' \in I(\delta_{i+1})$ such that $d_{i+1} < \delta_{i+1}$. That is, each element of the subvector is defined as a cointegrating relation. As $\boldsymbol{\vartheta}_i$ requires estimation in most applications, we impose the following structure on the cointegrating relations:

Assumption FC-M. *For each $i = 1, \ldots \underline{k}$, we have $d_{i+1} < \delta_{i+1}$ and one of following conditions:*

(a) $\quad 0 \le d_{i+1} < \delta_{i+1} < 1/2 \quad$ *and* $\quad d_{i+1} + \delta_{i+1} < 1/2.$

(b) $\quad 0 \le d_{i+1} < 1/2 < \delta_{i+1} < 2.$

Rather than exhaust all combinations of (d_{i+1}, δ_{i+1}), Assumption FC-F restricts attention to cointegrating relations that rid $(\varkappa_{i,t}, \boldsymbol{\chi}'_{i,t})$ of substantial persistence such that the cointegration residuals are (asymptotically) stationary long memory processes. However, $\underline{\boldsymbol{x}}_t = (x_{1,t}, \ldots, x_{\underline{k},t})'$ may still require fractional filtering in the LCM procedure. Note that the *weak* cointegration case in Assumption FC-F(a) is studied by Christensen & Nielsen (2006) and Christensen & Varneskov (2017), whereas the stronger cointegration case in (b) is analyzed by Robinson & Marinucci (2001, 2003).

Similarly to the fractional filtering step in Section 3.2, we do not adopt a specific estimator of the cointegration vector, $\boldsymbol{\vartheta}_i$, but rather assume to have an estimator $\widehat{\boldsymbol{\vartheta}}_i$, $i = 1, \ldots \underline{k}$, available, which satisfies the following consistency requirement:

Assumption FC-R. *Let $m_{i,\vartheta} = m_{i,\vartheta}(n)$ be a sequence with $n^{\delta_{i+1}-d_{i+1}}/m_{i,\vartheta} \to \underline{c}_i \geq 0$ as $n \to \infty$, then we assume to have an estimator with $\widehat{\boldsymbol{\vartheta}}_i - \boldsymbol{\vartheta}_i = O_p(1/m_{i,\vartheta})$, for all $i = 1, \ldots, \underline{k}$.*

The rate of convergence restriction in Assumption FC-R is very mild, requiring that it reflects the cointegration gap. As for the Assumptions D1-D3 for LCM inference, and consistent with cointegration frameworks in Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Christensen & Varneskov (2017), we further impose and state assumptions for $\Xi_{i,t} = (x_{i,t}, \boldsymbol{\chi}'_{i,t})'$, with conditions holding for all $i = 1, \ldots, \underline{k}$, and specify an additional cointegration identification restriction:

Assumption FC1. *The vector process $\Xi_{i,t}$, $t = 1, \ldots,$ is covariance stationary with mean, or initial value, $\boldsymbol{\mu}_{\Xi_i}$ and spectral density matrix $\boldsymbol{f}_{\Xi_i \Xi_i}(\lambda) \sim \boldsymbol{\Lambda}_{\Xi_i}^{-1} \boldsymbol{G}_{\Xi_i \Xi_i} \bar{\boldsymbol{\Lambda}}_{\Xi_i}^{-1}$ as $\lambda \to 0^+$, where $\boldsymbol{G}_{\Xi_i \Xi_i}$ is a real symmetric matrix of dimension $(\jmath_i + 1) \times (\jmath_i + 1)$, whose first element on the diagonal is strictly positive and lower right $\jmath_i \times \jmath_i$ submatrix has full rank, and the $(\jmath_i + 1) \times (\jmath_i + 1)$ diagonal scaling matrix is given by $\boldsymbol{\Lambda}_{\Xi_i} = \mathrm{diag}[(1 - e^{i\lambda})^{d_{i+1}}, (1 - e^{i\lambda})^{\delta_{i+1}}, \ldots, (1 - e^{i\lambda})^{\delta_{i+1}}]$ Moreover,*

(a) *There exists a $\boldsymbol{\vartheta}_i \neq \boldsymbol{0}$ such that $x_{i,t} = \varkappa_{i,t} - \boldsymbol{\vartheta}'_i \boldsymbol{\chi}_{i,t}$.*

(b) *There exists a $\varpi_i \in (0, 2]$ such that*

$$\left| \boldsymbol{f}_{\Xi_i \Xi_i}(\lambda) - \boldsymbol{\Lambda}_{\Xi_i}^{-1} \boldsymbol{G}_{\Xi_i \Xi_i} \bar{\boldsymbol{\Lambda}}_{\Xi_i}^{-1} \right| = O(\lambda^{\varpi_i} \boldsymbol{\Upsilon}_{\Xi_i} \boldsymbol{\Upsilon}'_{\Xi_i}), \qquad \lambda \to 0^+,$$

where $\boldsymbol{\Upsilon}_{\Xi_i} = (\lambda^{-d_{i+1}}, \lambda^{-\delta_{i+1}} \ldots, \lambda^{-\delta_{i+1}})'$ is a $(\jmath_i + 1) \times 1$ vector.

(c) *Let $\boldsymbol{G}_{\Xi_i \Xi_i}(1, h)$ be the $(1, h)$th element of the $(\jmath_i + 1) \times (\jmath_i + 1)$ spectrum matrix $\boldsymbol{G}_{\Xi_i \Xi_i}$, which is assumed to have $\boldsymbol{G}_{\Xi_i \Xi_i}(1, h) = \boldsymbol{G}_{\Xi_i \Xi_i}(h, 1) = 0$ for all $h = 2, \ldots, \jmath_i + 1$.*

Assumption FC2. *The vector sequence $\boldsymbol{D}_{\Xi_i}(L)\Xi_{i,t} = \boldsymbol{v}_{i,t} \boldsymbol{1}_{\{t \geq 1\}}$ where the $(\jmath_i + 1) \times (\jmath_i + 1)$ diagonal filtering matrix is given by $\boldsymbol{D}_{\Xi_i}(L) = \mathrm{diag}[(1 - L)^{d_{i+1}}, (1 - L)^{\delta_{i+1}}, \ldots, (1 - L)^{\delta_{i+1}}]$ and $\boldsymbol{v}_{i,t}$ is a linear process $\boldsymbol{v}_{i,t} = \boldsymbol{A}_{\Xi_i}(L)\boldsymbol{\epsilon}_{i,t}$, whose components satisfy:*

(a) *$\boldsymbol{A}_{\Xi_i}(L) = \sum_{j=0}^{\infty} \boldsymbol{A}_{\Xi_i,j} L^j$, $\det(\boldsymbol{A}_{\Xi_i}(1)) \neq \boldsymbol{0}$, $\sum_{j=0}^{\infty} (\sum_{h=j}^{\infty} \|\boldsymbol{A}_{\Xi_i,h}\|^2)^{1/2} < \infty$.*

(b) *The innovations, $\boldsymbol{\epsilon}_{i,t}$, are independent and identically distributed vectors with finite moments as well as with $\mathbb{E}[\boldsymbol{\epsilon}_{i,t}] = \boldsymbol{0}$ and $\mathbb{E}[\boldsymbol{\epsilon}_{i,t}\boldsymbol{\epsilon}'_{i,t}] = \boldsymbol{\Sigma}_{\Xi_i}$ where $\mathrm{rank}(\boldsymbol{\Sigma}_{\Xi_i}) = \jmath_i + 1$.*

Assumption FC3. *For $\boldsymbol{A}_{\Xi_i}(\lambda, h)$, the h-th row of $\boldsymbol{A}_{\Xi_i}(\lambda) = \sum_{j=0}^{\infty} \boldsymbol{A}_{\Xi_i,j} e^{\mathrm{i}j\lambda}$, its partical derivative satisfies $\|\partial \boldsymbol{A}_{\Xi_i}(\lambda, h)/\partial \lambda\| = O(\lambda^{-1} \|\boldsymbol{A}_{\Xi_i}(\lambda, h)\|)$ as $\lambda \to 0^+$, for $i = 1, \ldots, \jmath_i + 1$.*

The assumptions closely resembles Assumption D1-D3. The two main differences are the important cointegration identification condition in Assumption FC1(a) and the stronger assumptions on the linear process in Assumption FC2(b). The latter is maintained for comparability with the results in Robinson & Marinucci (2003). In particular, under the stated conditions, the NBLS estimator applied to $(\varkappa_{i,t}, \boldsymbol{\chi}'_{i,t})'$ will satisfy the requirement for Assumption FC-R, see, among others, Christensen & Nielsen (2006), Nielsen & Frederiksen (2011) and Christensen & Varneskov (2017) for the weak fractional cointegration case in Assumption FC-M(a) as well as Robinson & Marinucci (2003, Theorems 3-5) for the stronger cointegration case in Assumption FC-M(b). Under slightly different assumptions, including no constant or initial value, similar results are established for (semi-)parametric GLS band-spectrum estimators in Robinson & Hualde (2003, 2010), for an ELW likelihood estimator in Shimotsu (2012) and for the fractionally cointegrated VAR model in Johansen & Nielsen (2012).

Now, in this setting, we may estimate $x_{i,t}$ as,

$$ x_{i,t}^{\vartheta} = \varkappa_{i,t} - \widehat{\boldsymbol{\vartheta}}_i{}' \boldsymbol{\chi}_{i,t} \; = \; x_{i,t} + \check{x}_{i,t}, \qquad \check{x}_{i,t} \; \equiv \; (\boldsymbol{\vartheta}_i - \widehat{\boldsymbol{\vartheta}}_i)' \boldsymbol{\chi}_{i,t}, \tag{22} $$

and let $\underline{\check{\boldsymbol{x}}}_t = (\check{x}_{1,t}, \ldots, \check{x}_{\underline{k},t})'$ and $\check{\boldsymbol{x}}_t = (\underline{\check{\boldsymbol{x}}}'_t, 0, \ldots, 0)'$. Hence, the treatment of pre-estimation errors for regressors based on cointegration residuals resembles the endogenous error case in Section 4.1, having a similar additive error decomposition. However, complexities arise due to differences between the asymptotic orders of $\boldsymbol{\chi}_{i,t} \in I(\delta_{i+1})$ and $\widehat{\boldsymbol{\vartheta}}_i - \boldsymbol{\vartheta}_i = O_p(1/m_{i,\vartheta})$. As in Section 4.1, let us define the series $\widehat{\boldsymbol{u}}_{t-1}^{\vartheta} = \widehat{\boldsymbol{u}}_{t-1} + \check{\boldsymbol{u}}_{t-1}$ with $\check{\boldsymbol{u}}_t = \widehat{\boldsymbol{D}}_x(L)\check{\boldsymbol{x}}_t$, then we seek to establish asymptotic bounds for $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}^{\vartheta}(\ell, m) - \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell, m)$ and $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}^{\vartheta}(\ell, m) - \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}(\ell, m)$ as well as $\widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}^{\vartheta}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}(\ell_G, m_G)$ and $\widehat{G}_{\widehat{\eta}\widehat{\eta}}^{\vartheta}(\ell_G, m_G) - \widehat{G}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G)$ with $\widehat{\eta}_t^{\vartheta} = \widehat{e}_t - \widehat{\boldsymbol{B}}_{\vartheta}(\ell, m)' \widehat{\boldsymbol{u}}_{t-1}^{\vartheta}$ and $\widehat{\boldsymbol{B}}_{\vartheta}(\ell, m) = \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}^{\vartheta}(\ell, m)^{-1} \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}^{\vartheta}(\ell, m)$.

Assumption FC-F. *Assumption F holds for $(y_t, (\boldsymbol{x}_t^{\vartheta})')'$, and let $\underline{b} = \min_{i=1,\ldots,\underline{k}} \delta_{i+1} - d_{i+1} > 0$.*

Theorem 5. *Suppose the conditions of Theorems 1 and 2, with $b = \widehat{b} = 0$, as well as the cointegration Assumptions FC-M, FC-R, FC-F and FC1-FC3 hold, then, for some arbitrarily small $\epsilon > 0$,*

(a) $\lambda_m^{-1}\big(\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}^{\vartheta}(\ell, m) - \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell, m)\big) \leq O_p(1/\ell^{1+\underline{b}}) + O_p(1/(m^{1-\epsilon}\ell^{1+\epsilon}))$,

(b) $\sqrt{m}\lambda_m^{-1}\big(\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}^{\vartheta}(\ell, m) - \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}(\ell, m)\big) \leq O_p(m^{1/2}/\ell^{1+\underline{b}}) + O_p(1/(m^{1/2-\epsilon}\ell^{1+\epsilon}))$,

(c) $\widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}^{\vartheta}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}(\ell_G, m_G) \leq O_p(1/\ell_G^{1+\underline{b}}) + O_p(1/(m_G^{1-\epsilon}\ell_G^{1+\epsilon}))$,

(d) $\widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}^{\vartheta}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) \leq O_p(1/\ell^{1+\underline{b}}) + O_p(1/(m^{1-\epsilon}\ell^{1+\epsilon})) + O_p(1/\ell_G^{1+\underline{b}}) + O_p(1/(m_G^{1-\epsilon}\ell_G^{1+\epsilon}))$.

Corollary 2. *Suppose that the conditions of Theorem 5 hold and $\kappa/(2(1 + \underline{b})) < \nu$. The asymptotic limit results in Theorem 1 and 2 then still apply with $\boldsymbol{x}_t^{\vartheta} = \boldsymbol{x}_{t-1} + \check{\boldsymbol{x}}_{t-1}$ in lieu of $\boldsymbol{x}_{t-1}$.*

Theorem 5 and Corollary 2 demonstrate that, by the aid of trimming, the LCM procedure can accommodate regressors that have been pre-estimated and formed as a cointegration residuals. Specifically, by imposing $\kappa/(2(1+\underline{b})) < \nu$, trimming eliminates the estimation error biases, asymptotically, regardless of the cointegration strength, as conveyed by $\underline{b}$. Moreover, it alleviates higher-order biases. Of course, these results rest on having an estimator of the fractional integration orders, which satisfies Assumption FC-F. However, the latter has already been shown for the LW estimator, which is applicable to the cointegration residuals due to Assumption FC-M, by Velasco (2003) and Nielsen & Frederiksen (2011) under conditions similar to those in Assumptions FC1-FC3. Moreover, equivalent results have been provided for the tapered and exact LW by Shimotsu (2012) and a maximum likehood estimator by Johansen & Nielsen (2012) in slightly different settings.

Finally, it is important to note that the LCM procedure applies to fractionally integrated series with and without endogenous measurement errors *and* regressors from pre-estimated cointegrating relations as long as the mild trimming conditions in Assumptions T, T-G and Corollaries 1-2 are satisfied, thus providing a very flexible tool for analyzing predictability in persistent economics systems.

4.3 Discussion of Related Literature

The main benefits of using the LCM procedure, beyond avoiding concerns about spurious inference and regression balance, as detailed previously, may be summarized as follows:

(i) it accommodates general multivariate systems;

(ii) it allows for flexible persistence, captured by different fractional integration orders;

(iii) the inference procedure is uniformly Gaussian across persistence regimes as well as across alternative hypotheses with and without cointegration;

(iv) it allows the variables to have non-trivial means or initial values;

(v) it avoids imposing predictability a priori for validity of the inference;

(vi) it accommodates endogenous regressors as well as regressors that have been obtained as residuals from a pre-estimated fractional co-integrating relation.

In the light of these observations, we compare the LCM approach to alternative procedures in the literature. Specifically, we complement the discussion in the introduction with two remarks:

Remark 8. *As described in the introduction and Section 2.2, the null hypothesis that $\boldsymbol{x}_{t-1}$ contains no predictive information for y_t rules out cointegration and, hence, no information is lost by fractional filtering. Moreover, if there is cointegration between the variables under the alternative hypothesis, this improves the rate of convergence for the LCM procedure and lower its asymptotic variance, similarly to other frequency domain least squares procedures in cointegration settings; e.g., Robinson & Hualde (2003, 2010), Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Christensen & Varneskov (2017) as well as Johansen & Nielsen (2012) for a parametric approach. Notably, in contrast to*

those procedures, LCM achieves this feature without imposing predictability a priori for identification and validity of inference. Furthermore, as these alternatives, typically, require the variables of economic systems to have identical integration orders, they are not amenable for predictability testing in our general setting, failing to achieve (almost all of) the features (i)-(iv), depending on the procedure.

Remark 9. *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 setting presents a non-nested alternative to ours where, however, the dependent variable is restricted from exhibiting long-range dependence. Whereas the IVX inference is not Gaussian uniformly in the persistence regimes, the corresponding Wald significance test is robust to the specific form of persistence, which 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.*

5 Empirical Illustration: Forecasting Equity Market Volatility

The prediction of future realized equity market volatility using financial and macroeconomic indicators in VAR systems has gained renewed attention, as illustrated by the recent studies of, e.g., Christiansen et al. (2012), Paye (2012), Bansal et al. (2014), Dew-Becker et al. (2017) and Campbell et al. (2018). 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 inferential problems. In Section 7, we revisit the evidence through our robust LCM procedure.

5.1 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 index. 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.[10] 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. To this end, let $r_{t,i}$ denote the daily log-return

---

[10]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. (2018) 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 7.2.

on the S&P 500 for trading days $i = 1, \ldots, n_t$ in months $t = 1, \ldots, n$, and then,

$$\widetilde{V}_t \;=\; \sum_{i=1}^{n_t} r_{t,i}^2, \tag{23}$$

comprises our low-frequency (LF) realized variance measure.[11] 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. Specifically, we use high-frequency (HF) 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 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 B 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.[12] The source of the different series is also provided in Appendix B.

Table 1 presents full-sample summary statistics for the four series, as well as square-root and log-transformations of the LF realized variance.[13] Beyond the usual unconditional measures, we report a set of statistics that speak to the time series properties of the variables. Specifically, we provide estimates for the degree of fractional integration, using both the local Whittle (LW) estimator, cf. Künsch (1987), and the ELW estimator from Shimotsu (2010), which is robust against non-trivial means and remains valid for stationary as well as non-stationary variables. Furthermore, 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 $I(1)$.[14]

---

[11]The label LF indicates that no intra-daily, i.e., "high-frequency" financial data, are used in its construction.

[12]Our choice of variables matches the set employed in the final version of Campbell et al. (2018).

[13]The corresponding statistics for the subsample are qualitatively similar and omitted for brevity.

[14]Our baseline implementation of the KPSS test uses the Bartlett kernel and a bandwidth $\lfloor 8(n/10)^{1/4} \rfloor$, as studied by Hobijn, Franses & Ooms (2004). In addition to the results in Table 1, we have experimented with larger bandwidths, which, as shown by, e.g., Lee & Schmidt (1996) and Marmol (1998), make the KPSS test more robust against a fractional alternatives as well as considered the Epanechnikov (1969) and the Gasser, Müller & Mammitzsch (1985) optimal fourth-order kernel, also studied by Dew-Becker (2017). The results are qualitatively similar and left out for brevity.

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 Gaussian. The most noteworthy results, however, concern 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 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 ELW estimates corroborate these findings, suggesting that the DS is fractionally integrated with $d \simeq 0.8$, while TB is slightly more persistent with $d \simeq 0.9$. 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 plotting 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, corroborating our estimates of the (relative) size of the respective fractional integration orders.

5.2 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 $d \geq 0.8$. This suggests that a linear regression forecasting model for realized variance, even if $\mathcal{H}_0$ is rejected, violate the basic conditions for regression balance, c.f. the discussion in Section 2.2. Consequently, as documented by Tsay & Chung (2000), standard least squares procedures are likely to provide spurious inference.[15] Nonetheless, at this stage, we ignore such issues, as we seek to establish a benchmark for the predictive power obtained through commonly adopted OLS procedures.

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, 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 7.2, using the LCM approach. In any case, the fact that the predictor variables all display abnormal variation during the turbulent market conditions surrounding the financial crisis suggest they may carry (important) information about the future realized variance.

---

[15]See also our companion paper Andersen & Varneskov (2019b) for a detailed analysis of the issue.

Motivated by the extant literature and evidence in Figure 2, we run predictive regressions for the realized variance assuming a first-order VAR system, where $\boldsymbol{x}_{t-1}$ 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), Dew-Becker et al. (2017) and Campbell et al. (2018). Moreover, the subsample results, utilizing a high-frequency measure of volatility, demonstrate that they are robust to the choice of realized variance proxy.

Of course, this OLS analysis does not account for the degree of persistence in the realized variance and the regressors nor the indication of an unbalanced regression. Consequently, the inference may well be distorted. We next explore whether spurious inference is relevant for the present setting via simulations, based on reasonable parameter values and time series properties for the system conveyed by Table 1, and contrast them to the size and power properties of the LCM procedure.

6 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 bias, RMSE, size and power results for LCM in a bivariate setup and then proceed to study the properties of the LCM and OLS procedures in a more general predictive setting. Motivated by our empirical application, we focus on the size and power properties of tests for predictability without cointegration ($b = 0$) under $\mathcal{H}_A$, deferring the case with cointegration to the companion paper, Andersen & Varneskov (2019b). Importantly, albeit not surprisingly, the power of the LCM test further improves in the cointegration case, compared to the results below, because $\widehat{\boldsymbol{B}}(\ell, m)$ then enjoys a faster rate of convergence and lower asymptotic variance, as demonstrated by Theorem 1. Finally, we explore the robustness properties of LCM to the first-stage filtering.

6.1 Finite Sample Performance of the LCM Test

The properties of the LCM approach 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-1}, $t = 1, \ldots, n$, as,

$$(1-L)^{d_y}(1-\phi_y L)y_t = \rho u_{t-1} + \sqrt{1-\rho^2}v_t, \qquad (1-L)^{d_x}(1-\phi_x L)x_{t-1} = u_{t-1}, \qquad (24)$$

respectively, where u_t and v_t are i.i.d. standard Gaussian random variables.[16] This setup encompasses several distinct inference scenarios, depending on the parameter vector, $(\phi_y, \phi_x, \rho, d_y, d_x)$. First, we fix $\phi_y = \phi_x = 0.2$ and let $\rho = 0$ or $\rho = 0.2$, when examining the size and power properties, respectively. Second, we vary the integration orders d_y and d_x to generate alternative 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 series in Table 1, and they capture four of the inference cases examined theoretically for the OLS estimator by Tsay & Chung (2000).[17] Finally, we consider two different sample sizes $n = \{300, 650\}$, mirroring the size of the subsample ($n = 302$) and full sample ($n = 662$) in our empirical analysis, respectively.

Implementing the LCM test in Theorem 2 requires the choice of 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 that we may apply results from Hualde & Robinson (2011) and Nielsen (2015) to verify that Assumption F holds with $\varrho = 1$.[18] 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 $\kappa_G = 0.9$. The bandwidth rate κ is picked close to its upper bound to boost the efficiency of the inference, while satisfying the condition $n^{1/2}/m \to 0$ of Theorem 4, whereas the selection of the trimming rate ν is guided by 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 $\min_{i=2,\ldots,k+1} d_i \simeq 4/5$, 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, given the evidence from our empirical application.[19] 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.

Tables 3 and 4 summarize the simulation results. In particular, Table 3 demonstrates that the LCM coefficient estimates are accurate, having small biases and RMSEs across DGPs, and that its performance improves with sample size. Table 4 documents that the LCM test has excellent finite

---

[16]Note that this setting corresponds to having a latent regression relation (3) with $\boldsymbol{\mathcal{B}} = \rho$.

[17]As also detailed in Andersen & Varneskov (2019b), the least squares t-statistic, t_β, and R^2 have the following properties under the null hypothesis of no predictability: **DGP 1** has $t_\beta = O_p(n^{d_y+d_x-1/2})$ and $R^2 = O_p(n^{2(d_y+d_x-1)})$; **DGP 2** has $t_\beta = O_p(n^{d_y})$ and $R^2 = O_p(n^{2d_y-1})$; **DGP 3** has $t_\beta = O_p(n^{d_x})$ and $R^2 = O_p(n^{2d_x-1})$; and **DGP 4** has $t_\beta = O_p(n^{1/2})$ and $R^2 = O_p(1)$. Hence, the t-statistics diverge in all cases, and the R^2 is either slowly converging to zero or $O_p(1)$.

[18]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 7, 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.

[19]We have 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 similar to those reported below.

sample properties. For the smaller (sub-)sample size $n = 300$, the test is only slightly oversized, with rejection rates in the range 6-10% compared to the 5% nominal level, and the power is good, especially when selecting the wider bandwidth, $\kappa = 0.799$. For the larger sample size $n = 650$, the LCM test demonstrates both great size and power. Moreover, we note that the LCM test performs well across all DGPs, rendering it 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 the largest bandwidth ($\kappa = 0.799$), and this is achieved without sacrificing the size properties. Consequently, we use $\kappa = 0.799$ throughout.

6.2 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 designed to predict a persistent variable using persistent regressors. Specifically, we adopt a scenario similar to equation (24) with $\rho = 0$ and allow for additional exogenous processes of the same form as x_{t-1}. 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 least squares 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, displayed in Table 5, contain several interesting findings. First, for **DGP M1**, all OLS-based significance tests for individual coefficients of the persistent regressors are oversized, irrespective of the number of regressors included. In fact, the size distortions only grow as sample size increases, corroborating the theoretical results of Tsay & Chung (2000) (cf., footnote 15, **DGP 2**). Hence, the size distortions are substantial, raising serious concerns regarding the applicability of OLS inference. For example, for regressions involving only one exogenous predictor with $d_x = 0.8$ (mimicking DS), the 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 improve 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 the results in Tsay & Chung (2000). 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 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 accurate for $n = 650$, and essentially perfect for

$n = 1000$, regardless of the number of predictors included.

The results for **DGP M2** are similar to those described above, with all qualitative conclusions carrying over despite minor numerical differences. Notably, the size distortions are even greater for the individual and joint significance tests in this setting, consistent with the different divergence rates of the t-statistics provided in Tsay & Chung (2000); see **DGP 2** and **4** in footnote 15.

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 displays good size and power properties, suggesting it will deliver 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 determining whether our macroeconomic and financial state variables add auxiliary predictive power beyond past volatility in forecasting future realized variance.

6.3 Robustness to Parametric First-step Filtering

To further explore the robustness properties of the LCM procedure, we replace the first-step fractional ARMA$(1, 0)$ estimator with the semiparametric ELW from Shimotsu (2010). In particular, we implement the latter with a bandwidth $\lfloor n^{\varrho} \rfloor$ with $\varrho = \{0.71, 0.75\}$, the second-stage MBLS estimator with tuning parameters $\nu = \{0.21, 0.25, 0.30\}$ and $\kappa = \{0.65, 0.70, 0.749\}$, ensuring that the condition $\kappa < \varrho$ is always satisfied, and use the remaining configurations described in Section 6.1. The results, presented in Tables C.1-C.2 of the Supplementary Appendix, are very similar to those in Tables 3-4; the LCM procedure demonstrates excellent bias, RMSE, size and power properties. The main differences between implementations are that the fractional ARMA estimator generates lower RMSE for $n = 650$, slightly worse size for $n = 300$ and better power overall than the corresponding ELW results.

The comparison between the fractional ARMA and ELW implementations, however, rests on the former having correctly identified the AR(1) structure of y_t and x_{t-1}. Hence, as an additional robustness check, we simulate the latter as in equation (24), but with AR(2) dynamics where $\phi_{l,y} = \phi_{l,x} = 0.2$ for lags $l = 1, 2$. Moreover, we "incorrectly" fit a fractional ARMA$(1, 0)$ model or, alternatively, the ELW estimator in the first step and gauge the impact of filtering misspecification. The results for the two estimators are provided in Tables C.3-C.4, respectively, C.5-C.6 of the Supplementary Appendix. Importantly, there are no significant differences between these and the corresponding for AR(1) dynamics, demonstrating robustness of the second-stage MBLS estimator, despite the first-stage memory parameter estimates being slightly upward biased. Hence, the LCM inference is robust to the choice of fractional filtering procedure. Moreover, the results show that it is favorable to select κ close to its upper bound, $\kappa < 4/5$ for parametric filtering and $\kappa < \min(\varrho, 4/5)$ for semiparametric, since the noise arising from having estimated the ex-ante unknown integration orders is (much) smaller than the efficiency gains from using a wider range of the spectrum. Correspondingly, there is a bias-variance tradeoff in selecting the trimming rate ν: if selected too low, the first-step filtering bias from mean slippage (or endogenous regressors) may impact the coefficient estimate; if selected too high, this may hurt efficiency of the inference. In general, we recommend using the persistence-dependent lower

bound on the trimming rate in Assumption T as a rule-of-thumb guide when implementing the LCM procedure, similarly to the discussion of our empirical application in Section 6.1.

7 LCM Analysis of Predictive Power for Future RV

Section 6 documents that least squares predictive inference and testing procedures may be unreliable when the variables of the system are persistent. 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.[20] 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 6.

7.1 Empirical LCM Results

From Panel A of Table 6, we see that the coefficient estimates of the state variables 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 conventional significance levels. This contrasts sharply with the OLS results, which indicate that the state variables are jointly significant at 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. Again, this is contrary to the OLS results for joint significance, which are stronger in the subsample (significant at the 99% level). Finally, in Panel C, we apply the variance-stabilizing log-transformation on the RV series and repeat the exercise in Panel A. The qualitative implications are robust to transformation, that is, we find no significant evidence that the persistent state variables carry predictive power.

The stark difference between the LCM and OLS results is readily explained by theoretical and simulation results. The theory in Tsay & Chung (2000) and finite sample evidence in Section 6, 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 will fail to control size. In fact, these procedures will, incorrectly, reject the null hypothesis of no predictability with probability approaching one, as the sample size grows. Our LCM procedure, on the other hand, is valid irrespective of the persistence displayed by the variables of the VAR system. We conclude that there is no significant statistical evidence that any of the state variables contain relevant information for forecasting the future realized variance. Finally, since our LCM significance

---

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

tests for $\mathcal{B}$ are consistent with $\mathcal{H}_0$, there is no need to investigate how the persistent predictor variables may be transformed to comply with regression balance.[21]

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 5.2, 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 state variables. This hypothesis stipulates that uncertainty in the pricing of equities may reflect future shifts in more slowly moving economic 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 important 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 future realizations 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; the evidence is consistent with the realized variance predicting future changes in all three state variables. The results apply for both samples, demonstrating robustness with respect to alternative 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 an elevation in DS and a decline for TB and PE, consistent with the visual evidence in Figure 4. Obviously, the recent financial crisis is an extraordinary, yet important, economic event which may have an outsized impact on the inference, as also suggested by visual inspection from Figure 4. For robustness, we 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.

The LCM methodology is explicitly designed to accommodate persistent variation in VAR systems, so it is well positioned to uncover low-frequency ties between the realized variance and a set of macroeconomic state variables. Although we find that the latter do not forecast the former, we cannot reject forecast power in the reverse direction. Importantly, we have uncovered evidence of a balanced predictive relation. Specifically, given the differences in integration orders from Table 1, RV should be transformed via $\mathcal{Q}(L)$, when describing its impact on the state variables. We discuss other

---

[21]These findings are consistent with, and possibly provides an explanation for, the recent results in Berger, Dew-Becker & Giglio (2019), who no longer find the default spread to carry significant information about realized volatility over the following six months in a subsample from 1983-2014, after including different financial variables such as the VIX.

interpretations of predictability in Andersen & Varneskov (2019a, 2019b). These findings provide a challenge for Campbell et al. (2018) 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 crucial for their conclusion that shocks to the market variance are integral to understanding the role of macroeconomic fluctuations in driving the cross-sectional pricing across distinct asset classes. Our results suggest it may be useful to reassess this evidence. We defer an in-depth investigation of the implications for the prediction of macroeconomic and realized variance 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 issues with balancedness of the regression relation as well as a spurious regression problem for least squares estimation 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 procedure allows us to include variables in the system that are asymptotically stationary ($0 \leq d < 1/2$), non-stationary ($d \geq 1/2$), having endogenous measurement errors and being formed from a pre-estimated cointegrating relation. Moreover, if the regressors are significant, we accommodate predictive relations that may ($b > 0$) or may not ($b = 0$) be cointegrated. 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 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 model specification, inference, and forecasting.

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	ELW	AR-ϕ	AR-R^2	KPSS	MZ
RV	0.2897	0.2882	0.4288	0.1839	0.6298*	-23.664**
Sqrt-RV	0.4458	0.4447	0.6468	0.4184	0.9976**	-19.022**
Log-RV	0.5223	0.5409	0.7034	0.4947	1.2525**	-16.269**
DS	0.8216	0.8655	0.9699	0.9373	1.1527**	-15.400**
TB	0.8993	0.8714	0.9926	0.9814	1.5911**	-8.2203*
PE	1.0610	1.1149	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 statistics. Here, standard deviation, skewness, and excess kurtosis (relative to 3) are denoted "S.D.", "Skew", and "EKur", respectively. The LW and ELW semiparametric estimators of integration order d are implemented using a bandwidth $m = n^{0.7}$, with ELW being robust against a constant, as in Shimotsu (2010). 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 Bartlett kernel function and a bandwidth $\lfloor 8(n/10)^{1/4} \rfloor$, see, e.g., Hobijn et al. (2004) for details. 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.
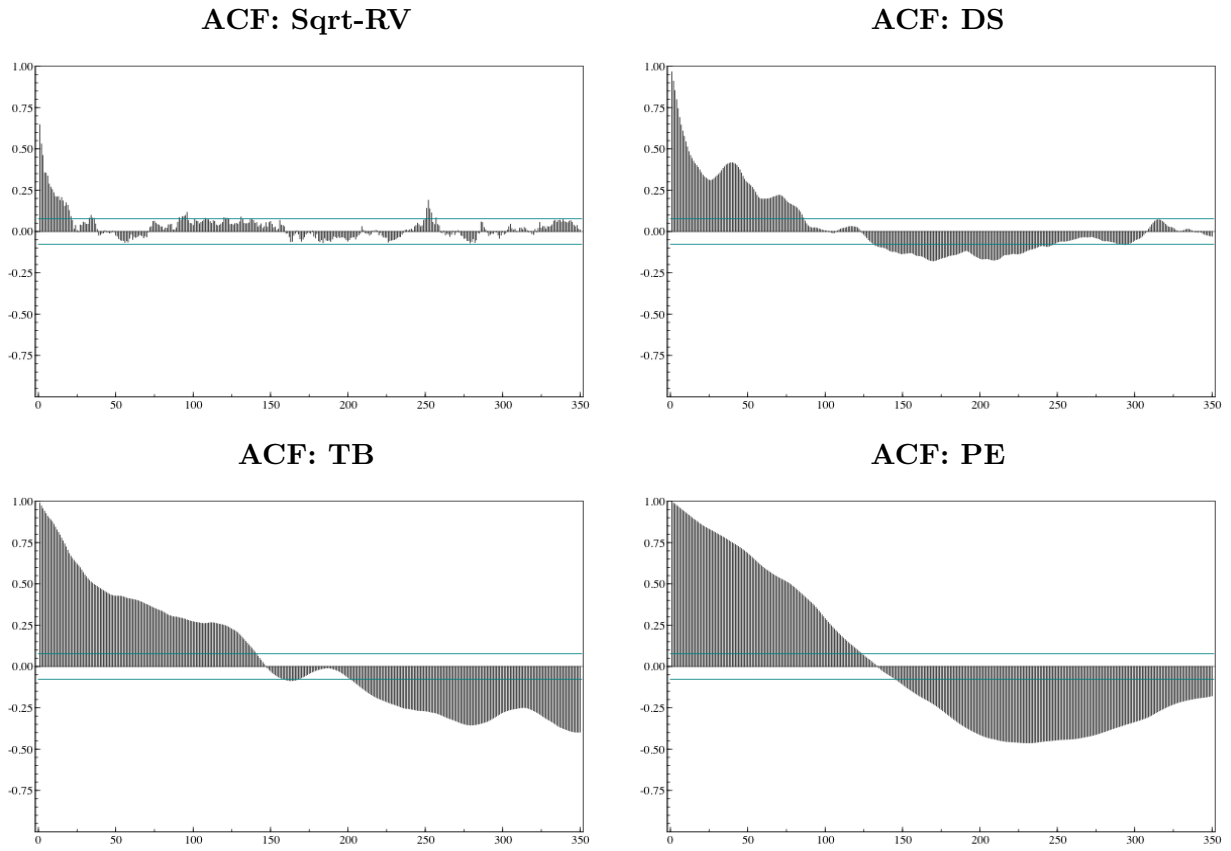
Figure 1: Autocorrelation functions. The sample autocorrelation functions are computed for the first 350 lags for each variable in the full sample, which spans the period from February 1960 through March 2015 ($n = 662$), where realized variance is estimated using daily log-returns. The variables are the square-root (sqrt) transformation of realized variance (RV), the default spread (DS), 3m T-bills (TB), and the price-earnings ratio (PE).
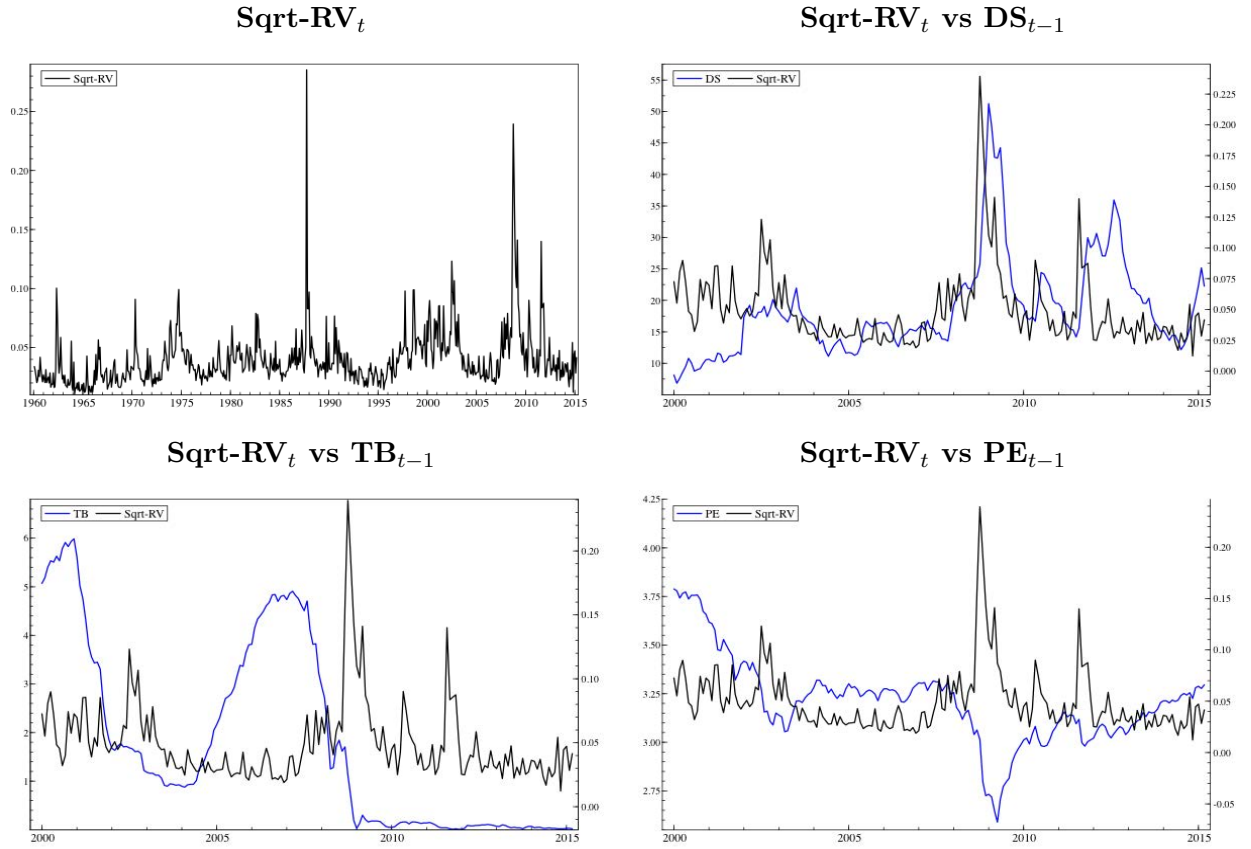
Figure 2: Plotted series. The upper left panel depicts the full sample of square-root transformed realized variance. The three remaining panels show the state variables (blue) along with the square-root RV estimates based on daily data (black) using the sample period January 2000 through March 2015 ($n = 183$). 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 (0.0004)	−0.0005 (0.0005)	−0.0012 (0.0005)	−0.0049 (0.0019)	0.0011 (0.0002)	0.0005 (0.0004)	−0.0035 (0.0013)	−0.0041 (0.0014)
RV_{t-1}	0.4288 (0.2005)	0.3919 (0.1852)	0.3918 (0.1849)	0.3838 (0.1842)	0.5939 (0.0679)	0.5653 (0.0730)	0.5628 (0.0712)	0.5558 (0.0726)
DS_{t-1}	-	0.0099 (0.0047)	0.0099 (0.0047)	0.0148 (0.0059)	-	0.0046 (0.0037)	0.0062 (0.0036)	0.0082 (0.0043)
TB_{t-1}	-	-	-	0.0179 (0.0083)	-	-	-	0.0091 (0.0079)
PE_{t-1}	-	-	0.0385 (0.0220)	0.1185 (0.0425)	-	-	0.1186 (0.0400)	0.1203 (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 (0.0516)	0.1315 (0.0516)	0.1299 (0.0516)	0.1032 (0.0516)	0.0826 (0.0680)	0.0510 (0.0680)	0.0283 (0.0680)	0.0147 (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.

Bias and RMSE of the Local Spectrum Estimator								
	DGP 1: Bias				**DGP 2: Bias**			
	$n = 300$		$n = 650$		$n = 300$		$n = 650$	
Implementation:	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$
$(\nu, \kappa) = (0.21, 0.799)$	0.0029	-0.0073	0.0020	-0.0004	0.0031	-0.0047	0.0020	0.0012
$(\nu, \kappa) = (0.25, 0.799)$	0.0025	-0.0073	0.0024	0.0001	0.0027	-0.0050	0.0023	0.0016
$(\nu, \kappa) = (0.30, 0.799)$	0.0027	-0.0071	0.0024	0.0002	0.0029	-0.0049	0.0024	0.0017
$(\nu, \kappa) = (0.25, 0.70)$	0.0028	-0.0050	0.0028	0.0012	0.0029	-0.0015	0.0027	0.0035
$(\nu, \kappa) = (0.25, 0.75)$	0.0023	-0.0067	0.0024	0.0005	0.0024	-0.0037	0.0024	0.0024
	DGP 3: Bias				**DGP 4: Bias**			
	$n = 300$		$n = 650$		$n = 300$		$n = 650$	
Implementation:	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$
$(\nu, \kappa) = (0.21, 0.799)$	0.0028	-0.0087	0.0020	-0.0010	0.0029	-0.0058	0.0020	0.0006
$(\nu, \kappa) = (0.25, 0.799)$	0.0027	-0.0086	0.0024	-0.0004	0.0029	-0.0059	0.0024	0.0011
$(\nu, \kappa) = (0.30, 0.799)$	0.0031	-0.0084	0.0024	-0.0004	0.0034	-0.0058	0.0024	0.0010
$(\nu, \kappa) = (0.25, 0.70)$	0.0030	-0.0067	0.0028	0.0002	0.0030	-0.0029	0.0028	0.0024
$(\nu, \kappa) = (0.25, 0.75)$	0.0025	-0.0081	0.0025	-0.0003	0.0026	-0.0048	0.0025	0.0016
	DGP 1: RMSE				**DGP 2: RMSE**			
	$n = 300$		$n = 650$		$n = 300$		$n = 650$	
Implementation:	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$
$(\nu, \kappa) = (0.21, 0.799)$	0.0829	0.0887	0.0553	0.0572	0.0836	0.0891	0.0556	0.0565
$(\nu, \kappa) = (0.25, 0.799)$	0.0829	0.0873	0.0552	0.0569	0.0834	0.0880	0.0555	0.0563
$(\nu, \kappa) = (0.30, 0.799)$	0.0821	0.0862	0.0556	0.0572	0.0826	0.0867	0.0559	0.0565
$(\nu, \kappa) = (0.25, 0.70)$	0.1129	0.1184	0.0788	0.0817	0.1142	0.1194	0.0795	0.0817
$(\nu, \kappa) = (0.25, 0.75)$	0.0961	0.1009	0.0649	0.0674	0.0969	0.1017	0.0652	0.0669
	DGP 3: RMSE				**DGP 4: RMSE**			
	$n = 300$		$n = 650$		$n = 300$		$n = 650$	
Implementation:	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$
$(\nu, \kappa) = (0.21, 0.799)$	0.0866	0.0951	0.0556	0.0576	0.0873	0.0946	0.0559	0.0568
$(\nu, \kappa) = (0.25, 0.799)$	0.0856	0.0914	0.0553	0.0572	0.0865	0.0917	0.0556	0.0566
$(\nu, \kappa) = (0.30, 0.799)$	0.0843	0.0901	0.0562	0.0576	0.0851	0.0899	0.0565	0.0569
$(\nu, \kappa) = (0.25, 0.70)$	0.1167	0.1248	0.0790	0.0822	0.1186	0.1258	0.0797	0.0822
$(\nu, \kappa) = (0.25, 0.75)$	0.0995	0.1060	0.0650	0.0677	0.1007	0.1064	0.0653	0.0672

Table 3: Bias and RMSE of the LCM estimator. This table displays the bias and RMSE of the local spectrum estimator, $\widehat{\mathcal{B}}(\ell, m)$, for $\rho = \{0, 0.2\}$ as a function of the trimming and bandwidth parameters, $\ell = n^\nu$ and $m = n^\kappa$, respectively. As described in Section 6, the tuning parameters are fixed according to the asymptotic theory and the DGPs are simulated as in Hong (1996) and Shao (2009). Two 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). 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. The simulations are carried out with 1,000 replications.
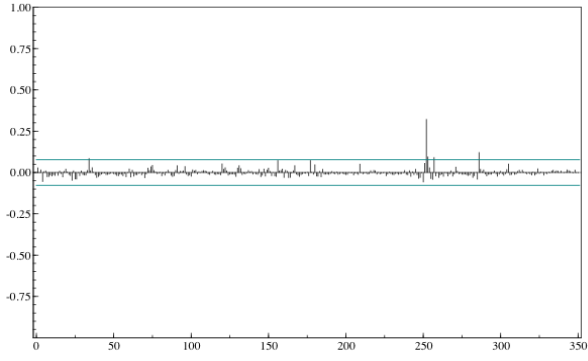
Size and Power of the Local Spectrum Test								
	DGP 1				**DGP 2**			
	$n = 300$		$n = 650$		$n = 300$		$n = 650$	
Implementation:	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$
$(\nu, \kappa) = (0.21, 0.799)$	9.20	73.00	5.80	96.10	9.70	74.30	5.90	96.40
$(\nu, \kappa) = (0.25, 0.799)$	9.50	73.30	6.40	96.10	9.90	74.80	6.50	96.60
$(\nu, \kappa) = (0.30, 0.799)$	8.60	73.80	5.90	96.00	9.00	74.60	5.70	96.40
$(\nu, \kappa) = (0.25, 0.70)$	9.90	51.90	6.30	76.30	10.30	54.30	6.60	77.50
$(\nu, \kappa) = (0.25, 0.75)$	9.50	62.20	5.60	89.20	9.60	65.20	5.60	89.40
	DGP 3				**DGP 4**			
	$n = 300$		$n = 650$		$n = 300$		$n = 650$	
Implementation:	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0$	$\rho = 0.2$
$(\nu, \kappa) = (0.21, 0.799)$	9.50	72.10	5.60	96.00	9.80	74.00	5.60	96.20
$(\nu, \kappa) = (0.25, 0.799)$	9.50	72.60	6.00	96.00	9.90	74.70	6.00	96.40
$(\nu, \kappa) = (0.30, 0.799)$	9.00	72.70	5.50	95.90	9.10	74.20	5.60	96.30
$(\nu, \kappa) = (0.25, 0.70)$	10.00	51.20	6.00	75.60	10.50	52.90	6.30	77.00
$(\nu, \kappa) = (0.25, 0.75)$	9.60	60.60	5.40	88.40	9.80	64.10	5.30	89.00

Table 4: 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 6, 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.
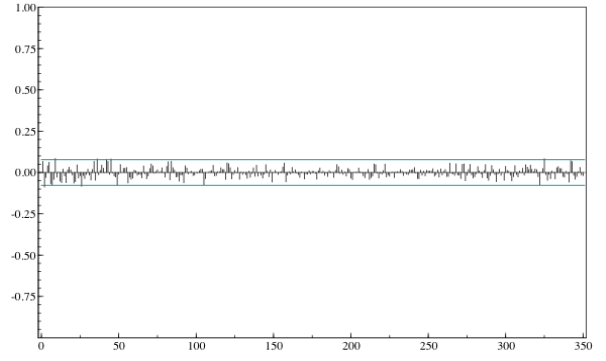
Test Size: LCM versus OLS

DGP M1

	n = 300				n = 650				n = 1000			
	100	100	100	100	100	100	100	100	100	100	100	100
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			
	100	100	100	100	100	100	100	100	100	100	100	100
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 5: 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 (24). 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.
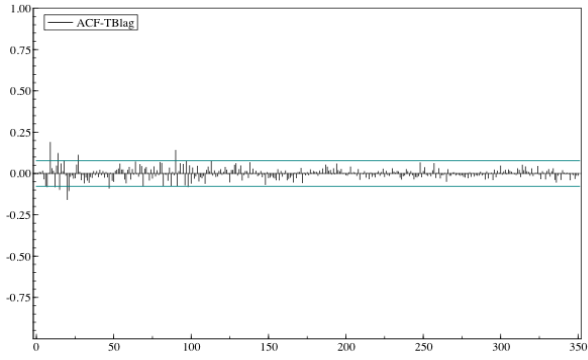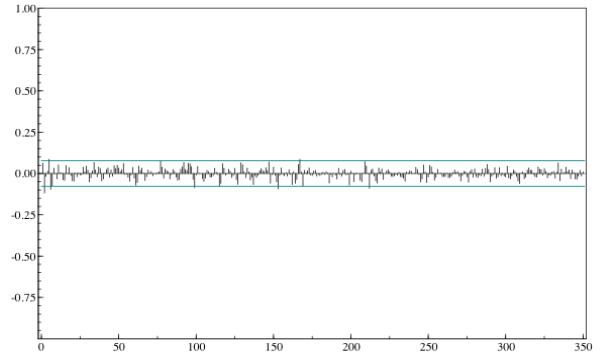
Figure 3: Autocorrelation functions. The sample autocorrelation functions are computed for the first 350 lags for the residual series of each variable after applying the ARFIMA filter to estimate the fractional integration order in the full sample, which spans the period from February 1960 through March 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.

43

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	0.0143	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	-0.0998	-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	0.2137
	$\mathcal{H}_1, \nu =$			$\mathcal{H}_2, \nu =$			$\mathcal{H}_3, \nu =$		
Panel C	0.21	0.25	0.30	0.21	0.25	0.30	0.21	0.25	0.30
DS_{t-1}	0.9223	0.6541	0.4596	0.1014	0.0814	-0.5309	0.4623	0.5918	-0.0731
TB_{t-1}	-	-	-	-	-	-	5.1313	7.3741	6.0333
PE_{t-1}	-	-	-	-62.68	-48.07	-76.14	-65.34	-52.03	-76.53
Wald	0.2370	0.1193	0.0589	3.0359	1.7895	4.2383	3.5262	2.8032	4.6146
$\mathbb{P}$-Wald	0.3736	0.2702	0.1918	0.7808	0.5913	0.8799	0.6826	0.5770	0.7977

Table 6: 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 for significance of the regressors. Specifically, Panel A reports results from the full sample where the realized variance is estimated using daily log-returns; Panel B using a subsample from February 1990 through March 2015 ($n = 302$) where high-frequency data is utilized; and Panel C considers a log-transformation of the series in Panel A. 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 6-7 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.
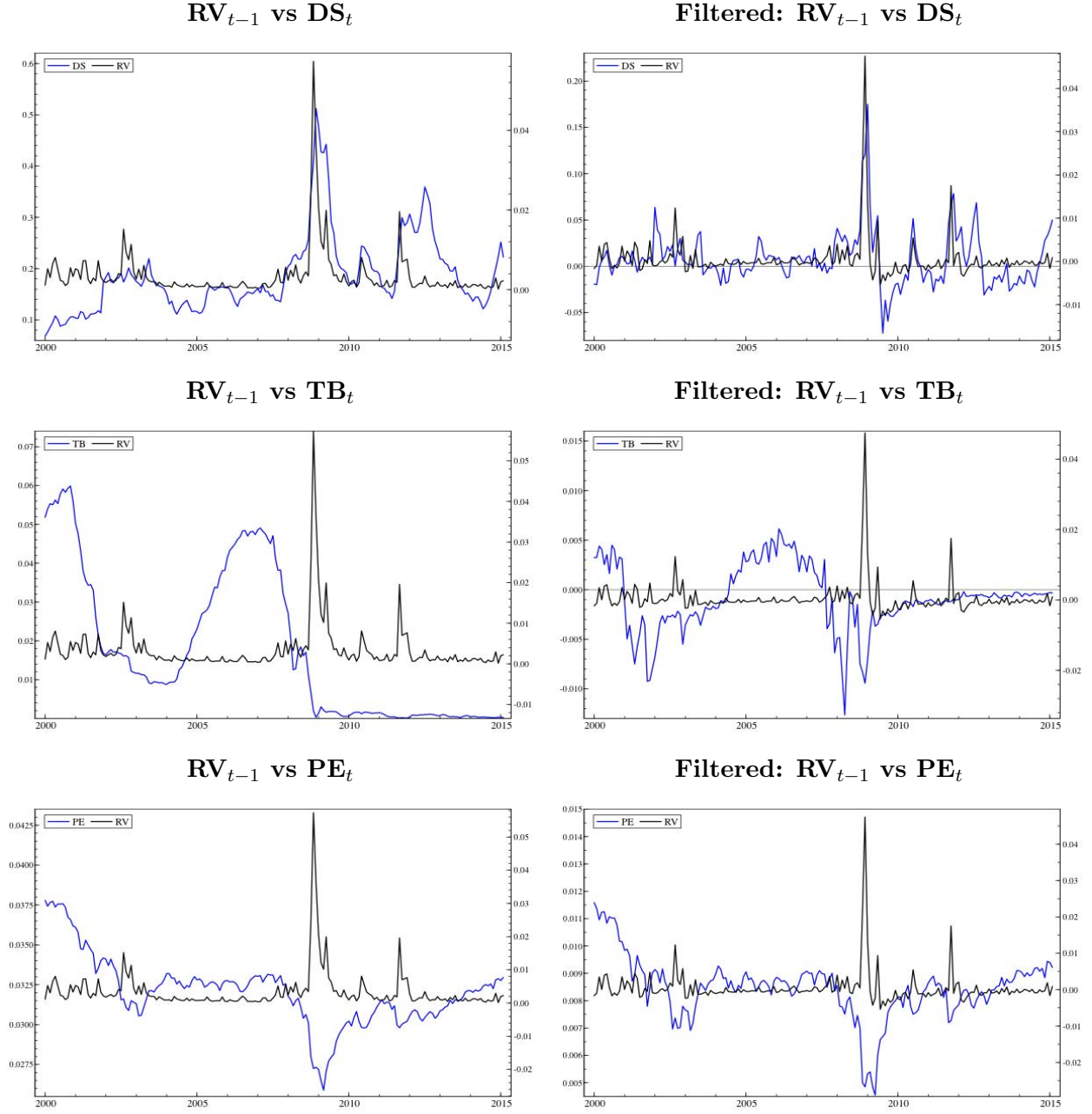
RV$_{t-1}$ vs DS$_t$ **Filtered: RV$_{t-1}$ vs DS$_t$**

RV$_{t-1}$ vs TB$_t$ **Filtered: RV$_{t-1}$ vs TB$_t$**

RV$_{t-1}$ vs PE$_t$ **Filtered: RV$_{t-1}$ vs PE$_t$**

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.

45

	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 6-7 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.

Andersen, T. G. & Varneskov, R. T. (2019*a*), Consistent local spectrum (LCM) inference for predictive return regressions. Unpublished manuscript, Copenhagen Business School and Northwestern University.

Andersen, T. G. & Varneskov, R. T. (2019*b*), Reflections on predictive regressions in persistent economic systems. Mimeo, Copenhagen Business School and Northwestern University.

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.

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.

Berger, D., Dew-Becker, I. & Giglio, S. (2019), 'Uncertainty shocks as second-moment news shocks', *Review of Economic Studies* **forthcoming**.

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. (2018), 'An intertemporal CAPM with stochastic volatility', *Journal of Financial Economics* **128**, 207–233.

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.

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.

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.

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. (2017), 'How risky is consumption in the long-run? benchmark estimates from a robust estimator', *Review of Financial Studies* **30**, 631–666.

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

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

Epanechnikov, V. A. (1969), 'Non-parametric estimation of a multivariate probability density', *Theory of Prob. and its applications* **14**, 153–158.

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.

Gasser, T., Müller, H.-G. & Mammitzsch, V. (1985), 'Kernels for nonparametric curve estimation', *Journal of the Royal Statistical Society Series B* **47**, 238–252.

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.

Hobijn, 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. (1996), *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford University Press, Oxford, United Kingdom.

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.

Lee, G. & Schmidt, P. (1996), 'On the power of the KPSS tet of stationarity against fractionally-integrated alternatives', *Journal of Econometrics* **73**, 285–302.

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.

Marmol, F. (1998), Searching for fractional evidence using combined unit root tests. Working Paper Series No. 98-39, Universidad Carlos III de Madrid.

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

McCloskey, A. & Perron, P. (2013), 'Memory parameter estimation in the presence of level shifts and deterministic trends', *Econometric Theory* **29**, 1196–1237.

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. (2017), Low-frequency econometrics, *in* B. Honore & L. Samuelson, eds, 'Advances in Economics and Econometrics: Eleventh World Congress of the Econometric Society', Vol. 2, Cambridge University Press, pp. 53–94.

Müller, U. & Watson, M. (2018), 'Long-run covariability', *Econometrica* **86**, 775 – 804.

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

Ng, S. & Perron, P. (1997), 'Estimation and inference in nearly unbalanced and nearly cointegrated systems', *Journal of Econometrics* **79**, 53–81.

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.

Nielsen, M. O. & Frederiksen, P. (2011), 'Fully modified narrow-band least squares estimation of weak fractional cointegration', *The Econometrics Journal* **14**(1), 77–120.

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.

Perron, P. & Qu, Z. (2010), 'Long memory and level shifts in the volatility of stock market return indices', *Journal of Business and Economic Statistics* **28**, 275–290.

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. (1991), 'Optimal inference in cointegrated systems', *Econometrica* **59**, 283–306.

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

Phillips, P. C. B. (2015), 'Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression', *Journal of Financial Econometrics* **13**, 521–555.

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

Qu, Z. (2011), 'A test against spurious long memory', *Journal of Business and Economic Statistics* **29**, 423–438.

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. & Hualde, J. (2003), 'Cointegration in fractional systems with unknown integration orders', *Econometrica* **71**, 1727–1766.

Robinson, P. M. & Hualde, J. (2010), 'Semiparametric inference in multivariate fractionaly cointegrated systems', *Journal of Econometrics* **157**, 492–511.

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.

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. (2012), 'Exact local whittle estimation of fractional cointegrated systems', *Journal of Econometrics* **169**, 266–278.

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

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.

Varneskov, R. T. & Perron, P. (2018), 'Combining long memory and level shifts in modeling and forecasting the volatility of asset returns', *Quantitative Finance* **18**(3), 371–393.

Velasco, C. (2003), 'Gaussian semi-parametric estimation of fractional cointegration', *Journal of Time Series Analysis* **24**, 345–378.

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