NBER WORKING PAPER SERIES

CONSISTENT LOCAL SPECTRUM (LCM) INFERENCE FOR PREDICTIVE RETURN REGRESSIONS

Torben G. Andersen
Rasmus T. Varneskov

Working Paper 28569
http://www.nber.org/papers/w28569

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2021

We thank Guido Kuersteiner and participants at the 2018 conference in honor of Peter C.B. Phillips at Yale University for helpful comments and suggestions. Financial support from CREATES, Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation, is gratefully acknowledged. Varneskov further acknowledges support from the Danish Finance Institute (DFI). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Consistent Local Spectrum (LCM) Inference for Predictive Return Regressions
Torben G. Andersen and Rasmus T. Varneskov
NBER Working Paper No. 28569
March 2021
JEL No. G12,G17

ABSTRACT

This paper studies the properties of predictive regressions for asset returns in economic systems governed by persistent vector autoregressive dynamics. In particular, we allow for the state variables to be fractionally integrated, potentially of different orders, and for the returns to have a latent persistent conditional mean, whose memory is difficult to estimate consistently by standard techniques in finite samples. Moreover, the predictors may be endogenous and "imperfect". In this setting, we provide a cointegration rank test to determine the predictive model framework as well as the latent persistence of returns. This motivates a rank-augmented Local Spectrum (LCM) procedure, which is consistent and delivers asymptotic Gaussian inference. Simulations illustrate the theoretical arguments. Finally, in an empirical application concerning monthly S&P 500 return prediction, we provide evidence for a fractionally integrated conditional mean component. Moreover, using the rank-augmented LCM procedure, we document significant predictive power for key state variables such as the price-earnings ratio and the default spread.

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

Return predictability remains a hotly debated topic. In the early financial economics literature, the fact that short-horizon equity-index returns are largely unpredictable and return innovations highly volatile was seen as a manifestation of a no-arbitrage condition, consistent with no predictability and efficient markets; see, e.g., Fama (1970). This view started to change in the 1980's with the recognition that the relevant risk factors may vary over time and across the business cycle, implying that expected stock returns must exhibit time-variation to retain an equilibrium risk-reward trade-off.

Theoretically, dynamic present value models stipulate that valuation ratios, such as the price-earnings, dividend-price, or book-to-market ratios predict future equity returns; see, e.g., Lettau & Ludvigson (2010) and Campbell (2018, Chapter 5). Similarly, equilibrium asset pricing models such as the long-run risk model (Bansal & Yaron 2004), dynamic disaster model (Gabaix 2012) or regime-switching CCAPM (Lettau, Ludvigson & Wachter 2008) suggests that returns are predictable by persistent state variables, such as the mean and volatility of consumption growth as well as the time-varying disaster recovery rate; see Neuhierl & Varneskov (2020). Nonetheless, the reliability of the empirical findings and the appropriate econometric methodology remains highly contentious. For example, the large-scale empirical study of Welch & Goyal (2008) concludes that skepticism regarding genuine out-of-sample predictability is warranted. From a methodological perspective, the primary complication is that many candidate regressors display a very high degree of persistence, inducing severe finite-sample biases under standard regularity conditions. These problems are only recently being addressed in a comprehensive manner, and the research continues unabated in the search for techniques that deliver better finite-sample performance and improved robustness.

This section first highlights the pitfalls that arise when applying standard regression inference for return predictions with persistent regressors, before reviewing potential solutions that have adopted local-to-unity and related asymptotic settings. Finally, we explain how these ideas map into the long memory framework developed in this paper and clarify what our main contributions are.

1.1 Standard Regression Inference

To illustrate the key methodological points in a concise manner, we follow Phillips (2015) by initially considering the simplest form of a predictive regression, relating the future asset returns, y_t, to a single lagged predictor, x_{t-1}, through a linear regression without an intercept,

$$y_t \;=\; \mathcal{B}x_{t-1} + v_t\,, \qquad t = 1,\ldots,n, \tag{1}$$

where the innovations, v_t, follow a martingale difference sequence (mds) with respect to the filtration generated by the past observables in the system.[1] Importantly, note that the notation and model

[1] These assumptions simplifies the exposition, but nothing of essence changes, if returns are allowed to exhibit weak dependence or to have an intercept. The mds assumption for the error term is consistent with the intuition that simple profitable strategies, unrelated to systematic risk exposures, should be absent in liquid financial markets. Weakly

specifications in this section are only expository. We will formalize our setting in Section 2.

If it is sensible to invoke standard assumptions, including weak dependence and stationarity of the returns and regressor, then it is straightforward to test for return predictability via the ordinary least squares (OLS) estimator $\widehat{\mathcal{B}}_{\text{OLS}} = \sum_{t=1}^{n} y_t x_{t-1} / \sum_{t=1}^{n} x_{t-1}^2$. The null hypothesis of no predictability implies that $\mathcal{B} = 0$, and a regular t-test for significance may be constructed. However, many relevant predictors are inherently stochastic and persistent. The impact of these features is studied by Stambaugh (1986), who amends the predictive regression with an AR(1) representation for the regressor dynamics, so that the inference problem is embedded within a closed system. In Stambaugh (1999), this approach is utilized to analyze predictive return regressions. Specifically, ignoring the intercept, the regressor obeys,

$$x_t \;=\; \phi_n x_{t-1} \,+\, w_t\,, \qquad t = 1, \ldots, n, \tag{2}$$

for a fixed initial value x_0, where $(\upsilon_t\,, w_t)'$ is an mds with $\mathbb{E}[\upsilon_t^2] = \sigma_{\upsilon\upsilon}^2$, $\mathbb{E}[w_t^2] = \sigma_{ww}^2$, and $\mathbb{E}[\upsilon_t w_t] = \sigma_{\upsilon w}$.

Often, x_t is assumed stationary, $\phi_n = \phi < 1$, even if the series is close to featuring a unit root.[2] Invoking results of Kendall (1954) and Marriott & Pope (1954), Stambaugh (1986) establishes the presence of a finite-sample bias, whenever the return and regressor innovations are correlated, that is, $\sigma_{\upsilon w} \neq 0$. Marriott & Pope (1954) show that this endogeneity bias asymptotically ($n \to \infty$), to first order, equals $-(\sigma_{\upsilon w}/\sigma_{ww}^2)(1 + 3\phi)/n$, if the mean of x_t is unknown a priori.[3] For common predictors like the dividend-price or the price-earnings ratio, the covariance $\sigma_{\upsilon w}$ is inevitably non-trivial due to the joint dependence of y and x on the price innovation, while, as noted previously, ϕ is often close to unity. Finally, because the return innovations typically are an order of magnitude larger than the innovations in the regressor, inflating $(\sigma_{\upsilon w}/\sigma_{ww}^2)$, the bias may be substantial. This motivates Stambaugh (1986) to implement a bias-correction, which is applied frequently in the subsequent literature.

Whether the endogeneity correction ensures satisfactory inference hinges on the quality of the asymptotic approximation to the distribution for the regression coefficient, $\widehat{\mathcal{B}}_{\text{OLS}}$. In this regard, the strong persistence of many candidate regressors points towards a potential "spurious regression" problem, although the absence of strong return correlation may alleviate this concern. Still, under the alternative hypothesis, $\mathcal{B} \neq 0$, the mean return inherits the persistence of the (true) regressor, even if it likely will be disguised by the large return innovations. The theoretical justification for predictability implies we should pay close attention to this scenario. Indeed, through extensive simulations under carefully calibrated, strictly stationary, alternatives, Ferson, Sarkissian & Simin (2003) demonstrate that a spurious regression problem is present, if the mean return is strongly persistent.[4] Moreover, by design, these simulations exclude correlations among the innovation series, so endogeneity and spurious regression features may constitute separate confounding challenges for inference in practice.

---

dependent return innovations, uncorrelated with past innovations to the regressor, may be accommodated through a one-sided long-run covariance correction term for most of the discussion below.

[2] The subscript n in the autoregressive coefficient ϕ_n is merely introduced for convenience here. It will be utilized in the exposition below, however, when we move beyond the strictly stationary setting.

[3] Alternatively, if the mean is known (zero in our setting), the bias is given by the smaller quantity, $-(\sigma_{\upsilon w}/\sigma_{ww}^2)(2\phi)/n$.

[4] They further demonstrate that the spurious regression problem is absent under the null hypothesis of no predictability.

The presence of a highly persistent mean return has implications beyond the need to adapt the finite-sample inference accordingly. On the one hand, it improves our ability to identify the true predictive relationship, as the signal-to-noise is enhanced, when we examine the "correct" regressor. On the other hand, the concern about misleading inference is exacerbated by the high correlation among many candidate regressors. If one is found significant, a number of others are also likely to display predictive ability. This implies that a significant regressor is not necessarily the "true" predictor, and the associated predictive relation should, at best, be viewed as providing an "imperfect" or noisy indicator for the conditional mean. In the parlance of Pastor & Stambaugh (2009), we have an imperfect predictor. It constitutes another feature we should seek to accommodate in the design of suitable inference techniques. An additional implication, stressed by Ferson et al. (2003), is that the existing evidence for predictability based on conventional inference procedures is subject to a substantial "data mining" problem. Because many potential regressors have been examined and there is a potentially significant inferential bias, many such predictors may appear significant – and by extrapolation, so will many other regressors with which the original predictor is correlated.

A common response to the problems noted above is to turn towards longer-horizon regressions, assuming the persistent signal would be more readily identified in that setting. However, the same issues surface in this setting, along with additional complications introduced by the use of overlapping observations. In fact, Boudoukh, Richardson & Whitelaw (2008), and more recently Kostakis, Magdalinos & Stamatogiannis (2015), find that no significant gains are obtained through this approach.

1.2 The Local-to-Unit Root Approach

The inferential problems associated with persistent regressors under the alternative, $\mathcal{B} \neq 0$, have spurred a large literature on techniques for improved asymptotic approximation schemes. A general representation enabling an analysis for autoregressive coefficients near unity takes the form,

$$\phi_n \;=\; 1 - \frac{C_\phi}{n^{\delta_\phi}}, \qquad C_\phi \geq 0, \quad 0 < \delta_\phi \leq 1. \tag{3}$$

In particular, for $C_\phi = 0$, we obtain the regular unit root model, $\phi_n = 1$, while $C_\phi > 0$ and $\delta_\phi = 1$ yields the local-to-unit-root (LUR) specification, $\phi_n = 1 - C_\phi/n$, which ensures that the asymptotic distribution captures the effect of having a root in the vicinity of unity, irrespective of sample size. The LUR representation for autoregressions is first analyzed in depth by Phillips (1986), while early developments for the predictive regression setting are provided by Cavanagh, Elliott & Stock (1995) and Valkanov (2003), with the latter focusing on applications in financial economics.

The LUR approximation to the asymptotic distribution in the near unit root scenario for the predictor has two important implications. First, the rate of convergence of $\widehat{\mathcal{B}}_{\text{OLS}}$ increases to n, reflecting the enhanced signal-to-noise ratio associated with unit root-style regressions. Second, inference generally becomes non-standard. Specifically, if $\sigma_{vw} \neq 0$, the interaction between the persistent regressor and the lagged return residual generates a random endogeneity bias that depends on C_ϕ. Under the

LUR specification, the deviation of the autoregressive root ϕ_n from unity shrinks at the same speed as the rate of convergence, rendering consistent inference for this coefficient infeasible. This implies that C_ϕ is an unidentified nuisance parameter, and the asymptotic distribution for $\mathcal{B}$ has a discontinuity around unity, relative to the stationary case ($\phi_n = \phi < 1$), complicating inference in the absence of prior knowledge about the underlying strength of the regressor persistence.

Various techniques have been developed to handle the above inference problem within the univariate regression setting. The most common procedure is the construction of Bonferroni bounds, combining the confidence intervals obtained across a range of relevant values for C_ϕ, as explored systematically by Campbell & Yogo (2006). The main shortcoming of this approach, as noted in Phillips (2014), is the lack of robustness to the stationary scenario, $\phi_n = \phi < 1$. The latter scenario will entail spurious rejections of the null hypothesis of no predictability with probability approaching one, as the sample size increases. Instead, Phillips (2014) advocates reliance on the usual (asymptotically centered) estimate for the autoregressive coefficient under stationarity in the construction of the LUR Bonferroni bounds, as Mikusheva (2007) shows this leads to uniformly valid confidence intervals for ϕ_n under a broad set of conditions. Moreover, the induced confidence bands are asymptotically valid and provide a good approximation to the ones obtained under stationary asymptotics.[5]

However, even if the the robust Bonferroni approach provides sensible inference in the case of highly persistent regressors in univariate predictive regressions, it falters for multivariate predictive regressions due to the complications associated with handling of multiple distinct localizing coefficients. Moreover, this limitation is shared by many of the other alternative inference techniques for univariate predictive regressions, as reviewed by Phillips (2015). Consequently, in the next section, we turn to an approach that has proven successful, also for cases involving multiple predictors.

1.3 The IVX Approach

A tractable approach to *multivariate* predictive return regressions with highly persistent regressors and potential endogeneity was obtained only following the developments of Magdalinos & Phillips (2009), who introduce endogenous instrumentation designed to eliminate the nonstandard asymptotics arising from the choice of $\delta_\phi = 1$ for the autoregressive coefficient in the regressor dynamics. This is achieved by ensuring the instrument induces less persistence than the LUR and unit root scenarios, yet retains a sufficiently high degree of time series dependence to annihilate the potentially severe finite-sample endogeneity bias and to secure a relatively fast convergence rate, as explained below.

1.3.1 Univariate IVX Estimation

We continue to illustrate the main points within the univariate setting for brevity, noting, however, that all aspects of the discussion may be extended to multivariate systems. The key deviation in this section is that prior knowledge about the nature of the persistence of the regressor is not assumed,

[5] For another procedure to obtain near optimal tests in the univariate setting, see Elliott, Müller & Watson (2015).

as the IVX framework allows the regressors to contain a unit root, a LUR representation, moderate integration ($C_\phi > 0$, $0 < \delta_\phi < 1$), and stationarity ($C_\phi > 0$, $\delta_\phi = 0$). Specifically, in this setting, the IVX procedure obtains valid inference by generating an instrument for $x_t = \sum_{s=1}^{t} \Delta x_s$ directly from the series itself through a filter that ensures a mild reduction in the degree of persistence,

$$\widetilde{z}_t \; = \; \sum_{s=1}^{t} \phi_{nz}^{t-s} \, \Delta x_s, \qquad \phi_{nz} \; = \; 1 - \frac{C_z}{n^{\beta_z}}, \quad 0 < \beta_z < 1, \quad C_z > 0. \tag{4}$$

When β_z is chosen below, but near, unity, $\widetilde{z}_t$ is at most mildly integrated, and its dynamics is governed exclusively through deliberate choices of C_z and β_z, which may, thus, be designed to generate a desirable limit distribution.[6] The IVX estimator is, then, simply the standard IV estimator, with $\widetilde{z}_t$ serving as instrument, $\widehat{\mathcal{B}}_{\text{IVX}} = \sum_{t=1}^{n} y_t \widetilde{z}_{t-1} / \sum_{t=1}^{n} x_{t-1} \widetilde{z}_{t-1}$. In the unit root and LUR scenarios, the estimation error for OLS, $\sum_{t=1}^{n} \upsilon_t \, x_{t-1} / \sum_{t=1}^{n} x_{t-1}^2$ will have asymptotically dependent numerator and denominator, generating a non-standard limiting distribution. In contrast, the lower degree of dependence associated with the moderately integrated IVX instrument is sufficient to ensure asymptotic independence and a tractable limit distribution, as shown in Phillips & Magdalinos (2007). Specifically, letting the errors obey a mds, then, under suitable regularity conditions, $n^{(1+\beta_z)/2}(\widehat{\mathcal{B}}_{\text{IVX}} - \mathcal{B}) \xrightarrow{\mathbb{D}} MN(0, \sigma_{\text{IVX}}^2)$. The asymptotic variance, σ_{IVX}^2, is generally stochastic, if the IVX instrument is moderately integrated, but a feasible, consistent estimator may readily be constructed using the standard linear regression approach, as detailed in Phillips (2015), and a standard t-test may be constructed. Consequently, the IVX instrumentation restores standard inference for return regressions, in cases where the predictor possesses an unknown degree of integration and may be an $I(1)$ or LUR process.

The main cost of the IVX approach is the lower rate of convergence, $n^{(1+\beta_z)/2}$, compared to n for the $I(1)$ or LUR scenarios. This suggests picking a value for β_z near unity, while still ensuring a finite sample performance, that avoids mimicking the nonstandard unit root asymptotics. The extensive simulation evidence in Kostakis et al. (2015) demonstrates that picking $\beta_z = 0.95$ is sufficient to ensure reliable inference and induce good power properties in many typical settings.

1.3.2 Multivariate IVX Estimators

As noted previously, the IVX methodology can be generalized to return regressions with multiple predictors. However, this does require the imposition of additional assumptions. For example, Kostakis et al. (2015) provide theory for the multivariate regressor case, but impose that the unknown localizing coefficient is identical for all regressors. That is, they can display memory characteristics ranging from strictly stationary to non-stationary unit root processes, but they all possess the identical degree of persistence. Given the range of predictors used in empirical work, including near-unit root valuation ratios, macroeconomic variables, lagged returns, and realized volatility measures, it is a very strong

[6] To see this, note that $\widetilde{z}_t = z_t - (C_\phi / n^{\delta_\phi}) \sum_{s=1}^{t} \phi_{nz}^{t-s} x_{s-1}$, where $z_t = \phi_{nz} z_{t-1} + w_t$, implying $\widetilde{z}_t$ equals z_t, except for a term that is asymptotically negligible. The notion of moderate deviation from unity was introduced by Phillips & Magdalinos (2007) to capture slightly wider deviations from a unit root than accomplished through LUR specifications.

requirement. Phillips & Lee (2016) show that results can be obtained for mixed localization coefficients on the regressors, including the presence of both moderately integrated and moderately explosive regressors, but their general setting does require imposition of various bounds on the size of the IVX parameter β_z relative to the set of (unknown) localizing coefficients for the regressors, which does not include the strictly stationary case. Likewise, non-trivial conditions must be imposed on the specification of the linear set of restrictions imposed on the autoregressive coefficient matrix for the usual multivariate Wald test. Although their findings, combined with the Monte Carlo results in Kostakis et al. (2015), suggest that the IVX ultimately can deal with multiple regressors possessing mixed and wide ranging degrees of persistence and long run properties, a fully unified theory is still not established, as explicitly discussed in the concluding section of Phillips & Lee (2016).

Besides these caveats, Xu (2020) points to the issue of potential cointegration among the multiple regressors employed within a predictive return regression. This can easily arise, especially if more than one of the typical valuation ratios are used, as they all represent scaled versions of the stock price level.[7] Xu (2020) proceeds to show that the Kostakis et al. (2015) approach can be robust to an unknown degree of cointegration among the regressors, but it requires a strong assumption, namely that the regressors are "perfect" in the sense of Pastor & Stambaugh (2009).

1.3.3 Extensions and Related Inference Principles

The IVX principle induces tractable inference procedures within highly persistent regression systems through the use of instruments that proxy the original predictors, but are engineered to display a lower degree of persistence. This bears a resemblance to prior insights, noting that asymptotic normal inference will obtain for parameters expressed as coefficients on stationary regressors, even within $I(1)$ systems, see, e.g., Park & Phillips (1989) and Sims, Stock & Watson (1990). The same line of reasoning inspired the idea of adding lagged regressors and/or regressands to linear regression systems in settings, where there is uncertainty about the orders of integration among the variables. For example, if a specific regressor is assumed to have a root close to unity, one may include an additional lag of this persistent regressor or, alternatively, its first difference, as an additional regressor.[8]

The idea of variable addition has been adopted for predictive regressions with unknown degrees of persistence for either the regressand, the regressors or both. Breitung & Demetrescu (2015) compare the size and power properties of IVX and related variable addition techniques in a LUR setting; Ren, Tu & Yi (2019) adopt a similar setting with potentially strongly dependent regressors and add an extra lag of all regressors to obtain the slower, standard rate of convergence, $\sqrt{n}$, along with χ^2-distributed Wald tests. Likewise, Liu, Yang, Cai & Peng (2019) consider univariate predictive regressions, where

[7] In fact, Lettau & Ludvigson (2001) directly employ a theoretically motivated cointegrating relation to generate a predictive regressor, the so-called cay variable, involving aggregate consumption, income and wealth.

[8] The point is illustrated in Hamilton (1994, Chapter 18) for scenarios subject to potential spurious regression issues in a unit root setting, while Choi (1993) explores inference in AR systems with $I(1)$ processes. These procedures are studied more broadly for inference in possibly (co-)integrated VAR systems by, e.g., Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996). Moreover, Bauer & Maynard (2012) show how an infinite order VAR system can accommodate unknown strong persistence in an additional set of forcing variables via the same type of variable augmentation.

the regressand cannot be stationary under the alternative of predictability, if the regressor is strongly dependent. They augment the regression with the first-differenced predictor and an additional lagged predictor, and then conduct inference through an empirical likelihood approach, obtaining standard χ^2 distributed test statistics. This particular method is, however, quite unwieldy in multivariate settings. Moreover, Lin & Tu (2020) study the univariate regression case, where the regressand is strongly persistent, while the (persistent) predictor is imperfect, so that the persistence spills over into the regression residuals. They propose a robust inference strategy by including both a lagged regressand and predictor as extra regressors. Not surprisingly, this generates the usual rate of $\sqrt{n}$ convergence for the slope coefficient, allowing for regular inference procedures. Their results also hold if the system displays ("perfect", in the sense of Pastor & Stambaugh (2009)) cointegration. Finally, Georgiev, Harvey, Leybourne & Taylor (2020) develops a fixed regressor wild bootstrap test for whether the predictive regression is invalid in a setting where the regressors are persistent and, possibly, imperfect such that the persistence spills over to the residuals, leading to potential spurious inference.

1.4 Final Observations: Bridging the Gap to LCM

In summary, a variety of econometric issues continue to complicate the analysis of multivariate predictive return regressions. The predictors may possibly be "imperfect", and they may display unknown and differing degrees of persistence. The issue of imperfect predictors looms particularly large, as this feature, intuitively, provides a realistic characterization of the type of scenario encountered in practice. To alleviate this issue, it is tempting to include a large set of regressors to maximize the ability to span the most persistent conditional mean component of the regressand. However, currently, there is no uniform approach that can handle inference for the multivariate, imperfect predictor case.

In our previous work Andersen & Varneskov (2020), we develop a different asymptotic framework for analyzing predictive regressions within persistent systems. Specifically, we assume that all variables are fractionally integrated of potentially different orders, and that the regression may, or may not, feature cointegration. Let L and $(1-L)^d$ be the usual lag and fractional differencing operators, then, drawing parallels to the predictive systems (1)-(4), we stipulate a predictive relation of the form,

$$y_t = \mathcal{B}(1-L)^{d_x - d_y} x_{t-1} + \upsilon_t, \quad (1-L)^{d_x} x_{t-1} = u_{t-1}, \tag{5}$$

where $u_{t-1} \in I(0)$ is weakly dependent, and $\upsilon_t \in I(d_y - b)$ with $0 \leq b \leq d_y$ captures the possibility of cointegration (when $b > 0$). As a result, it follows that $y_t \in I(d_y)$ and $x_{t-1} \in I(d_x)$ may exhibit either weak or strong dependence by allowing their fractional integration orders to fall within a wide range $0 \leq d < 2$, for $d = \{d_y, d_x\}$. Importantly, the framework in Andersen & Varneskov (2020) is not confined to univariate predictive regressions (with trivial means or initial values), but accommodates diverse persistence (i.e., d's) among the predictors, thus providing a flexible setting to analyze systems with various financial and macroeconomic variables. This feature corresponds to having different localization coefficients in the LUR setting (3). Andersen & Varneskov (2020) propose a two-step Local speCtruM (LCM) approach that delivers asymptotically Gaussian inference, regardless of persistence of the variables and cointegration in the predictive relation, by first stripping the persistence of the variables using a consistent estimate of their integration orders and subsequently by applying a robust, medium band least squares (MBLS) estimator. However, while tackling the issue of "spurious" inference in persistent systems, they do not consider scenarios where the predictors may be "imperfect".

Hence, in this paper, we extend the framework in Andersen & Varneskov (2020) to further allow for imperfect regressors (in the spirit of Pastor & Stambaugh (2009)) that may exhibit general forms of endogeneity. That is, we tackle empirically relevant scenarios where the regressors may be imperfect, persistent and endogenous, for which, as discussed above, there is currently no uniform solution in the literature. However, as the LCM procedure critically relies on consistent estimation of the fractional integration orders of the variables, this problem turns out to be particularly difficult for return regressions, since the signal-to-noise ratio of the conditional mean return to its innovations is too "low" for standard univariate time series techniques to detect (strong) serial dependence in finite samples. We overcome this issue by proposing a new rank testing procedure, that allows us to discriminate between the "imperfect" and "perfect" regressor scenario and to determine the persistence of the conditional return mean. Our, important, identifying condition is that the integration of the returns belongs to the set of integration orders from the multiple candidate predictors. That is, the set of predictors have been chosen "sensibly". If this is the case, the procedure can verify that the conditional mean is, indeed, persistent and distinguish between inference scenarios. Once we have determined the return persistence, we may implement the two steps of the LCM procedure, without modification.

We establish the asymptotic properties of our new rank test and rank-augmented LCM procedure in an endogenous, imperfect, and persistent regressor setting, demonstrating that the asymptotic distribution theory is Gaussian, regardless of the inference scenario; stationary versus non-stationary persistence and perfect versus imperfect predictors. Moreover, we examine the finite sample properties of predictability tests using OLS, IVX and LCM procedures. Specifically, we find that OLS and IVX may suffer from considerable size distortions in our long memory setting, thus providing "spurious" inference. Importantly, we also show that our rank selection procedure has considerable finite sample power to detect a persistent conditional mean return, and that our rank-augmented LCM procedure is (almost) as efficient as if we had known the true persistence of the system ex-ante, i.e., as an oracle implementation of LCM. Finally, in an empirical application to monthly S&P 500 return prediction, we find corroborating evidence that returns contain a fractionally integrated conditional mean component. In addition, by applying the rank-augmented LCM procedure, we find key state variables, such as the price-earnings ratio and the default spread, to possess significant predictive power for future returns.

The paper proceeds as follows. Section 2 introduces the setting, draws parallels to the imperfect regressor model of Pastor & Stambaugh (2009) and describes the LCM procedure. Section 3 provides our new rank test and rank-augmented LCM procedure as well as examines their asymptotic properties. Section 4 contains the simulation study, and Section 5 provides the empirical analysis of return predictions. Finally, Section 6 concludes. The Appendix contains additional theory and proofs.

2 Predictive Returns Regressions with Persistent Variables

This section introduces a predictive regression framework for asset returns, where all the variables may exhibit fractional integration of potentially different orders. The framework is inspired by the persistent economic systems studied by Andersen & Varneskov (2020) as well as the predictive system for expected returns with imperfect predictors developed by Pastor & Stambaugh (2009). Finally, we motivate and review the Local speCtruM (LCM) approach, introduced in the former.

2.1 Predictive System and Assumptions

Suppose we observe a $(k+1)\times 1$ vector $\boldsymbol{Z}_t = (y_t, \boldsymbol{\mathcal{X}}_{t-1}')'$ at times $t = 1,\dots,n$, where y_t contains the asset returns and $\boldsymbol{\mathcal{X}}_{t-1}$ is a vector of candidate predictors, which has a multi-component structure,

$$\boldsymbol{\mathcal{X}}_{t-1} = \boldsymbol{x}_{t-1} + \boldsymbol{c}_{t-1}, \qquad \boldsymbol{x}_t \perp\!\!\!\perp \boldsymbol{c}_s, \quad \text{for all } t,s, \tag{6}$$

with $\boldsymbol{x}_{t-1}$ capturing the most persistent signal, and $\boldsymbol{c}_{t-1} \in I(0)$ being mean-zero and collecting either measurement errors, additional weakly dependent components embedded in the variables, or both. Moreover, let us define $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_{t-1}')'$, which is assumed to obey a Type II fractional model,

$$\boldsymbol{D}(L)(\boldsymbol{z}_t - \boldsymbol{\mu}) \;=\; \boldsymbol{v}_t\,\mathbf{1}_{\{t\geq 1\}}, \tag{7}$$

where $\boldsymbol{\mu}$ is a $(k+1)\times 1$ vector of nonrandom, unknown finite numbers, capturing either the means or initial values of $\boldsymbol{z}_t$, the vector process $\boldsymbol{v}_t = (e_t, \boldsymbol{u}_{t-1}')'$ is weakly dependent, and,

$$\boldsymbol{D}(L) = \text{diag}\left[(1-L)^{d_1},\dots,(1-L)^{d_{k+1}}\right], \quad \text{with} \quad (1-L)^d \;=\; \sum_{i=0}^{\infty} \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}\,L^i, \tag{8}$$

where $\Gamma(\,\cdot\,)$ is the gamma function.[9] In this setting, in which all variables may exhibit high degrees of persistence, the predictive relation between y_t and the observable regressors $\boldsymbol{\mathcal{X}}_{t-1}$ will be defined through the weakly dependent components of the persistent signals. Specifically, we assume,

$$e_t = \varphi_{t-1} + \eta_t^{(b)}, \quad \varphi_{t-1} = \boldsymbol{\mathcal{B}}'\boldsymbol{u}_{t-1} + \xi_{t-1}, \quad \boldsymbol{u}_t \perp\!\!\!\perp \xi_s, \quad \text{for all } t,s, \tag{9}$$

where $\eta_t^{(b)} = (1-L)^b \eta_t$ for some constant $b \geq 0$ and $\eta_t \in I(0)$, and with $\xi_{t-1} \in I(0)$. Importantly, however, by combining the relations (7) and (9), this is tantamount to a balanced predictive model for asset returns,

$$y_t = a + \boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{Q}}(L)\boldsymbol{x}_{t-1} + \xi_{t-1}^{(-d_1)} + \upsilon_t, \quad t = 1,\dots,n. \tag{10}$$

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 $\boldsymbol{\mu} = (\mu_y, \boldsymbol{\mu}_x')'$ as well as $\upsilon_t = (1-L)^{b-d_1}\eta_t$ and $\xi_{t-1}^{(-d_1)} = (1-L)^{-d_1}\xi_{t-1}$.

[9] Formal assumptions on the components of the system are stated below.

The predictive system (6)-(10) encompasses most multivariate fractionally integrated systems in the literature, in addition to features specific to the problem of predicting asset returns. To see this, suppose $\boldsymbol{c}_{t-1} = \mathbf{0}$ and $\xi_{t-1} = 0$, $\forall t$, as well as $0 \leq b \leq d_1$, then the most persistent components of the explanatory variables are directly observable, the predictive relation is well-defined and balanced, and the system may ($b > 0$) or may not ($b = 0$) feature (fractional) cointegration. By relaxing these restrictions, however, the system more accurately describe the inferential issues surrounding return regressions. In particular, $\boldsymbol{c}_{t-1}$ is included to accommodate endogeneity, multiple components and measurement errors in the regressors, rendering their signals latent, φ_{t-1} captures the possibility that the predictors may imperfectly describe the conditional mean, and, by letting $b = d_1$, the return regression have a weakly dependent innovation that may dominate the persistent signal in finite samples. We will detail these points, provide examples and draw parallels to the extant literature, particularly to Pastor & Stambaugh (2009) and Andersen & Varneskov (2020), in the next section.

Before proceeding, we impose some formal structure on the system. The conditions mirror those imposed by Andersen & Varneskov (2020) and the assumptions for the semiparametric fractional cointegration analyses in, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Christensen & Varneskov (2017), but with subtle differences due to the differing model features. To this end, let "$\sim$" signify that the ratio of the left- and right-hand-side tends to one in the limit, element-wise. We then impose assumptions in terms of $\boldsymbol{q}_t = (\boldsymbol{u}'_{t-1}, \eta_t)'$ and $\boldsymbol{\zeta}_{t-1} = (\boldsymbol{c}'_{t-1}, \xi_{t-1})'$ rather than $\boldsymbol{v}_t$, when exploring the asymptotic properties for the LCM procedure below.

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

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

Assumption D3. *For $\boldsymbol{A}(\lambda, i)$, the 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 $i = 1, \ldots, k+1$.*

Assumption C. *Suppose $\boldsymbol{\zeta}_{t-1} = \boldsymbol{\zeta}_{t-1}\mathbf{1}_{\{t \geq 1\}}$ is a mean-zero $(k+1) \times 1$ vector satisfying the same Assumption D1-D3 as $\boldsymbol{u}_{t-1}$, except that it has η_t co-spectrum $\boldsymbol{f}_{\zeta\eta}(\lambda) \sim \boldsymbol{G}_{\zeta\eta}$, as $\lambda \to 0^+$, where the constant vector $\boldsymbol{G}_{\zeta\eta}$ may have non-zero entries. Moreover, let $\boldsymbol{u}_t \perp\!\!\!\perp \boldsymbol{\zeta}_s$ for all $t, s \geq 1$. Finally, if the i-th element of the vector $\boldsymbol{\zeta}_{t-1}$ is trivial, that is, if $\boldsymbol{\zeta}_{t-1}(i) = 0$ for all $t \geq 1$, then the co-spectrum condition $\boldsymbol{G}_{\zeta\eta}(i) = \boldsymbol{G}_{\zeta\zeta}(i, g) = \boldsymbol{G}_{\zeta\zeta}(g, i) = 0$ for $g = 1, \ldots, k+1$, is naturally also required.*

Assumption M. *Let* $0 \leq d_1 \leq 1$ *and* $0 \leq d_i \leq 2$ *for all* $i = 2, \ldots, k+1$. *Define* $\underline{d}_x = \min(d_i;\, 2 \leq i \leq k+1)$, $\underline{d} = \min(d_1, \underline{d}_x)$, $\bar{d}_x = \max(d_i;\, 2 \leq i \leq k+1)$, *and let* $\underline{d}_x > 0$ *and* $b = d_1$.

Assumptions D1-D3 are standard in the literature studying fractional (co-)integration. Specifically, D1 and D3 impose a rate of convergence for the spectral density $\boldsymbol{f}_{qq}(\lambda)$ as $\lambda \to 0^+$, which depends on the smoothness parameter $\varpi \in (0, 2]$. In addition, 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, condition D2 specifies linearity, martingale and moment conditions for $\boldsymbol{q}_t$, allowing for general multivariate dependence among the variables, thus accommodating flexible lead-lag and predictive structures.

Whereas D1 allows the latent predictive signals, $\boldsymbol{x}_{t-1}$, to exhibit mild endogeneity (as $\lambda \to c > 0$) through $\boldsymbol{u}_{t-1}$, Assumption C lets the observable explanatory variables exhibit stronger forms of endogeneity, that is, to display non-trivial correlations with the innovations to asset returns. These correlations are captured via the co-spectrum between the less persistent component (and/or measurement errors) $\boldsymbol{c}_{t-1}$ and the innovations η_t, which, furthermore, may both be non-trivially correlated with the "conditional mean errors" from the, possibly, imperfect predictors, ξ_{t-1}. This treatment of endogenous predictors is similar in spirit to Stambaugh (1999) and Pastor & Stambaugh (2009).

Assumption M imposes a mild structure on the memory of the system. Specifically, we restrict the persistent component of returns to maximally exhibit unit root persistence, whereas the explanatory variables can be explosive, $d_i > 1$. In general, however, the assumptions accommodate flexible persistence among the variables; if $0 < d_i < 1/2$, the variable is (asymptotically) stationary with long memory; if $d_i \geq 1/2$, the variable is non-stationary, but has a well-defined mean for $d_i < 1$. This flexibility is particularly useful for characterizing the properties of multivariate predictive 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.

Finally, we impose $b = d_1$, which implies $\upsilon_t = \eta_t$ and, consequently, that the return prediction model exhibits (fractional) cointegration, if ξ_{t-1} is trivial. Hence, we equip returns with a persistent conditional mean and weakly dependent innovations. This is consistent with a vast literature, that finds limited serial correlation in return innovations; see, e.g., the introduction for references.

Remark 1. *Assumption M stipulates that* $\underline{d}_x > 0$, *i.e., that all predictors have long memory. This condition is necessary, when the requisite elements of* $\boldsymbol{c}_{t-1}$ *are non-trivial. That is, we obtain identification of the persistent predictive signals through differences in memory relative to their weakly dependent components (and by using the LCM approach). We can accommodate cases, where* $d_i = 0$, *when* $\boldsymbol{c}_{t-1}(i) = 0$, $\forall t \geq 1$, *which is analogous to assuming exogeneity in OLS settings. Our assumption is reminiscent of the approach in Pastor & Stambaugh (2009), who also, as will be explained below, utilize memory differentials to identify the conditional mean properties of asset returns, but within a more standard weakly dependent setting. Importantly, our empirical application in Section 5 illustrates that popular return predictors from recent macro-finance models, e.g., Bansal, Kiku, Shaliastovich & Yaron (2014) and Campbell, Giglio, Polk & Turley (2018), exhibit strong persistence and may be* *characteristized as either stationary or non-stationary fractionally integrated processes. Hence, despite Assumption M deviating from the literature by requiring fractional integration, rather than weak, local-to-unity or $I(1)$ dependence, this assumption has a solid empirical foundation.*

2.2 Return Regressions: Dynamics and Implications

The predictive system (6)-(10) has several distinct features. First, the regression model is balanced, irrespective of the forecasting prowess of the predictors, that is, $y_t \in I(d_1)$ under both $\mathcal{H}_0 : \boldsymbol{\mathcal{B}} = \mathbf{0}$ and $\mathcal{H}_A : \boldsymbol{\mathcal{B}} \neq \mathbf{0}$. The null hypothesis, $\mathcal{H}_0$, allows for the scenario, where the regressors imperfectly span the conditional mean, i.e., $\varphi_{t-1} = \xi_{t-1} \neq 0$. Under the alternative hypothesis, $\mathcal{H}_A$, the fractional filter adjusts the persistence of the "latent" signals, $\boldsymbol{x}_{t-1}$, to ensure regression balance. If the system is balanced, then $\boldsymbol{Q}(L) = \boldsymbol{I}_k$, a k-dimensional identity matrix, and the adjustment is trivial. To further appreciate the mechanics of the fractional filter, consider a scenario where the conditional mean component has $d_1 = 0.8$, thereby being nonstationary with a well-defined mean. Then, if we observe an explanatory variable with $d_x = 1.8$, the regressor must be transformed to match the persistence of the conditional mean. In this case, the predictor requires a simple difference transformation.

Second, even under $\mathcal{H}_A$, the regressors may be imperfect, that is, ξ_{t-1} may be non-trivial. This captures a scenario, where the predictors contain information about the conditional mean, but fail to fully span its variation. In contrast, if the predictors are "perfect", we have $\varphi_{t-1} = \boldsymbol{\mathcal{B}}'\boldsymbol{u}_{t-1}$.

Third, the system accommodates endogenous regressors through, $\boldsymbol{c}_{t-1}$, which is independent of the persistent signal, $\boldsymbol{x}_{t-1}$. To motivate this model feature, let us draw a parallel to the long-run risk model of Bansal & Yaron (2004), where persistent shocks to the mean and volatility of consumption growth determine the conditional equity premium. In our setting, the persistence of the risk factors is captured by fractionally integrated processes instead of persistent first-order autoregressive (AR) ones, whose half-lives have been stipulated to exceed 52 months (coefficients of 0.979 and 0.987). Moreover, Bansal & Yaron (2004) assume, that these shocks are independent of the innovations to consumption growth. In contrast, we can accommodate a second component in both factors, that are less persistent, but allowed to exhibit non-trivial correlation with the return innovations. These components contain no information about the conditional equity premium, but facilitates richer system dynamics.[10]

Fourth, the model facilitates non-trivial correlation between unspanned component of the conditional mean, ξ_{t-1}, and the observable explanatory variables (again, through $\boldsymbol{c}_{t-1}$) as well as with the innovations to asset returns, $v_t = \eta_t$. This allows for endogeneity through different channels.

Finally, the model allows asset returns to have a weakly dependent component η_t, which may have a "large" volatility relative to the persistent conditional mean, thereby producing a "low" signal-to-noise ratio for the return regression and rendering predictability hard to detect empirically. This feature is consistent with a comprehensive empirical literature, that find limited return serial correlation, yet

[10] A multi-component structure of the conditional mean of consumption growth is consistent with the dynamic decomposition in, e.g., Ortu, Tamoni & Tebaldi (2013), who show that consumption growth has a very persistent component with low volatility as well as a less persistent "error" component with high volatility. Moreover, multi-factor volatility models are used extensively in financial econometrics; see, e.g., Andersen & Benzoni (2012) and many references therein.

predictive power from highly persistent financial and macroeconomic variables; see, e.g., Welch & Goyal (2008), Lettau & Ludvigson (2010) and the many references therein. Likewise, many prominent asset pricing theories, e.g., the present value, long-run risk and dynamic disaster models, stipulate the existence of a persistent conditional mean return with a "low" signal-to-noise ratio.

Altogether, these features mimic the qualitative implications of the predictive system for asset returns in Pastor & Stambaugh (2009), despite arising in our fractionally integrated setting rather than their first-order AR economy. The following remark outline these similarities.

Remark 2. *Pastor & Stambaugh (2009) analyze an asset return system with imperfect predictors, whose components follow stationary AR(1) processes. Adapted to our notation, it takes the form,*

$$\begin{aligned} y_t &= \varphi_{t-1} + \eta_t, \quad \varphi_{t-1} = a_\varphi + \boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{X}}_{t-1} + \xi_{t-1}, \\ \varphi_t &= (1-\phi)\mu_\varphi + \phi\varphi_{t-1} + w_t, \quad \boldsymbol{\mathcal{X}}_t = (\boldsymbol{I}_k - \boldsymbol{\mathcal{A}})\boldsymbol{\mu}_{\mathcal{X}} + \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{X}}_{t-1} + \boldsymbol{u}_t, \end{aligned}$$

where $0 < \phi < 1$*, the eigenvalues of* $\boldsymbol{\mathcal{A}}$ *are inside the unit circle, and the innovation vector* $(\eta_t, w_t, \boldsymbol{u}_t')'$ *is i.i.d. Gaussian. The system features return predictability via the conditional mean (since* $\phi > 0$*), endogenous regressors, and it accommodates imperfect predictors, when* $\varphi_{t-1} \neq a_\varphi + \boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{X}}_{t-1}$*. Moreover, if the predictors are imperfect, this generates unspanned return persistence, as we obtain by inclusion of the component* $\xi_{t-1}^{(-d_1)}$ *in equation* (10)*. Finally, their key identifying assumption for* $\boldsymbol{\mathcal{B}}$ *is* $0 < \phi < 1$*, allowing the persistent conditional mean to be disentangled from the noise. If this assumption fails, they need exogenous regressors. It is analogous to assuming* $\underline{d}_x > 0$ *in Assumption M.*

The model (6)-(10) features four competing hypotheses for the return dynamics:

(i) $\boldsymbol{\mathcal{B}} = \mathbf{0}$ and ξ_{t-1} is trivial, $\forall t = 1, \ldots, n$; returns are not predictable.

(ii) $\boldsymbol{\mathcal{B}} = \mathbf{0}$ and ξ_{t-1} non-trivial, $\forall t = 1, \ldots, n$; returns are not predictable by $\boldsymbol{\mathcal{X}}_{t-1}$.

(iii) $\boldsymbol{\mathcal{B}} \neq \mathbf{0}$ and ξ_{t-1} non-trivial, $\forall t = 1, \ldots, n$; returns are predictable, and $\boldsymbol{\mathcal{X}}_{t-1}$ is "imperfect".

(iv) $\boldsymbol{\mathcal{B}} \neq \mathbf{0}$ and ξ_{t-1} trivial, $\forall t = 1, \ldots, n$; returns are predictable, and $\boldsymbol{\mathcal{X}}_{t-1}$ is "perfect".

The hypotheses (i) and (ii) imply that $\boldsymbol{\mathcal{X}}_{t-1}$ possess no predictive power for returns, but they have different dynamic implications; namely, returns are $I(0)$ and $I(d_1)$, respectively. Moreover, the first hypothesis stipulates, that returns are not predictable by *any* persistent regressor, whereas the second allows for predictability with, however, the "wrong" set of predictors having been examined. The vast empirical literature on return predictability and extensive theoretical developments (again, see the introduction) suggest that, in many settings, we should focus on the null hypothesis given by scenario (ii) rather than (i), especially if examining a set of predictor variables sequentially in single-regressor models. In addition, hypotheses (iii) and (iv) also carry different dynamic implications. Specifically, both hypotheses imply $y_t \in I(d_1)$, but (iii) has regression errors, that are comprised of $\xi_{t-1}^{(-d_1)} \in I(d_1)$ and $\eta_t \in I(0)$ processes, while (iv) describes a fractional cointegration model with $I(0)$ innovations.

The hypotheses imply different inference regime for persistent variables, for which standard OLS is known to deliver spurious inference; see, e.g., Granger & Newbold (1974), Phillips (1987), and Tsay & Chung (2000). Estimation and inference is further complicated by the fact, that the particular scenario as well as the persistence properties of $\boldsymbol{z}_t = (y_t, \boldsymbol{x}'_{t-1})'$ are unknown ex-ante. For example, if we know that y_t and $\boldsymbol{\mathcal{X}}_{t-1} = \boldsymbol{x}_{t-1}$ form a fractional cointegration model (i.e., the signals are significant, observable and "perfect"), one may readily apply inference procedures such as Robinson & Marinucci (2003), Robinson & Hualde (2003), Christensen & Nielsen (2006) and Johansen & Nielsen (2012). Generally, however, we do not know, a priori, which of the hypotheses capture the inference scenario, i.e., whether the regressors are endogenous and/or the predictors are "perfect", and we need to estimate the persistence of $\boldsymbol{z}_t$, which is complicated due to the "low" signal-to-noise ratio for the returns.

As exemplified in Remark 2 and the introduction, related issues have been examined in different predictive settings, assuming stationary first-order AR dynamics, (near) local-to-unity, unit root or locally-explosive persistence. In contrast, we assume a flexible long memory system with similar qualitative features, and we analyze the return predictability via the LCM approach. Moreover, compared with Andersen & Varneskov (2020), we allow for "imperfect" predictors and the simultaneous presence of endogeneity and cointegration.[11] Hence, all subsequent results are new.

Specifically, we provide a (cointegration) rank testing framework that facilitates discriminating between hypotheses (i)-(iv) and allows us to determine the persistence of the conditional mean asset returns. Moreover, we propose a rank-augmented LCM procedure to study return predictability. These are developed with hypotheses (ii)-(iv) in mind, that is, thinking about inference scenarios, where returns have a persistent mean component, and the predictors are either insignificant, imperfect or perfect. However, we emphasize that both our rank test and rank-augmented LCM procedure remain valid in scenario (i), and we provide comments regarding this case throughout.

2.3 The Local Spectrum Approach

The motivation behind the LCM inference and testing procedure is readily conveyed by considering decompositions of the spectral density for the observable regressors, $\boldsymbol{\mathcal{X}}_{t-1}$, and their co-spectrum with the asset returns, y_t. Specifically, using that $\boldsymbol{f}_{xc}(\lambda) \sim \boldsymbol{0}$, as $\lambda \to 0^+$, we may write,

$$\boldsymbol{f}_{\mathcal{X}\mathcal{X}}(\lambda) \sim \boldsymbol{\Lambda}_{xx}^{-1}\, \boldsymbol{G}_{uu}\, \overline{\boldsymbol{\Lambda}}_{xx}^{-1} + \boldsymbol{G}_{cc}, \tag{11}$$

$$\boldsymbol{f}_{\mathcal{X}y}(\lambda) \sim \boldsymbol{\Lambda}_{xx}^{-1}\, \boldsymbol{G}_{uu}\, \boldsymbol{\mathcal{B}}\, \overline{\Lambda}_{yy}^{-1} \;+\; \boldsymbol{f}_{x\xi}^{(-d_1)}(\lambda) \;+\; \boldsymbol{f}_{x\eta}(\lambda) \;+\; \boldsymbol{G}_{c\xi}\, \overline{\Lambda}_{yy}^{-1} \;+\; \boldsymbol{G}_{c\eta}, \tag{12}$$

for $\lambda \to 0^+$, where $\overline{\Lambda}_{yy}$ and $\overline{\boldsymbol{\Lambda}}_{xx}$ are the complex conjugates of Λ_{yy}, respectively, $\boldsymbol{\Lambda}_{xx}$, defined as,

$$\Lambda_{yy} = (1 - e^{\mathrm{i}\lambda})^{d_1}, \qquad \boldsymbol{\Lambda}_{xx} \;=\; \operatorname{diag}\left[(1 - e^{\mathrm{i}\lambda})^{d_2}, \ldots, (1 - e^{\mathrm{i}\lambda})^{d_{k+1}}\right].$$

[11] Andersen & Varneskov (2020) study the asymptotic properties of LCM approach in a general predictive setting. However, when examining the the effect of regressor endogeneity on the inference, they assume cointegration is absent.

These decompositions are intuitive. First, $\boldsymbol{f}_{\mathcal{X}\mathcal{X}}(\lambda)$ shares the multi-component structure of the observable regressors $\boldsymbol{\mathcal{X}}_{t-1}$, with the spectral density of the persistent signal dominating the frequencies in the vicinity of the origin. However, the speed of divergence may differ across elements, depending on the fractional integration orders of the regressors. Second, $\boldsymbol{f}_{\mathcal{X}y}(\lambda)$ not only contains information about the forecasting prowess of the regressors, $\boldsymbol{\mathcal{B}}$, the first term dominates the remaining ones at lower frequency ordinates as $\lambda \to 0^+$. Moreover, $\boldsymbol{G}_{c\xi}\overline{\Lambda}_{yy}^{-1} + \boldsymbol{G}_{c\eta}$ captures an endogeneity-induced bias, which may be persistent and even diverge (when $d_1 > 0$) as $\lambda \to 0^+$, however, at a slower rate than the first term. Finally, the co-spectra $\boldsymbol{f}_{x\xi}^{(-d_1)}(\lambda)$ and $\boldsymbol{f}_{x\eta}(\lambda)$ introduce sampling errors for estimators of $\boldsymbol{\mathcal{B}}$, with their respective asymptotic orders differing due to $\xi_{t-1}^{(-d_1)} \in I(d_1)$ and $\eta_t \in I(0)$.

In general, the (co-)spectral densities in equations (11) and (12) diverge with rates depending on the, possibly, different integration orders of the predictors and asset returns. In contrast, the corresponding co-spectral densities for the unobserved weakly dependent components of the predictive system, $\boldsymbol{u}_{t-1}$ and e_t, are,

$$\boldsymbol{f}_{uu}(\lambda) \sim \boldsymbol{G}_{uu} \quad \text{and} \quad \boldsymbol{f}_{ue}(\lambda) \sim \boldsymbol{G}_{uu}\boldsymbol{\mathcal{B}} + \boldsymbol{f}_{u\xi}(\lambda) + \boldsymbol{f}_{u\eta}^{(d_1)}(\lambda), \tag{13}$$

which are both asymptotically bounded and convey similar information about $\boldsymbol{\mathcal{B}}$. This suggests that inference based on $\boldsymbol{u}_{t-1}$ and e_t may circumvent issues regarding balance, degeneracy of point estimates and spurious inference, motivating Andersen & Varneskov (2020) to introduce the LCM procedure, which consists of two main steps. First, the procedure carries out *fractional filtering* of the observed variables $\boldsymbol{Z}_t = (y_t, \boldsymbol{\mathcal{X}}_{t-1}')'$ to obtain an estimate of $\boldsymbol{v}_t = (e_t, \boldsymbol{u}_{t-1}')'$. Second, it uses medium band least squares (MBLS) estimation for robust inference. These steps are detailed next, together with additional subtleties created by the specific problem of predicting asset returns.

Step 1: Fractional Filtering. As we seek to retain flexibility, allowing for different estimators of the fractional integration orders, we abstain from dedicating a specific estimator and, instead, assume to have one available, $\widehat{d}_i$ for $i = 1, \ldots, k+1$, that satisfies mild consistency requirements.

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

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

Assumption F is very mild, essentially only requiring existence of an estimator which, under appropriate assumptions on equation (7), is consistent. However, since we accommodate both (asymptotically) stationary and non-stationary variables in Assumption M, the estimator must apply for a wide range of d_i. Examples include the semi-parametric exact local Whittle (ELW), see Shimotsu & Phillips (2005) and Shimotsu (2010), the trimmed ELW (TELW) by Andersen & Varneskov (2020), and parametric (long) fractional ARIMA(p, d, q) models using information criteria to determine the short-memory dynamics; see, e.g., Hualde & Robinson (2011) and Nielsen (2015).

Once we obtain the filtering matrix, $\widehat{\boldsymbol{D}}(L)$, the estimates for $\boldsymbol{v}_t$ are,

$$\widehat{\boldsymbol{v}}_t^c \;\equiv\; (\widehat{e}_t, (\widehat{\boldsymbol{u}}_{t-1}^c)')' \;=\; \widehat{\boldsymbol{D}}(L)\boldsymbol{Z}_t, \tag{14}$$

where $\widehat{\boldsymbol{u}}_{t-1}^c = \widehat{\boldsymbol{u}}_{t-1} + \widehat{\boldsymbol{c}}_{t-1}$, with $\widehat{\boldsymbol{u}}_{t-1} = \widehat{\boldsymbol{D}}_x(L)\boldsymbol{x}_{t-1}$ and $\widehat{\boldsymbol{c}}_{t-1} = \widehat{\boldsymbol{D}}_x(L)\boldsymbol{c}_{t-1}$. Similarly, we define $\widehat{\boldsymbol{v}}_t \equiv (\widehat{e}_t, \widehat{\boldsymbol{u}}_{t-1}')'$, which is the equivalent (albeit, unobservable) estimate of $\boldsymbol{v}_t$, without an endogenous component in the regressors. We will, then, utilize frequency domain estimation to extract asymptotically identical information from $\widehat{\boldsymbol{v}}_t^c$ as for $\widehat{\boldsymbol{v}}_t$. Moreover, we leave the mean, or initial value, of the variables unspecified at the filtering stage. Instead, we account for their residual impact on the mean in a Type-II fractional model, $\widehat{\boldsymbol{D}}(L)\,\boldsymbol{\mu}\boldsymbol{1}_{\{t\geq 1\}}$, in a unified manner during second stage estimation.

Step 2: Medium band least squares estimation. We estimate and draw inference about $\boldsymbol{\mathcal{B}}$ using a frequency-domain least squares estimator and $\widehat{\boldsymbol{v}}_t^c$. To define the former, we let,

$$\boldsymbol{w}_h(\lambda_j) \;=\; \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \boldsymbol{h}_t\, e^{\mathrm{i}t\lambda_j}, \qquad \boldsymbol{I}_{hk}(\lambda_j) \;=\; \boldsymbol{w}_h(\lambda_j)\,\overline{\boldsymbol{w}}_k(\lambda_j), \tag{15}$$

be the discrete Fourier transform (DFT) and cross-periodogram, respectively, where $\boldsymbol{h}_t$ and $\boldsymbol{k}_t$ are generic (and compatible) vector time series, and $\lambda_j = 2\pi j/n$ denotes the Fourier frequencies. Moreover, we define the trimmed discretely averaged co-periodogram (TDAC), using the real part of $\boldsymbol{I}_{hk}(\lambda_j)$, as,

$$\widehat{\boldsymbol{F}}_{hk}(\ell, m) \;=\; \frac{2\pi}{n} \sum_{j=\ell}^{m} \Re(\boldsymbol{I}_{hk}(\lambda_j)), \qquad 1 \leq \ell \leq m \leq n, \tag{16}$$

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

$$\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) \;=\; \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}^c(\ell, m)^{-1}\, \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}^c(\ell, m), \tag{17}$$

for which $\ell, m \to \infty$ and $\ell/m + m/n \to 0$, as $n \to \infty$. The MBLS estimator has some distinct advantages for predictive inference and testing with persistent variables. Specifically, by combining sample-size-dependent trimming with a bandwidth $m/n \to 0$, equation (17) turns out to be first-order equivalent to,

$$\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) \;=\; \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1}\, \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}(\ell, m), \tag{18}$$

that is, the corresponding estimator based on $\widehat{\boldsymbol{v}}_t$. In other words, trimming and a local bandwidth suffice to annihilate biases arising as a result of endogenous regressors. Intuitively, this follows from the MBLS estimator utilizing frequencies, that are asymptotically "close" to the origin, which, as shown by the decompositions (11) and (12), are dominated by information about $\boldsymbol{\mathcal{B}}$, whereas the higher frequencies are more prone to endogenous regressor biases. Moreover, the trimming and bandwidth sequences aid in asymptotic elimination of the residual impact from the filtered mean component (mean slippage contamination), occurring at lower frequencies, and first-stage estimation errors from the filtering procedure, occurring at higher frequencies. This suggests that LCM procedure, particularly the second step, should be well-suited to draw inference regarding return predictability.

The main obstacle for using LCM to analyze return regressions is the fractional filtering step. It is complicated due to the low signal-to-noise ratio of the conditional mean relative to the weakly dependent innovations; the return serial dependence is limited, although some highly persistent series often provide significant predictive power for the returns. This suggest that we cannot draw inference about d_1 in finite samples using standard univariate time series techniques and, in fact, we verify these results in both our simulation study and the empirical analysis below. Consequently, the next section provides a new (cointegration) rank testing framework, that not only facilitates discriminating between the model hypotheses (i)-(iv), but also allows us to determine the persistence of the conditional mean return component. Subsequently, in Section 4, we document that this multivariate procedure overcomes the shortcomings of univariate time series techniques in realistic finite sample settings.

3 LCM Rank Testing and Inference

This section provides a new rank test for fractional cointegration, that facilitates discriminating between the model hypotheses (i)-(iv). First, we establish the properties of the test, requiring that Assumption F holds, i.e., we can consistently estimate the fractional integration order for the predictors and the asset returns. As argued above, this assumption is unlikely to hold in finite samples, using standard univariate techniques, as the signal-to-noise ratio of the conditional mean to the return innovations is too "low". Hence, we subsequently outline how a sequence of rank tests may be used to deduce, whether returns have a persistent conditional mean and to determine its fractional integration order. Finally, we establish central limit theory for a rank-augmented LCM (RLCM) procedure.

3.1 LCM Rank Testing for Cointegration

Initially, we suppose that the returns are equipped with a conditional mean, and we know its fractional integration order, $0 \leq d_1 \leq 1$. Then our filtering, heuristically, implies,

$$\begin{cases} (1-L)^{d_1} y_t \ \simeq \ \boldsymbol{\mathcal{B}}' \boldsymbol{u}_{t-1} + \eta_t^{(d_1)} & \text{under models (i) and (iv),} \\ (1-L)^{d_1} y_t \ \simeq \ \boldsymbol{\mathcal{B}}' \boldsymbol{u}_{t-1} + \xi_{t-1} + \eta_t^{(d_1)} & \text{under models (ii) and (iii),} \end{cases} \tag{19}$$

with, again, $\xi_{t-1} \perp\!\!\!\perp \boldsymbol{u}_{t-1}$ by Assumption C. Hence, we can apply this decomposition to test for the presence of ξ_{t-1}. The interpretation of the test, however, depends on the magnitude of d_1. If the returns do not feature a persistent mean component, $d_1 = 0$, in line with scenario (i), then we cannot distinguish ξ_{t-1} and $\eta_t^{(d_1)} = \eta_t$, which are both $I(0)$. As noted in Remark 2, this corresponds to identification failure (when $\phi = 0$) in the imperfect predictor model of Pastor & Stambaugh (2009). However, given the extensive empirical and theoretical evidence on return predictability, our primary focus is on the persistent mean return case, $d_1 > 0$, corresponding to scenarios (ii)-(iv). In these cases, we may utilize the low-frequency spectrum to design a cointegration rank test for the presence of ξ_{t-1} using $\widehat{\boldsymbol{v}}_t^c = (\widehat{e}_t, (\widehat{\boldsymbol{u}}_{t-1}^c)')'$. This test design works, since $\eta_t^{(d_1)}$ is a lower-order residual and has a degenerate spectral density for $\lambda \to 0^+$, as discussed below equations (11) and (12).[12]

Formally, to design a rank testing procedure, we leverage insights from equation (19) and use the fractionally filtered series, $\widehat{\boldsymbol{v}}_t^c$. Hence, we must accommodate estimation errors from filtering, mean-slippage, as well as bias and errors induced by regressor endogeneity, in analogy to the challenges detailed for the second-stage MBLS in Section 2. To this end, we turn to the trimmed long-run covariance estimator,

$$\widehat{\boldsymbol{G}}_{\widehat{v}\widehat{v}}^c(\ell_G, m_G) \;=\; \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \Re\left(\boldsymbol{I}_{\widehat{v}\widehat{v}}^c(\lambda_j)\right), \tag{20}$$

where we use separate bandwidth and trimming functions, $m_G = m_G(n)$ and $\ell_G = \ell_G(n)$. This class of long-run covariance estimators is used for inference and testing in Andersen & Varneskov (2020) and is akin to those in Christensen & Varneskov (2017). If we restrict $\ell = 1$, the estimator also resembles those employed by Robinson & Yajima (2002) and Nielsen & Shimotsu (2007) to design semiparametric tests for fractional cointegration rank in LW and ELW settings, respectively. However, we face additional challenges due to the, possibly, endogenous regressors, fractional filtering induced mean-slippage and estimation errors as well as the lower-order filtering error $\eta_t^{(d_1)}$. Hence, we seek appropriate conditions to prevent either feature from impacting the limiting properties. Moreover, while Andersen & Varneskov (2020) establish consistency of the trimmed estimator (20) for the covariance matrix $\boldsymbol{G}_{\psi\psi}$ with $\boldsymbol{\psi}_{t-1} \equiv (\varphi_{t-1}, \boldsymbol{u}_{t-1}')'$ – either in the case with weak endogeneity, as in Assumption D1, or for the case of stronger endogeneity, but absent cointegration – we now require an associated central limit theory, covering models (ii)-(iv), to design a suitable rank test for (iv).[13]

Assumption T-G. *let the bandwidth $m_G \asymp n^{\kappa_G}$ and $\ell_G \asymp n^{\nu_G}$, with $0 < \nu_G < \kappa_G < \varrho \leq 1$. Then, for some arbitrarily small $\epsilon > 0$, the following cross-restrictions are imposed on ℓ_G, m_G, m_d and n,*

$$\frac{m_G^{1+2\varpi}}{n^{\varpi}} + \frac{n}{\ell_G^2\sqrt{m_G}}\left(\frac{1}{\sqrt{m_G}} + \left(\frac{m_G}{n}\right)^{\underline{d}}\right) + \frac{\ell_G}{\sqrt{m_G}} + \left(\frac{m_G}{n}\right)^{d_x}\frac{\sqrt{m_G}}{\ell_G^{1+\epsilon}} + \to 0, \quad \text{as} \quad n \to \infty.$$

The first condition in Assumption T-G is familiar from semiparametric frequency domain estimation, e.g., Robinson & Yajima (2002) and Nielsen & Shimotsu (2007). For the empirically relevant vector ARFIMA process (with $\varpi = 2$), it requires $\kappa_G < 4/5$. In contrast, the last three conditions impose joint bounds on the bandwidth and trimming rates. Specifically, two and four stipulate a lower bound on the trimming to eliminate the bias from mean-slippage and regressor endogeneity,

[12] We explain in the next section how to use sequential rank tests to verify that $d_1 > 0$, indeed, holds. Still, it is worth noting that our rank test will work, even when $d_1 = 0$; indicating full rank due to the presence of both ξ_{t-1} and η_t.

[13] Again, we use the definition $\boldsymbol{\psi}_{t-1} \equiv (\varphi_{t-1}, \boldsymbol{u}_{t-1}')'$ to indicate that weakly dependent return innovations have no asymptotic impact on the limit theory, when $d_1 > 0$ in scenarios (ii)-(iv).

respectively, and condition three restricts the loss of information. Taken together,

$$\frac{1-\kappa_G}{2} \vee \frac{(1-\kappa_G)(1-\underline{d})+\kappa_G/2}{2} \vee \frac{\kappa_G}{2} - (1-\kappa_G)\underline{d}_x < \nu < \frac{\kappa_G}{2}\,, \tag{21}$$

These bounds are quite restrictive, if $0 < \underline{d} \leq \underline{d}_x$ is small. Moreover, we require $\underline{d}_x > 0$ for ν to be defined on an open interval, again, illustrating the importance of this identifying condition, when the regressors are endogenous. The fourth restriction can be dispensed with if $\boldsymbol{c}_{t-1} = \mathbf{0}$, $\forall t = 1, \ldots, n$. It is important to note that, if the conditional mean of returns and the regressors are strongly persistent, e.g., $\underline{d} \simeq 1$, the lower bound simplifies to $(1-\kappa_G)/2 \vee \kappa_G/4 \vee 3\kappa_G/2 - 1 < \nu$, which is very mild.

Theorem 1. *Suppose Assumptions D1-D3, C, M, F and T-G hold,* $0 < d_1 \leq 1$*, and* $n^{1/2}/m_G \to 0$*. Then, by letting* $\boldsymbol{G}_{\psi\psi}^{(i)}$ *be the* $i = 1, \ldots, k+1$ *column of* $\boldsymbol{G}_{\psi\psi}$*, it follows,*

$$m_G^{1/2}\text{vec}\left(\widehat{\boldsymbol{G}}^c_{\widehat{v}\widehat{v}}(\ell_G, m_G) - \boldsymbol{G}_{\psi\psi}\right) \xrightarrow{\mathbb{D}} N\left(\mathbf{0}, \left(\boldsymbol{G}_{\psi\psi} \otimes \boldsymbol{G}_{\psi\psi} + \left(\boldsymbol{G}_{\psi\psi} \otimes \boldsymbol{G}^{(1)}_{\psi\psi}, \ldots, \boldsymbol{G}_{\psi\psi} \otimes \boldsymbol{G}^{(k+1)}_{\psi\psi}\right)\right)/2\right).$$

Theorem 1 shows that the trimmed long-run covariance estimator attains asymptotic properties mirroring those of Robinson & Yajima (2002, Propositions 2-3) and Nielsen & Shimotsu (2007, Lemmas 4-5). Hence, despite the additional challenges in the current environment, we may utilize their procedures to study the (cointegration) rank of $\boldsymbol{G}_{\psi\psi}$ and, thus, whether predictive model (iv) should replace models (ii) or (iii). Moreover, the analysis simplifies, as we do not seek to determine an exact cointegration rank of a system, but rather to test the null hypothesis $\widetilde{\mathcal{H}}_0 : \text{rank}(\boldsymbol{G}_{\psi\psi}) = k+1$ against the specific alternative $\widetilde{\mathcal{H}}_A : \text{rank}(\boldsymbol{G}_{\psi\psi}) = k$, because $\boldsymbol{u}_{t-1}$ is of full rank.

Remark 3. *A result analogous to Theorem 1 still holds, if* $b = d_1 = 0$*, as for scenario (i). Specifically, we need to write* $\boldsymbol{\psi}_t \equiv \boldsymbol{v}_t$ *and* $\boldsymbol{G}_{\psi\psi} \equiv \boldsymbol{G}_{vv}$*, since* $\eta_t^{(b)} = \eta_t$ *and* ξ_{t-1} *are both* $I(0)$*. Moreover, we can relax the second trimming condition in Assumption T-G, which is required to eliminate the lower-order error (when* $d_1 > 0$*); namely,* $\eta_t^{(d_1)}$*. In scenario (i),* $\boldsymbol{G}_{\psi\psi}$ *will be of full rank, and this will be indicated by our subsequent rank selecetion procedures with probability approaching 1. Hence, despite being developed with models (ii)-(iv) and* $d_1 > 0$ *in mind, our approach still applies for scenario (i).*

Next, to estimate the rank, we let δ_i and $\widehat{\delta}_i$, $i = 1, \ldots, k+1$ denote the eigenvalues of the covariance matrices $\boldsymbol{G}_{\psi\psi}$ and $\widehat{\boldsymbol{G}}^c_{\widehat{v}\widehat{v}}(\ell_G, m_G)$, listed in descending order, $0 < \delta_k < \cdots < \delta_1$, with $0 < \delta_{k+1} < \delta_k$ and $\delta_{k+1} = 0$ under $\widetilde{\mathcal{H}}_0$ and $\widetilde{\mathcal{H}}_A$, respectively. Then, for $r = 0$ and $r = 1$ indicating the rank reduction under the two hypotheses, we follow Robinson & Yajima (2002) and Nielsen & Shimotsu (2007) and estimate r as,

$$\widehat{r} = \underset{\varrho \in \{0,1\}}{\text{argmin}}\, \mathcal{L}(\varrho), \qquad \mathcal{L}(\varrho) = \vartheta(n)(k+1-\varrho) - \sum_{i=1}^{k+1-\varrho} \widehat{\delta}_i, \tag{22}$$

for some $\vartheta(n) > 0$, which is assumed to obey the conditions:

Assumption V. *The sequence* $\vartheta(n)$ *satisfies* $\vartheta(n) + \frac{1}{\vartheta(n)\sqrt{m_G}} \to 0$ *as* $n \to \infty$*.*

Theorem 2. *Suppose the conditions of Theorem 1 and Assumption V hold, then,*

$$\lim_{n\to\infty} \mathbb{P}(\widehat{r} = r) = 1.$$

The rank selection procedure is consistent and, thus, facilitates discrimination between the predictive models (ii) or (iii) and the cointegration model (iv). Moreover, as discussed in, e.g., Phillips & Ouliaris (1988), Robinson & Yajima (2002) and Nielsen & Shimotsu (2007), the rank selection procedure may be implemented using the corresponding (trimmed) correlation matrix estimator,

$$\widehat{\boldsymbol{P}}^{c}_{\widehat{v}\widehat{v}}(\ell_G, m_G) \equiv \mathrm{diag}(\widehat{\boldsymbol{G}}^{c}_{\widehat{v}\widehat{v}}(\ell_G, m_G))^{-1/2}\, \widehat{\boldsymbol{G}}^{c}_{\widehat{v}\widehat{v}}(\ell_G, m_G)\ \mathrm{diag}(\widehat{\boldsymbol{G}}^{c}_{\widehat{v}\widehat{v}}(\ell_G, m_G))^{-1/2}. \tag{23}$$

In line with their numerical evidence, we find that the correlation-based procedure performs substantially better in (unreported) finite sample simulation settings, which we return to in Section 4.

As discussed above, the validity of the described rank testing procedure depends on a reliable finite sample estimate of d_1 via Assumption F and, to this end, we cannot rely on univariate time series techniques. Hence, we continue by demonstrating how a sequence of rank tests may be used to determine d_1, in addition to discriminating between the hypotheses $\widetilde{\mathcal{H}}_0$ and $\widetilde{\mathcal{H}}_A$.

Remark 4. *In Appendix A.1, we describe an alternative rank test based on Theorem 1 and feasible inference for the eigenvalues, inspired by Phillips & Ouliaris (1988). This performance of this testing procedure, however, falls short of the corresponding one based on equations* (22)-(23) *in realistically calibrated finite-sample simulation settings. Hence, these results are omitted for brevity.*

3.2 Assessing Cointegration and Persistence in Returns

The vector of (latent) regressors, $\boldsymbol{u}_{t-1}$, is required to have full rank by Assumption D1, implying that the consistent LCM-based rank selection procedure based on equations (22)-(23) may be applied bivariately, sequentially pairing each regressor with the returns. Because we cannot estimate d_1 reliably in finite samples via standard univariate techniques, we implement a restricted version of the test by replacing $\widehat{d}_1$ with $\widehat{d}_i$, $i = 2, \ldots, k+1$, corresponding to the relevant regressor, assuming that,

$$d_1 \in \{d_2, \ldots, d_{k+1}\}, \quad \text{and still} \quad 0 < d_1 \leq 1, \tag{24}$$

i.e., the vector of regressors has been chosen "sensibly". For example, the regressors may include theory-guided state variables from the long-run risk, dynamic disaster and present-value models. Then, under equation (24), an estimate $\widehat{r}_i = 1$ with $0 < \widehat{d}_i \leq 1$, $i = 2, \ldots, k+1$, constitutes evidence in favor of fractional cointegration using the $(i-1)$th regressor. Similarly, a corresponding estimate $\widehat{r}_i = 0$ indicates that returns have *not* been over-differenced, and that the predictive relation is "imperfect", suggesting that we may "search" among the integration orders of the candidate predictors. However, the procedure must be implemented thoughtfully. An indiscriminate inclusion of (irrelevant) predictors, or regressors with integration orders larger than 1, will render the imposition of the maximal integration order among the regressors in our testing procedure for the conditional mean return problematic. The issue is, of course, that it will tend to generate rejections of the full rank hypothesis due to the ex-ante restricted integration order in the test $d_1 = d_i$, $i = 2, \ldots, k+1$, being too large. Specifically, suppose we include some auxiliary $I(d_x)$ predictor with $d_x > d_1$, then, heuristically,

$$\begin{cases} (1-L)^{d_x} y_t \simeq \boldsymbol{\mathcal{B}}'(1-L)^{d_x-d_1}\boldsymbol{u}_{t-1} + \eta_t^{(d_x)} & \text{under model (iv)}, \\ (1-L)^{d_x} y_t \simeq \boldsymbol{\mathcal{B}}'(1-L)^{d_x-d_1}\boldsymbol{u}_{t-1} + (1-L)^{d_x-d_1}\xi_t + \eta_t^{(d_x)} & \text{under models (ii) and (iii)}, \end{cases}$$

generating a rank estimate $\widehat{r}_x = 1$ with probability approaching one, because the spectral density of the, in this case, over-differenced return series will be degenerate as $\lambda \to 0^+$.

To guard against this type of mechanical over-differencing, we advocate a screening procedure. Let $\mathcal{R}$ denote the set of regressors, that generate estimates $\widehat{r}_i = 1$ with $0 < \widehat{d}_i \leq 1$ in a bivariate rank test with returns – indicating cointegration. Our sequential screening procedure now takes the form,

Step 1. Carry out k bivariate rank tests with the returns. If $\mathcal{R} = \emptyset$, suggesting no cointegration, choose $d_1 = \max_{i=2,\ldots,k+1}(d_i \,|\, 0 < d_i \leq 1)$ and stop. If $\mathcal{R} \neq \emptyset$, proceed to Step 2.

Step 2. Impose $d_1 = \max_{i=2,\ldots,k+1}(d_i \,|\, \boldsymbol{u}_{t-1}(i-1) \in \mathcal{R})$ and carry out all k bivariate rank test. If $\widehat{r}_i = 0$ is maintained for at least *one* variable, stop. If not, choose the second-largest d_i among the variables in $\mathcal{R}$ and repeat step 2. Continue until $\widehat{r}_i = 0$ is maintained for at least *one* variable.

Step 2 exploits the fact that a single non-rejection indicates the returns, for the given selection $d_1 \in (0, 1]$, have *not* been over-differenced. The identical argument motivates selecting $d_1 = \max_{i=2,\ldots,k+1}(d_i \,|\, 0 < d_i \leq 1)$, if $\mathcal{R} = \emptyset$. In fact, if $\widehat{r}_i = 0$ is estimated for just a single persistent ($d_1 > 0$) candidate predictor, even an irrelevant one, it provides consistent evidence against a weakly dependent return series. That is, if we find, for fractionally filtered returns, that the rank is non-degenerate relative to a fractionally integrated series, then the conditional mean return must display fractional persistence, as $(1-L)^{d_1}\eta_t$ will be a lower-order residual for all $d_1 > 0$. Moreover, by implementing the rank selection procedure sequentially, we may directly assess the (induced) degree of persistence of the conditional mean, and which of the inference scenarios (ii)-(iv) apply.

Remark 5. *To illustrate the workings of two-step procedure, we consider an example. Suppose there are two candidate predictors $\boldsymbol{\mathcal{X}}_{t-1} = (\mathcal{X}_{1,t-1}, \mathcal{X}_{2,t-1})'$ with integration orders $(d_2, d_3)' = (0.45, 0.80)'$. Moreover, let $\mathcal{X}_{1,t-1}$ be insignificant and $\mathcal{X}_{2,t-1}$ be "perfect", implying that model (iv) applies and $d_1 = d_3$. The integration orders $(d_2, d_3)'$ are estimated consistently via Assumption F. In this case, our bivariate rank test for returns and $\mathcal{X}_{2,t-1}$ using $\widehat{d}_3$ will (asymptotically) indicate cointegration, belonging to the set $\mathcal{R}$. The test for $\mathcal{X}_{1,t-1}$ using $\widehat{d}_2$ may or may not indicate cointegration.*[14] *Regardless, since*

[14]The reason is that returns have not been filtered sufficiently, $(1-L)^{d_2-d_1}(\boldsymbol{\mathcal{B}}\boldsymbol{u}_{t-1} + \xi_t) \in I(d_1 - d_2)$, with $d_1 > d_2$. Hence, the first element of the vector in the rank test may be of "large enough" asymptotic order relative to the second fractionally filtered element of the vector, which is $I(0)$, to indicate (spurious) cointegration in finite samples.

$d_3 > d_2$, we repeat the test for $\mathcal{X}_{1,t-1}$ with $\widehat{d}_3$ in Step 2, which will now (asymptotically) indicate full rank. Hence, the procedure stops here; we select $d_1 = d_3$ and know that inference scenario (iv) applies.

Remark 6. *If scenario (i) describes the return model, then we will find $\widehat{r}_i = 1$ for all $i = 2, \dots k+1$ and $d_i > 0$, with probability approaching one, since returns are over-differenced for any $d_i > 0$. Hence, our two-step procedure implicitly provide information about the validity of this model.*

3.3 Limit Theory for LCM

Beyond reliable estimates for $\widehat{\boldsymbol{D}}(L)$, satisfying Assumption F, obtained either via univariate time series and/or rank-augmented techniques, we also require trimming and bandwidth conditions for the second-stage MBLS estimator, along the lines of Assumption T-G. As noted, we impose $b = d_1$ throughout. We state the requisite conditions in terms of b for comparability with Andersen & Varneskov (2020).

Before proceeding, note, again, that we develop LCM inference with scenarios (ii)-(iv) in mind, but, as described below, the asymptotic results pertain equally to model (i).

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

$$\frac{m^{1+2\varpi}}{n^{2\varpi}} + \frac{\ell^{1+\varpi+b}}{n^{\varpi}\, m^{1/2+b}} + \frac{n^{1/2+b}}{m_d^{1/2}\, m^b\, \ell} + \frac{n^{1-2\underline{d}+b}}{m^{1/2-2\underline{d}+b}\, \ell^2} + \frac{n^b}{m^{1/2+b}} + \frac{m^{1/2+\underline{d}_x-b}}{n^{\underline{d}_x-b}\, \ell} \to 0.$$

The restrictions in Assumption T are mild. The first term is standard for semiparametric estimation in the frequency domain, see, e.g., Robinson (1995) and Lobato (1999), while the remaining conditions are specific to the second-stage MBLS estimator, adopted in the LCM procedure. Specifically, condition two, implying $\nu < (\varpi + \kappa(1/2 + b))/(1 + \varpi + b)$, restricts the loss of information from trimming frequencies; three, $(1 - \varrho)/2 + b(1 - \kappa) < \nu$ in conjunction with $0 < \nu < \kappa < \varrho \leq 1$, eliminates errors from estimating the integration orders; four, $(1 - \kappa/2 - (2\underline{d} - b)(1 - \kappa))/2 < \nu$, alleviates the low-frequency bias from mean-slippage following fractional filtering; five, $b/(1/2 + b) < \kappa$ imposes a mild bound on the bandwidth; six, $\kappa/2 - (1 - \kappa)(\underline{d}_x - b) < \nu$ eliminates the endogeneity bias.[15]

If we consider the empirically relevant vector ARFIMA process (with $\varpi = 2$) and select κ close to its upper bound 4/5, conditions two and four imply $(3/5 - (2\underline{d} - b)/5)/2 < \nu < 4/5$. The lower bound is strictly decreasing in $2\underline{d} - b \geq 0$ (as assumed below), implying that its most restrictive scenario is obtained when $\underline{d} = 0$, equaling 3/10. The third condition is (essentially) trivial, if we adopt a parametric first-stage estimator with $\varrho = 1$ and κ close to 4/5. If the estimator is semiparametric, however, and we select $\kappa < \varrho$ as well as ϱ arbitrarily close to 4/5, the additional lower bound requirement on the trimming rate becomes $1/10 + b/5 \leq 3/10 < \nu$. Finally, if the regressors are endogenous and we select κ close to 4/5 for efficiency, we require $2/5 - (\underline{d}_x - b)/5 < \nu$, with most conservative bound being

[15] We note that the trimming and bandwidth functions in Assumption T are mutually consistent for all values of $0 < \underline{d}_x < 2$ and $0 \leq d_1 \leq 1$ if the (implied) condition $\max(0, (1 - 3\kappa/2)/(1 + \kappa/2)) < \varpi \leq 2$ holds.

obtained when $\underline{d}_x - b = 0$. Intuitively, we require stronger trimming to obtain the same asymptotic efficiency in the presence of endogenous regressors, if the excess persistence of the system is small.

Theorem 3. *Suppose Assumptions D1-D3, C, M, F and T hold as well as the conditions* $0 < d_1 \leq 1$, $b \leq \underline{d}$, $n^{1/2}/m \to 0$, *and* $\max(0, (1-3\kappa/2)/(1+\kappa/2)) < \varpi \leq 2$, *then,*

$$\begin{cases} \sqrt{m}\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \boldsymbol{\mathcal{B}}\right) \xrightarrow{\mathbb{D}} N\big(\mathbf{0}, \boldsymbol{G}_{uu}^{-1} G_{\xi\xi}/2\big), & \textit{under models (ii) and (iii)}, \\ \sqrt{m}\,\lambda_m^{-b}\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \boldsymbol{\mathcal{B}}\right) \xrightarrow{\mathbb{D}} N\big(\mathbf{0}, \boldsymbol{G}_{uu}^{-1} G_{\eta\eta}/(2(1+2b))\big), & \textit{under model (iv)}. \end{cases}$$

Theorem 3 demonstrates that the LCM procedure, possibly augmented with the rank test to determine d_1, is asymptotically Gaussian for both the predictive models (ii) and (iii) and the cointegration model (iv). The asymptotic distribution theory differs, however, in the two cases. First, when the regressors are "imperfect", ξ_{t-1} is an asymptotic order larger than η_t and drives the limit theory. The convergence rate is $\sqrt{m}$, 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). Second, if the regressors are "perfect", ξ_{t-1} is trivial and the limit theory is determined by η_t. The rate is $\sqrt{m}\lambda_m^{-b} \asymp \sqrt{m}(n/m)^b$ and the asymptotic variance is scaled by $1/(2(1+2b))$. Hence, cointegration improves efficiency of the MBLS estimator, in analogy with super consistency properties.

Despite the limit theory differing across models (ii)-(iii) and (iv), it remains Gaussian in both cases, regardless of whether the variables are (asymptotically) stationary or non-stationary, whether there is cointegration, and irrespective of the cointegration being weak ($b < 1/2$) or strong ($b \geq 1/2$). This is unique within a fractional cointegration context, as similar uniformity applies neither to OLS, the narrow band least squares (NBLS) estimator, nor for maximum likelihood inference in the fractionally cointegrated VAR model, where the inference is Gaussian under stationary and exhibits different forms of non-Gaussianity in non-stationary cases; see Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Johansen & Nielsen (2012).[16] Similarly, the Gaussian limit theory for the MBLS estimator without fractional filtering in Christensen & Varneskov (2017) holds only for stationary systems with weak cointegration. Intuitively, the Gaussian limits in Theorem 3 follow from having fractionally filtered the variables such that the inference, after eliminating various errors and biases through trimming, becomes reminiscent of the ELW inference in Shimotsu & Phillips (2005).

Moreover, the asymptotic distribution theory of the LCM procedure is correctly centered, thus free from bias due to persistent and endogenous regressors, which are detailed by Stambaugh (1999), Pastor & Stambaugh (2009) and Phillips & Lee (2013). Interestingly, since the fractional filtering lowers the asymptotic order of the weakly dependent innovations, η_t, regardless of the inference scenario, the LCM procedure may also provide finite sample improvements by alleviating attenuation biases. An

[16]Such methods generally do not accommodate non-trivial means, or initial values as well as strong endogeneity among the regressors that may or may not be "perfect". Moreover, the limit theory of these alternatives rely on the presence of cointegration. Finally, as demonstrated by Andersen & Varneskov (2020, Theorem 5), the LCM procedure can accommodate regressors that are generated from pre-estimated fractional cointegration residuals. Consequently, the LCM procedure remains desirable in this context, delivering added robustness along with a fast convergence rate.

additional advantage of the Gaussian limit theory is that feasible inference and testing is standard, once we obtain a consistent estimator of the asymptotic variance in the requisite inference scenario. Specifically, the latter can be determined by estimating b using our rank-selection procedure from Section 3.2. We detail how to draw feasible inference in Appendix A.2.

Remark 7. *We impose $0 < d_1 < 1$ for Theorem 3, but can accommodate the case $d_1 = 0$, with appropriate changes to the asymptotic variance for both models (ii)-(iii) and (iv). In particular, for the former, we have to replace $G_{\xi\xi}$ with $G_{\xi\xi} + G_{\eta\eta}$. Moreover, since cointegration no longer features in model (iv), by $b = d_1 = 0$, the result is readily obtained by setting $b = 0$ in the limit theory, thereby slowing down the convergence rate. Similarly, the condition $\underline{d} - b \geq 0$ is equivalent to the "balanced cointegration" requirement in Andersen & Varneskov (2020, Eq. (8)), implying that the cointegration cannot be stronger than the persistence of the regressors. This condition is not strictly binding, but simplifies the tuning parameter restrictions on ℓ and m in Assumption T considerably.*

Remark 8. *The conditions $\sqrt{n}/m_G \to 0$ and $\sqrt{n}/m \to 0$, as $n \to \infty$, in Theorems 1 and 3, respectively, are not strictly binding, but are imposed for ease of exposition. Specifically, they are used to bound the endogenous regressor bias in auxiliary Lemmas B.1(a)-(d) (cf., Appendix B). Define,*

$$\bar{f}(m,n) \equiv 1 \vee (m/n)^{\underline{d}_x} n^{1/2}/m, \quad \bar{f}_G(m,n) \equiv 1 \vee (m_G/n)^{\underline{d}_x} n^{1/2}/m_G,$$

then the conditions can be relaxed to $\sqrt{n}/(m^{1-\epsilon}\sqrt{m_d}) + \sqrt{n}/(m_G^{1-\epsilon}\sqrt{m_d}) \to 0$, for some small $\epsilon > 0$, if multiplying the bounds in Lemma B.1(a)-(b) and (c)-(d) with $\bar{f}(m,n)$ and $\bar{f}_G(m,n)$, respectively. This feature is important, as we also entertain the selection $m_G \asymp n^{\kappa_G}$, with $\kappa_G = 2/5$ for our cointegration rank procedure, which is valid, albeit with stronger cross-restrictions on the tuning parameters.

While the asymptotic properties for the LCM procedure are highly desirable, it is prudent to study its finite sample performance for return regressions in realistic settings, in particular, the interplay between fractional filtering, MBLS estimation and rank testing. This is examined next.

4 Return Regressions and Numerical Evidence

This section illustrates inferential problems surrounding return regressions in a transparent numerical setting. Specifically, we explore the effects of increasing the noise-to-signal ratio of the return regressions for estimates of its fractional integration order as well as the size and power properties of predictability tests relying on either OLS, IVX or LCM inference. In particular, the size is assessed within an imperfect predictor specification. Moreover, we examine the size and power properties of the cointegration rank selection procedure based on equations (22)-(23). Finally, we study the bias and RMSE of LCM estimates of $\boldsymbol{\mathcal{B}}$, with and without applying a rank-augmented estimate of d_1.

4.1 Simulation Setting

We study inference problems for return regressions in a setting reminiscent of the ones in Hong (1996) and Shao (2009), albeit allowing the variables to exhibit non-stationary fractional integration. Specifically, we suppose $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{\mathcal{X}}_{t-1}$ are univariate (written as $\mathcal{B}$ and $\mathcal{X}_{t-1}$, respectively) and stipulate that $\mathcal{X}_{t-1} = x_{t-1}$, which renders the signal of the persistent regressor directly observable and excludes endogeneity. Then, we generate fractional ARMA$(0,0)$ processes as,

$$(1-L)^{d_1}(y_t - \mu_y) = \varphi_{t-1} + \eta_t^{(d_1)}, \quad \varphi_{t-1} = \mathcal{B}u_{t-1} + \mathcal{B}_\xi \xi_{t-1}, \quad (1-L)^{d_2}(x_{t-1} - \mu_x) = u_{t-1}, \tag{25}$$

and $\eta_t^{(d_1)} = (1-L)^{d_1}\eta_t$, where $\zeta_t \sim$ i.i.d. $N(0, \sigma_\zeta^2)$ for $\zeta_t \in \{\eta_t, \xi_{t-1}, u_{t-1}\}$. Moreover, to highlight the impact of the noise-to-signal ratio for drawing inference about return persistence and predictability, we fix $d_1 = d_2 = d$, set $\mu_y = \mu_x = 1/2$ and, without loss of generality, $\sigma_\xi = \sigma_u = 1$, while varying the volatility of the return innovations, σ_η. This ensures that the dynamic properties of the predictive system are captured solely by d, $\mathcal{B}$, $\mathcal{B}_\xi$, and σ_η. We consider two long-memory regimes: $d = 0.45$ and $d = 0.80$, corresponding to a stationary predictable return component versus one that is non-stationary, yet mean-reverting. As we vary $\sigma_\eta \in [0, 25]$, the noise-to-signal of the predictive relation is altered, possibly rendering the persistence of the conditional mean undetectable in finite samples.[17]

Initially, we entertain univariate predictions using x_{t-1}, but fix $\mathcal{B} = 0$ and $\mathcal{B}_\xi = 1.2$ in equation (25), implying that asset returns are comprised of a persistent mean, but the empiricist employs an irrelevant "imperfect" predictor, so that the persistence "spills over" into the residuals. In this scenario, we assess if and when σ_η is sufficiently large to induce "incorrect" inference regarding the fractional integration order of the returns, $\widehat{d_1} \simeq 0$, as is generally found empirically. Moreover, we examine the size properties of predictability tests with, seemingly, $I(0)$ returns using either OLS, IVX or LCM. These three inference procedures are all "misspecified," in the sense that OLS and IVX inference generally does not apply to fractionally integrated systems, as discussed in the introduction, whereas LCM is implemented using the "wrong" fractional integration order for the returns.

We implement IVX with parameters $C_{\text{IVX}} = 1$ and $\beta_{\text{IVX}} = 0.95$ to construct the self-filtered instrument, an additional deterministic instrument $\sin((t-1)\pi/n)$, $t = 1, \ldots, n$, and Eicker-White standard errors, in line with the recommendations in Breitung & Demetrescu (2015) and Kostakis et al. (2015). Similarly, we employ Eicker-White inference for OLS.[18] Moreover, we implement LCM using trimming and bandwidth parameters $(\nu, \kappa) = (0.20, 0.60)$. These are similar to the ones considered in Andersen & Varneskov (2020) and reflect the dynamic properties of returns and, especially, the persistent predictor variables in Section 5. Specifically, the bandwidth is chosen locally $(m/n \to 0)$ to avoid placing excessive weight on the higher-frequency errors from η_t and the trimming reflects condition two in Assumption T, with $\underline{d}_x = 0.3$ in the empirical analysis. Despite the results not

[17] We have run similar experiments with ARMA$(1,0)$ short-run dynamics. The corresponding results, when allowing for mild autoregressive dynamics in the processes, are almost identical to those presented in Figures 1-2 below.

[18] We have also carried out OLS-based testing for return predictability using Newey & West (1987) standard errors. The results are almost identical to those presented in Figure 1, and thus omitted for brevity.

being reported, we note that, importantly, the results are qualitatively robust to varying the tuning parameters ν and κ by ± 0.05 and ± 0.10, respectively. The first stage estimates of d_1 and d_2 are constructed using the TELW estimator of Andersen & Varneskov (2020), with corresponding trimming and bandwidth parameters $\ell_d = \lfloor n^{0.3} \rfloor$ and $m_d = \lfloor n^{0.7} \rfloor$. Moreover, the significance tests for LCM is implemented using the feasible inference procedure in Andersen & Varneskov (2020), see also Appendix A.2, where the consistent spectrum estimator of the asymptotic variance is implemented with $\nu_G = 0.20$ and $\kappa_G = 0.60$. Finally, we consider a sample size $n = 650$, mimicking the one for the empirical analysis ($n = 661$), a 5% nominal test size, and using 1000 replications.

4.2 Persistence, Size, Bias and RMSE

The estimated integration order of the returns, $\widehat{d}_1$, and the OLS, IVX and (misspecified, depending on $\widehat{d}_1$) LCM test sizes are displayed as functions of σ_η in Figure 1, whereas Figure 2 provides bias and RMSEs of the corresponding LCM coefficient estimates, benchmarked against an oracle implementation of LCM, where the true d_1 and d_2 are treated as known in the fractional filtering. Several features are noteworthy. One, the estimated persistence decreases as a function of σ_η, eventually implying failure to reject $d_1 = 0$. This occurs, not surprisingly, more rapidly for the weaker signal, $d = 0.45$, than for the stronger one, $d = 0.80$, illustrating that returns may, possibly, have a persistent component that is hard to identify using standard univariate time series techniques.

Two, OLS and IVX are oversized for a wide range of σ_η, even when the estimated fractional integration order, seemingly, suggests that the return series is $I(0)$. This is akin to the well-established spurious inference problem, arising when applying least squares to fractionally integrated processes, e.g., Tsay & Chung (2000), and the size distortions for return regressions, when applying persistent AR(1) predictors, e.g., Ferson et al. (2003). The current results demonstrate, that similar problems may arise for return regressions with "imperfect" predictors in fractionally integrated settings. Moreover, not only is OLS-based tests unreliable, generating serious size distortions, but similar problems arise for IVX, although the procedure, otherwise, is equipped to handle local-to-unity regressors.

Three, whereas the size of misspecified LCM-based tests appears considerably more accurate, the bias and RMSE of the coefficient estimates depend critically on whether the regressors are significant, i.e., whether $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$ or $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$. When $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$, the LCM estimates are unbiased, but, not surprisingly, less efficient that the oracle ones. In contrast, when the regressors are significant, $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$, the coefficient estimates are severely biased, and this bias raises the RMSE, in particular, for smaller values of σ_η. The bias is intuitive; the latent signal of y_t has not been fractionally filtered (since $\widehat{d}_1 \simeq 0$) and the resulting higher asymptotic order of the conditional mean, thus, blows up the estimate, since $\mathcal{B} > 0$ and $\widehat{\boldsymbol{u}}_{t-1}^c$ has been filtered "correctly".

At face value, these results are discouraging. OLS and IVX suffer from large size distortions, and our original LCM procedure is also ill-equipped to analyze significant return regressors. Hence, to explore whether it provides a potential remedy, we now examine the finite-sample properties of our rank test and the rank-augmented LCM procedure.

4.3 Rank Testing and Rank-augmented LCM Estimation

We examine the properties of the rank selection procedure based on equations (22) and (23) using the setting above, with $d \in \{0.45, 0.80\}$ and either $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$ or $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$, corresponding to the hypotheses $\widetilde{\mathcal{H}}_0 : r = 0$ and $\widetilde{\mathcal{H}}_A : r = 1$, respectively. Specifically, we implement the test with tuning parameters $\nu_G = 0.20$ and $\kappa_G \in \{0.40, 0.50, 0.60\}$. Moreover, we let the sequence in Assumption V be specified as $\vartheta(n) = n^{-\varkappa}$ and examine $\varkappa \in \{0.10, 0.20, 0.30\}$. Finally, to gauge the large(r) sample properties of the procedure, we consider $n = 650$ as well as $n = 2000$. Tables 1-2 provide rejection frequencies of $\widetilde{\mathcal{H}}_0$ in favor of $\widetilde{\mathcal{H}}_A$, when $\widetilde{\mathcal{H}}_0$ is correct and, similarly, Tables 3-4 display corresponding rejection rates, when $\widetilde{\mathcal{H}}_A$ is correct, reflecting "size and power" properties, respectively.[19]

First, we observe the that the "size" of the tests is very good, except when combining the tuning parameter selections $\kappa_G = 0.40$ and $\varkappa = 0.10$. For example, when $\kappa_G = 0.40$, $\varkappa = 0.20$ and $n = 650$, the procedure selects the wrong rank in approximately 5% of the simulations, reminiscent of a 5% test size. Second, we observe that the selection procedure generally has good "power" properties, especially when $\kappa_G < 0.60$ and $\varkappa \leq 0.20$, far exceding what is achieved by the corresponding univariate significance tests for the TELW estimator in Figure 1. Again, if considering $\kappa_G = 0.40$ and $\varkappa = 0.20$, the rejection rates are substantially above 5% for both $d = 0.45$ and $d = 0.80$ in Tables 3 and 4, respectively. Moreover, the rejection rates remain non-trivial for large values of the return innovations, σ_η, and they are, not surprisingly, larger for the stronger signal, $d = 0.80$, than the weaker, $d = 0.45$. Finally, when the sample size increases to $n = 2000$, the rejection frequencies in Tables 1 and 2 converge to zero as expected, and the "power" generally improves in Tables 3 and 4.

The finite sample results in Tables 1-4 are striking, demonstrating that a persistent conditional mean of asset returns can be identified (with good power) through a multivariate rank test, even if standard univariate techniques suggest serial dependence is absent. In particular, the tuning parameter selections $\kappa_G = 0.40$ and $\varkappa = 0.20$ balance "size" and "power" well. Hence, relying on these choices, we next seek to augment the LCM procedure with a rank test to determine the memory of the conditional mean return, as described in Section 3.2. Specifically, to mirror the 2-step testing procedure, if the test indicates full rank, when restricting $\widehat{d}_1 = \widehat{d}_2$, with $\widehat{d}_2$ computed by the TELW estimator, the return series has *not* been over-differenced and the restricted memory estimate is maintained. In contrast, if the rank test indicates cointegration, we test for over-differencing by simulating a fractional noise process, as for x_{t-1} in equation (25), but independent of y_t, and perform a rank test with $\widehat{e}_t$ and the filtered residuals from the auxiliary variable with, naturally, its own estimated integration order. If this test maintains full rank, the estimated series $\widehat{e}_t$ has *not* been over-differenced and we use the restricted memory estimate. If the test, once again, rejects full rank, the series has been over-differenced, and we implement the LCM procedure with the TELW estimate for $\widehat{d}_1$.[20]

[19] While Tables 1-2 report false rejection rates for the null hypothesis, akin to test size, the rank selection test is not calibrated to generate any specific rejection rate asymptotically, so lower (false) rejection rates are simply better. Nonetheless, the usual tradeoff between size and power properties applies, so comparing the "size" tables with the "power" reflected in Tables 3-4 provides the best guide towards identifying desirable test configurations.

[20] It is important to note that even though we only consider one candidate predictor and one auxilliary variable, this mimics

Once the rank-augmented estimate of $\widehat{d}_1$ has been obtained, the remaining filtering and estimation steps of the LCM procedure are the same. In Figure 3, we compare the bias and RMSE of this rank-augmented LCM (RLCM) procedure to the oracle implementation of LCM, thus mirroring the structure of Figure 2. The results, however, are in stark contrast. Whereas the unrestricted LCM procedure suffers from a pronounced bias, when $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$, the RLCM procedure is unbiased and essentially as efficient as the oracle version of LCM, demonstrating that our rank-testing step successfully ameliorates the estimation errors from using a misspecified estimate of d_1. That is, the RLCM procedure overcomes the challenges introduced by the "low" signal-to-noise ratio for the conditional mean return, and our rank selection procedure (22)-(23) represents an effective technique for detecting "hidden" persistence in the conditional mean and determining its integration order.

5 Empirical Illustration: Forecasting Equity Market Returns

This section analyzes predictions of monthly S&P 500 index returns via persistent and popular state variables in the macro-finance literature using OLS, IVX and RLCM procedures. Specifically, we examine the predictive content of the regressors in Bansal et al. (2014) and Campbell et al. (2018).

5.1 Data Description

We employ a data set of monthly observations for log-returns and corresponding realized variance (RV) measures of the aggregate U.S. stock market, proxied by the S&P 500 index, spanning the period from March 1960 through March 2015, which amounts to $n = 661$ observations. Specifically, following, e.g., Andersen & Bollerslev (1998), Barndorff-Nielsen & Shephard (2002), and Andersen, Bollerslev, Diebold & Labys (2003), RV is constructed by summing up daily squared returns for each month. Moreover, inspired by the VAR system in Campbell et al. (2018), we include the default spread (DS), three-month U.S. Treasury bills (TB), and price-earnings ratio (PE) as additional state variables. They have all been argued to be successful predictors of equity index returns, see, e.g., Lettau & Ludvigson (2010) and Campbell (2018, Chapters 5.3-5.4). The construction of these variables follows literature standards, with the DS being defined as the difference between logarithmic percentage yields on Moody's BAA and AAA bonds, TB is log-transformed, and PE is constructed as the log-ratio of the S&P 500 index to the ten-year trailing moving average of the aggregate S&P 500 constituent earnings. The DS and TB data are obtained from the website of the Federal Reserve Bank of St. Louis, while the PE data stem from Robert Shiller's website, see Shiller (2000).

5.2 RLCM Analysis of Return Predictability

First, we estimate the fractional integration order of returns and the four state variables; RV, DS, TB and PE. Specifically, we adopt the TELW estimator of Andersen & Varneskov (2020) and the

the two-step selection procedure more generally since we implement bivariate tests and only require one non-rejection of cointegration (for some $d_i > 0$) among all our regressors to conclude that the returns have *not* been overdifferenced.

exact local Whittle (ELW) estimator with a correction for the mean, or initial value, of Shimotsu & Phillips (2005) and Shimotsu (2010).[21] The results, reported in the top half of Table 5, show that the returns are, seemingly, $I(0)$, RV is stationary and fractionally integrated, and the remaining three state variables are non-stationary long-memory processes. However, as argued earlier, these results do not exclude returns from having a "latent" persistent conditional mean.

We proceed by implementing the sequential bivariate, LCM-based, rank selection procedure described in Section 3.2, using the TELW estimates from Table 5 and tuning parameters $(\nu_G, \kappa_G, \varkappa) = (0.20, 0.40, 0.20)$, as advocated in Section 4. The results are provided in the bottom half of Table 5. From Step 1 of the procedure, we find that $\widetilde{\mathcal{H}}_0 : r = 0$ is rejected for RV and PE. Hence, we select the larger fractional integration order for PE and continue with Step 2. Once restricting the memory in the second step, we maintain $\widetilde{\mathcal{H}}_0$ (i.e., full rank) for the three remaining predictors and, thus, stop there. These findings have striking implications. First, they provide consistent evidence that asset returns contain a fractionally integrated conditional mean, which we cannot detect using standard univariate time series techniques. Second, the conditional mean cointegrates with PE, suggesting that the latter is a "perfect" predictor of returns, as described in Section 2.2. Interestingly, this is consistent with dynamic present-value models for stock returns, e.g., Campbell (2018, Chapter 5.3), for which significance of standard predictability tests is often illusive; see, among others, Welch & Goyal (2008) and Lettau & Ludvigson (2010). Third, the rank selection procedure suggests that the remaining three candidate predictors are "imperfect", if at all significant. Fourth, this has implications for the statistical properties of our subsequent RLCM estimates, as demonstrated by Theorem 3. Specifically, it suggests that the limit theory for model (iv), with its super consistency properties, applies to PE, whereas the limit theory for models (ii)-(iii) may be used to test significance of the remaining predictors. We rely on these insights, when drawing feasible inference, as described in Appendix A.2.

Next, we estimate the coefficients of the four candidate predictors and test for their significance using OLS and IVX with Eicker-White inference, as described in Section 4, as well as our rank-augmented LCM (RLCM) procedure, where the integration order of the returns is fixed to that for PE.[22] The results are reported in Table 6. There are several interesting findings. First, using both OLS and IVX, we only find RV to contain statistically significant information about future returns. However, it has a negative coefficient, running counter to a traditional risk-return trade-off. In contrast, we find a positive predictive risk-return relation for RLCM, albeit insignificant. Second, using RLCM, we find significant predictability for both DS and PE. The positive sign for the former is consistent with a risk-return trade-off, and the negative for the latter reflects return-valuation theory (Campbell 2018). The sign of the corresponding coefficient estimates for OLS and IVX are similar, but the magnitudes are smaller, and the results are insignificant. Third, the significance for TB also improves using RLCM, but not sufficiently to render it a significant predictor at conventional significance levels.

These results are much sharper than typically obtained through return prediction studies, especially

[21] The TELW estimator, similarly to the mean-corrected ELW of Shimotsu (2010), is more robust to the mean, or initial value, of the process. Both estimators are valid for stationary and non-stationary fractionally integrated processes.

[22] We implement LCM with $\nu = 0.20$ and $\kappa = 0.60$ to reduce the impact from the contemporaneous return innovations.

at shorter horizons, e.g., Welch & Goyal (2008), Lettau & Ludvigson (2010), Campbell (2018, Chapter 5) and references therein. We attribute this to the various advantages of our RLCM procedure. First, uncovering the persistence of the conditional mean via rank testing, we may adequately filter returns, reducing the impact of the "large" contemporaneous innovations, thus mitigating its asymptotic and finite sample effect. Second, by letting $\ell/m + m/n \to 0$ as $n \to \infty$, we further reduce the impact from the error $\eta_t \in I(0)$ by sampling in a part of the spectrum, where the signal-to-noise ratio is larger. Finally, as discussed in Section 2.2, the (rank-augmented) LCM procedure is robust to endogenous innovations, which typically generate severe biases, e.g., Stambaugh (1999), Pastor & Stambaugh (2009) and Phillips & Lee (2013). These LCM features all alleviate critical attenuation biases, and we see from Table 6 that the coefficient estimates from RLCM is larger than those from OLS and IVX for DS, PE and TB. In contrast, in Andersen & Varneskov (2020), LCM is found to provide robust and reliable inference for return volatility forecasting, and to negate prior claims of auxiliary forecast power for a number of macro-finance variables. The critical difference across the applications is, that there are no evidence of latent unidentifiable integrated components in the return volatility series.

6 Conclusion

This paper studies the properties of predictive regressions for asset returns in economic systems governed by persistent vector autoregressive dynamics and considers robust estimation and inference. In particular, the dynamic properties of the state variables are captured by fractionally integrated processes, potentially of different orders, and returns have a latent persistent conditional mean, whose memory cannot be consistently estimated in finite samples. The latter feature is consistent with the typical findings in empirical studies, for which standard time series techniques almost invariably indicate only weak dependence in the return dynamics. We further allow for the regressors in the system to be endogenous and "imperfect". In this setting, we provide a cointegration rank test to determine the suitable predictive model framework as well as the latent persistence of the conditional mean return. By leveraging this additional source of information, we provide a rank-augmented LCM procedure, which is consistent and delivers asymptotic Gaussian inference. Simulations illustrate the theoretical arguments as well as favorable finite sample properties of the rank test and rank-augmented LCM procedure. Finally, in an empirical application to monthly S&P 500 return predictions, we find consistent evidence, that returns contain a (latent) fractionally integrated conditional mean component. Moreover, by applying the rank-augmented LCM procedure, we find strong predictive power for key economic state variables such as the price-earnings ratio and the default spread.

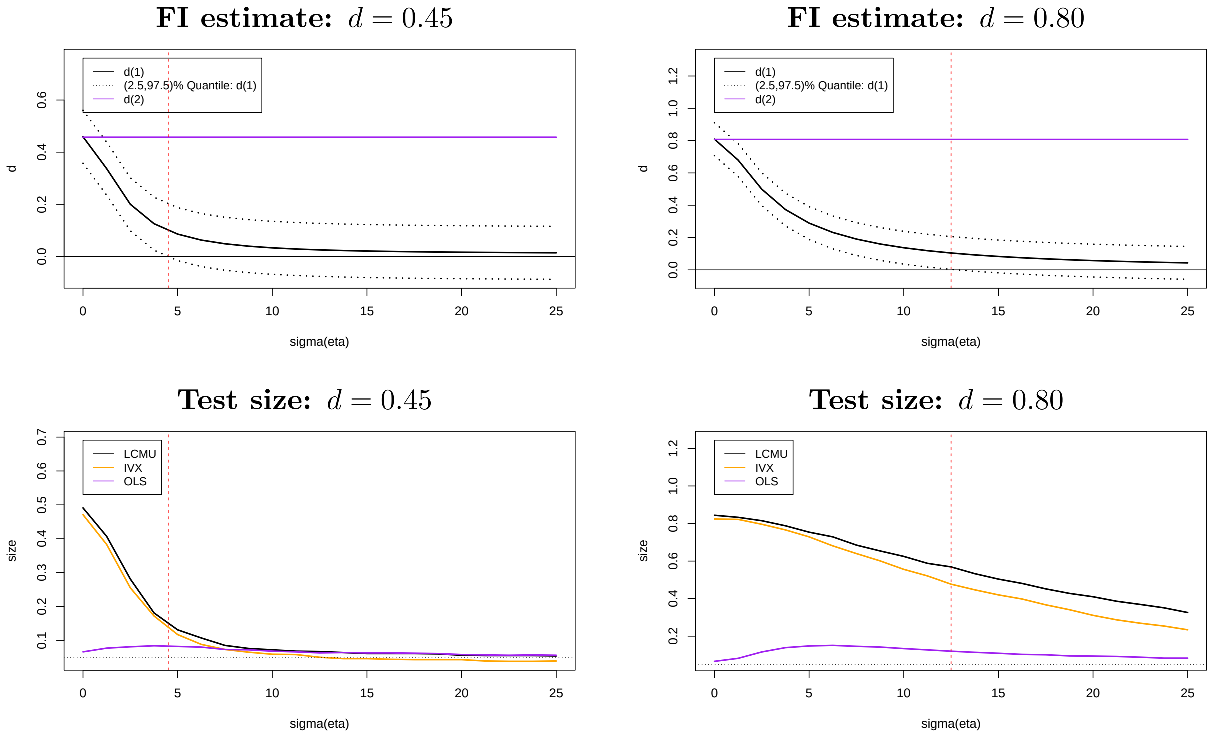

Figure 1: Fractional integration estimation and size. The top panels provide estimates of d_1 and d_2 as a function of the standard deviation of the weakly dependent return innovations, σ_η. Moreover, 95% confidence intervals are provided for $\widehat{d}_1$. The estimates are constructed using the TELW estimator with tuning parameters $\ell_d = \lfloor n^{0.3} \rfloor$ and $m_d = \lfloor n^{0.7} \rfloor$. The dotted vertical line highlights the value of σ_η where the empirical (unrestricted) estimate, $\widehat{d}_1$, is no longer significantly different from zero. The bottom panels provide the size of OLS, IVX and unrestricted LCM (LCMU) significance tests for $\beta = 0$. LCM is implemented with $(\nu, \kappa) = (0.2, 0.6)$ as well as $(\nu_G, \kappa_G) = (0.2, 0.6)$ for feasible inference; see Andersen & Varneskov (2020) and Appendix A.2 for details. Inference for OLS and IVX is drawn using Eicker-White standard errors. The left- and right-hand-side panels have $d = 0.45$ and $d = 0.80$. Finally, we consider a sample size $n = 650$, a 5% nominal test size and use 1000 replications.

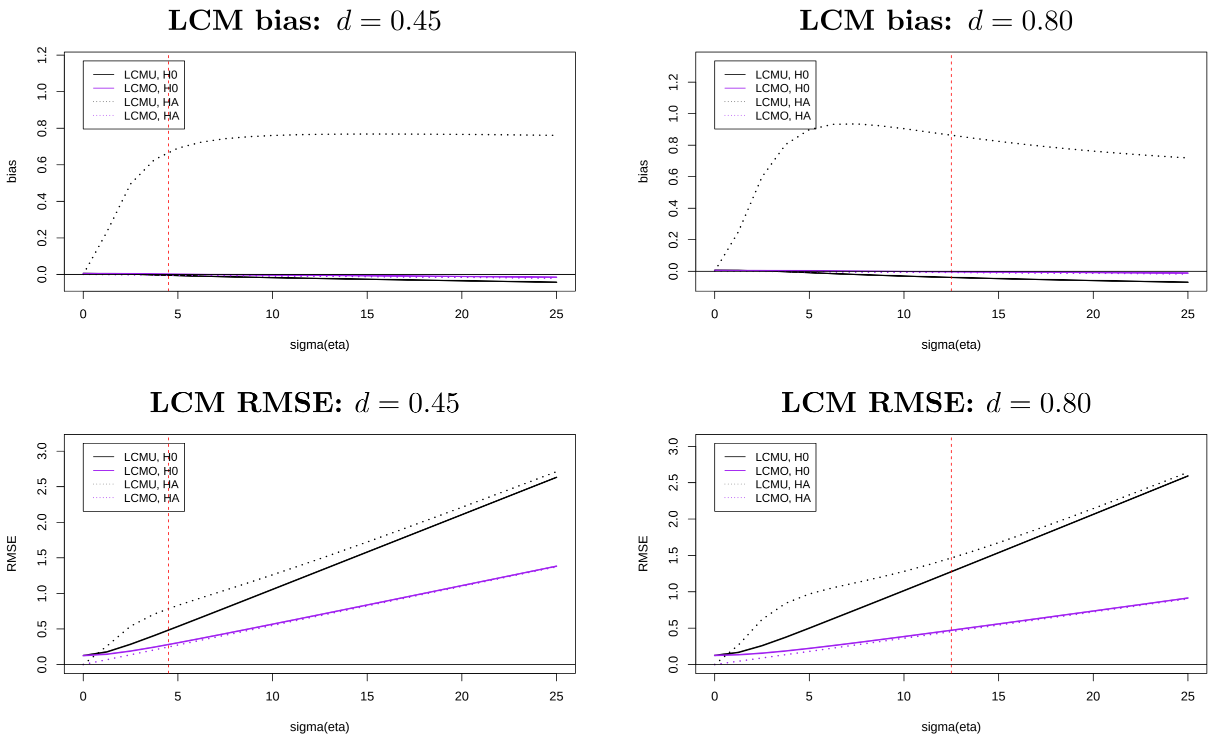

Figure 2: Bias and RMSE of LCM. The two top panels illustrate bias of LCM coefficient estimates in two scenarios when either $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$ $(\mathcal{H}_0)$ or $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$ $(\mathcal{H}_A)$ as a function of the standard deviation of the weakly dependent return innovations, σ_η. The two bottom panels provide corresponding RMSEs. Two versions of LCM is considered: An unrestricted LCM (LCMU), which uses the (biased) estimates $\widehat{d}_1$ and $\widehat{d}_2$ from Figure 1; an oracle LCM (LCMO), where $d_1 = d_2 = d$ is treated as known in the fractional filtering. Both versions of LCM are implemented with $(\nu, \kappa) = (0.2, 0.6)$. The left- and right-hand-side panels have $d = 0.45$ and $d = 0.80$, respectively. The dotted vertical line highlights the value of σ_η where the empirical (unrestricted) estimate, $\widehat{d}_1$, is no longer significantly different from zero. Finally, we consider a sample size $n = 650$ and use 1000 replications.

"Size" Properties of the Rank Test: $d = 0.45$									
Panel A	$\kappa_G = 0.40$, $\varkappa =$			$\kappa_G = 0.50$, $\varkappa =$			$\kappa_G = 0.60$, $\varkappa =$		
$\sigma_\eta =$	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
0	0.2780	0.0530	0.0080	0.0540	0.0010	0.0000	0.0010	0.0000	0.0000
2.5	0.2820	0.0530	0.0130	0.0660	0.0010	0.0000	0.0010	0.0000	0.0000
5	0.2870	0.0580	0.0100	0.0610	0.0010	0.0000	0.0020	0.0000	0.0000
7.5	0.3000	0.0530	0.0100	0.0570	0.0000	0.0000	0.0020	0.0000	0.0000
10	0.2980	0.0530	0.0110	0.0550	0.0000	0.0000	0.0020	0.0000	0.0000
12.5	0.2950	0.0560	0.0090	0.0550	0.0000	0.0000	0.0020	0.0000	0.0000
15	0.2960	0.0570	0.0090	0.0530	0.0000	0.0000	0.0010	0.0000	0.0000
17.5	0.2940	0.0600	0.0090	0.0520	0.0000	0.0000	0.0010	0.0000	0.0000
20	0.2970	0.0630	0.0090	0.0520	0.0000	0.0000	0.0010	0.0000	0.0000
22.5	0.2980	0.0620	0.0090	0.0530	0.0000	0.0000	0.0010	0.0000	0.0000
25	0.3020	0.0620	0.0090	0.0530	0.0000	0.0000	0.0010	0.0000	0.0000
Panel B	$\kappa_G = 0.40$, $\varkappa =$			$\kappa_G = 0.50$, $\varkappa =$			$\kappa_G = 0.60$, $\varkappa =$		
$\sigma_\eta =$	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
0	0.1360	0.0040	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000
2.5	0.1120	0.0050	0.0000	0.0030	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.1140	0.0100	0.0000	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000
7.5	0.1110	0.0090	0.0000	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.1120	0.0100	0.0000	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000
12.5	0.1160	0.0100	0.0000	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000
15	0.1180	0.0100	0.0000	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000
17.5	0.1230	0.0110	0.0000	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.1250	0.0110	0.0010	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000
22.5	0.1250	0.0100	0.0010	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000
25	0.1230	0.0110	0.0010	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000

Table 1: "Size" results. This tables show the frequency of rejecting full rank $\widetilde{\mathcal{H}}_0 : r = 0$ in favor of finding reduced rank $\widetilde{\mathcal{H}}_A : r = 1$ when $\widetilde{\mathcal{H}}_0$ is correct, using the LCM-based rank selection procedure in (22) and (23). This is in analogy with the size properties of a test. The memory of the system is $d = 0.45$, and the standard deviation of the weakly dependent return innovations, σ_η, is varied in $[0, 25]$. The rank test is implemented with the restricted estimate $\widehat{d}_2$ for both y_t and x_t, and the trimming rate $\nu_G = 0.20$ is fixed. Moreover, we consider tuning parameter selections $\kappa_G = \{0.4, 0.5, 0.6\}$ and $\varkappa = \{0.1, 0.2, 0.3\}$. The analysis uses two sample sizes, $n = 650$ and $n = 2000$, in Panels A and B, respectively, and 1000 replications. Finally, the dashed horizontal line highlights the value of σ_η in Figure 1, where the empirical (unrestricted) estimate, $\widehat{d}_1$, is no longer significantly different from zero.

"Size" Properties of the Rank Test: $d = 0.80$									
Panel A	$\kappa_G = 0.40$, $\varkappa =$			$\kappa_G = 0.50$, $\varkappa =$			$\kappa_G = 0.60$, $\varkappa =$		
$\sigma_\eta =$	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
0	0.2780	0.0530	0.0080	0.0540	0.0010	0.0000	0.0010	0.0000	0.0000
2.5	0.2860	0.0470	0.0070	0.0550	0.0020	0.0000	0.0030	0.0000	0.0000
5	0.2990	0.0560	0.0110	0.0530	0.0010	0.0000	0.0020	0.0000	0.0000
7.5	0.3030	0.0540	0.0090	0.0530	0.0010	0.0000	0.0020	0.0000	0.0000
10	0.2950	0.0580	0.0120	0.0540	0.0010	0.0000	0.0020	0.0000	0.0000
12.5	0.2970	0.0580	0.0130	0.0530	0.0010	0.0000	0.0020	0.0000	0.0000
15	0.2960	0.0620	0.0100	0.0560	0.0000	0.0000	0.0000	0.0000	0.0000
17.5	0.3010	0.0620	0.0110	0.0560	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.3000	0.0650	0.0110	0.0570	0.0000	0.0000	0.0000	0.0000	0.0000
22.5	0.3010	0.0630	0.0090	0.0560	0.0000	0.0000	0.0000	0.0000	0.0000
25	0.3030	0.0650	0.0080	0.0560	0.0000	0.0000	0.0000	0.0000	0.0000
Panel B	$\kappa_G = 0.40$, $\varkappa =$			$\kappa_G = 0.50$, $\varkappa =$			$\kappa_G = 0.60$, $\varkappa =$		
$\sigma_\eta =$	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
0	0.1360	0.0040	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000
2.5	0.1290	0.0030	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.1270	0.0050	0.0000	0.0030	0.0000	0.0000	0.0000	0.0000	0.0000
7.5	0.1220	0.0050	0.0000	0.0030	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.1200	0.0070	0.0000	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000
12.5	0.1090	0.0070	0.0010	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000
15	0.1050	0.0070	0.0010	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000
17.5	0.1050	0.0060	0.0010	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.1080	0.0060	0.0010	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000
22.5	0.1080	0.0070	0.0010	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000
25	0.1070	0.0070	0.0010	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2: "Size" results. This tables show the frequency of rejecting full rank $\widetilde{\mathcal{H}}_0 : r = 0$ in favor of finding reduced rank $\widetilde{\mathcal{H}}_A : r = 1$ when $\widetilde{\mathcal{H}}_0$ is correct, using the LCM-based rank selection procedure in (22) and (23). This is in analogy with the size properties of a test. The memory of the system is $d = 0.80$, and the standard deviation of the weakly dependent return innovations, σ_η, is varied in $[0, 25]$. The rank test is implemented with the restricted estimate $\widehat{d}_2$ for both y_t and x_t, and the trimming rate $\nu_G = 0.20$ is fixed. Moreover, we consider tuning parameter selections $\kappa_G = \{0.4, 0.5, 0.6\}$ and $\varkappa = \{0.1, 0.2, 0.3\}$. The analysis uses two sample sizes, $n = 650$ and $n = 2000$, in Panels A and B, respectively, and 1000 replications. Finally, the dashed horizontal line highlights the value of σ_η in Figure 1, where the empirical (unrestricted) estimate, $\widehat{d}_1$, is no longer significantly different from zero.

“Power” Properties of the Rank Test: $d = 0.45$									
Panel A	$\kappa_G = 0.40$, $\varkappa =$			$\kappa_G = 0.50$, $\varkappa =$			$\kappa_G = 0.60$, $\varkappa =$		
$\sigma_\eta =$	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2.5	1.0000	1.0000	0.9990	1.0000	1.0000	0.9760	1.0000	0.9800	0.3950
5	0.9850	0.8930	0.6910	0.9720	0.6530	0.1370	0.8470	0.0630	0.0000
7.5	0.8740	0.6140	0.2990	0.7760	0.1950	0.0120	0.3650	0.0010	0.0000
10	0.7370	0.4070	0.1400	0.5450	0.0680	0.0030	0.1410	0.0000	0.0000
12.5	0.6260	0.2670	0.0940	0.3830	0.0280	0.0020	0.0690	0.0000	0.0000
15	0.5460	0.1980	0.0620	0.2710	0.0160	0.0000	0.0390	0.0000	0.0000
17.5	0.5040	0.1540	0.0460	0.2300	0.0110	0.0000	0.0230	0.0000	0.0000
20	0.4480	0.1340	0.0350	0.1950	0.0090	0.0000	0.0190	0.0000	0.0000
22.5	0.4220	0.1210	0.0270	0.1670	0.0080	0.0000	0.0150	0.0000	0.0000
25	0.3910	0.1100	0.0220	0.1490	0.0030	0.0000	0.0080	0.0000	0.0000
Panel B	$\kappa_G = 0.40$, $\varkappa =$			$\kappa_G = 0.50$, $\varkappa =$			$\kappa_G = 0.60$, $\varkappa =$		
$\sigma_\eta =$	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5030
5	1.0000	0.9930	0.9200	0.9990	0.8620	0.1480	0.9820	0.0110	0.0000
7.5	0.9890	0.8370	0.4440	0.9320	0.1930	0.0010	0.3830	0.0000	0.0000
10	0.9240	0.5540	0.1590	0.6620	0.0270	0.0000	0.0610	0.0000	0.0000
12.5	0.8000	0.3330	0.0620	0.4060	0.0060	0.0000	0.0100	0.0000	0.0000
15	0.6830	0.2200	0.0360	0.2590	0.0020	0.0000	0.0010	0.0000	0.0000
17.5	0.5740	0.1470	0.0190	0.1760	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.4900	0.1030	0.0120	0.1180	0.0000	0.0000	0.0000	0.0000	0.0000
22.5	0.4280	0.0770	0.0070	0.0960	0.0000	0.0000	0.0000	0.0000	0.0000
25	0.3810	0.0580	0.0050	0.0740	0.0000	0.0000	0.0000	0.0000	0.0000

Table 3: “Power” results. This tables show the frequency of rejecting full rank $\widetilde{\mathcal{H}}_0 : r = 0$ in favor of finding reduced rank $\widetilde{\mathcal{H}}_A : r = 1$ when $\widetilde{\mathcal{H}}_A$ is correct, using the LCM-based rank selection procedure in (22) and (23). This is in analogy with the power properties of a test. The memory of the system is $d = 0.45$, and the standard deviation of the weakly dependent return innovations, σ_η, is varied in $[0, 25]$. The rank test is implemented with the restricted estimate $\widehat{d}_2$ for both y_t and x_t, and the trimming rate $\nu_G = 0.20$ is fixed. Moreover, we consider tuning parameter selections $\kappa_G = \{0.4, 0.5, 0.6\}$ and $\varkappa = \{0.1, 0.2, 0.3\}$. The analysis uses two sample sizes, $n = 650$ and $n = 2000$, in Panels A and B, respectively, and 1000 replications. Finally, the dashed horizontal line highlights the value of σ_η in Figure 1, where the empirical (unrestricted) estimate, $\widehat{d}_1$, is no longer significantly different from zero.

“Power” Properties of the Rank Test: $d = 0.80$									
Panel A	$\kappa_G = 0.40$, $\varkappa =$			$\kappa_G = 0.50$, $\varkappa =$			$\kappa_G = 0.60$, $\varkappa =$		
$\sigma_\eta =$	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9960
5	1.0000	1.0000	1.0000	1.0000	1.0000	0.9560	1.0000	0.6670	0.0320
7.5	1.0000	0.9920	0.9620	0.9980	0.8850	0.4240	0.8590	0.0640	0.0000
10	0.9890	0.9500	0.7860	0.9630	0.5560	0.0900	0.5330	0.0030	0.0000
12.5	0.9700	0.8340	0.5820	0.8550	0.3070	0.0260	0.2740	0.0000	0.0000
15	0.9230	0.7120	0.4040	0.7310	0.1560	0.0070	0.1570	0.0000	0.0000
17.5	0.8570	0.5870	0.2910	0.6030	0.0740	0.0040	0.0920	0.0000	0.0000
20	0.7990	0.4860	0.1980	0.5020	0.0510	0.0010	0.0580	0.0000	0.0000
22.5	0.7400	0.4070	0.1430	0.4310	0.0350	0.0010	0.0420	0.0000	0.0000
25	0.6900	0.3450	0.1130	0.3580	0.0240	0.0000	0.0290	0.0000	0.0000
Panel B	$\kappa_G = 0.40$, $\varkappa =$			$\kappa_G = 0.50$, $\varkappa =$			$\kappa_G = 0.60$, $\varkappa =$		
$\sigma_\eta =$	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9950	0.1370
7.5	1.0000	1.0000	1.0000	1.0000	1.0000	0.9430	0.9980	0.1800	0.0000
10	1.0000	0.9990	0.9960	1.0000	0.9710	0.3960	0.9000	0.0000	0.0000
12.5	1.0000	0.9970	0.9760	0.9980	0.7380	0.0820	0.5080	0.0000	0.0000
15	0.9990	0.9900	0.8850	0.9880	0.3870	0.0190	0.2120	0.0000	0.0000
17.5	0.9970	0.9640	0.7600	0.9240	0.1870	0.0020	0.0740	0.0000	0.0000
20	0.9920	0.8890	0.5780	0.8300	0.0870	0.0000	0.0190	0.0000	0.0000
22.5	0.9830	0.8210	0.4490	0.6940	0.0420	0.0000	0.0050	0.0000	0.0000
25	0.9700	0.7330	0.2930	0.5640	0.0220	0.0000	0.0030	0.0000	0.0000

Table 4: “Power” results. This tables show the frequency of rejecting full rank $\widetilde{\mathcal{H}}_0 : r = 0$ in favor of finding reduced rank $\widetilde{\mathcal{H}}_A : r = 1$ when $\widetilde{\mathcal{H}}_A$ is correct, using the LCM-based rank selection procedure in (22) and (23). This is in analogy with the power properties of a test. The memory of the system is $d = 0.80$, and the standard deviation of the weakly dependent return innovations, σ_η, is varied in $[0, 25]$. The rank test is implemented with the restricted estimate $\widehat{d}_2$ for both y_t and x_t, and the trimming rate $\nu_G = 0.20$ is fixed. Moreover, we consider tuning parameter selections $\kappa_G = \{0.4, 0.5, 0.6\}$ and $\varkappa = \{0.1, 0.2, 0.3\}$. The analysis uses two sample sizes, $n = 650$ and $n = 2000$, in Panels A and B, respectively, and 1000 replications. Finally, the dashed horizontal line highlights the value of σ_η in Figure 1, where the empirical (unrestricted) estimate, $\widehat{d}_1$, is no longer significantly different from zero.

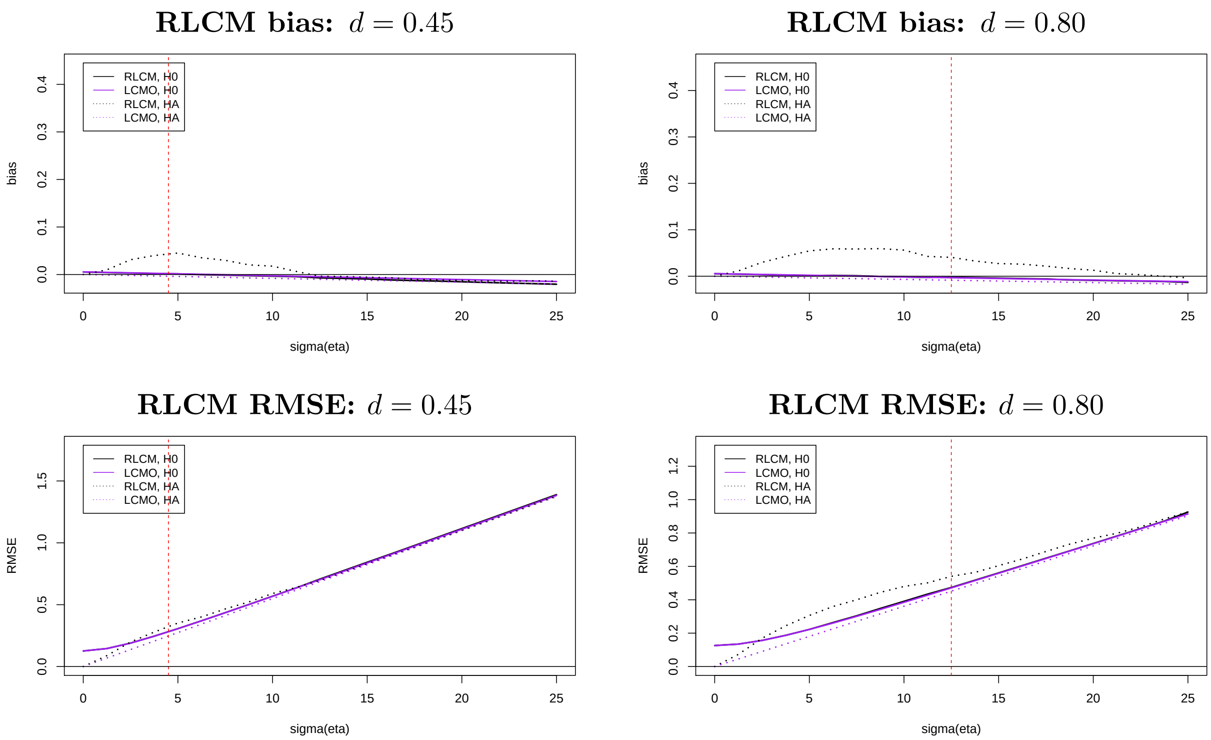

Figure 3: Bias and RMSE of RLCM. The two top panels illustrate bias of LCM coefficient estimates in two scenarios when either $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$ $(\mathcal{H}_0)$ or $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$ $(\mathcal{H}_A)$ as a function of the standard deviation of the weakly dependent return innovations, σ_η. The two bottom panels provide corresponding RMSEs. Two versions of LCM is considered: A rank-augmented LCM (RLCM), which uses a rank-test based estimate of $\widehat{d}_1$ as well as the TELW estimate $\widehat{d}_2$ from Figure 1; an oracle LCM (LCMO), where $d_1 = d_2 = d$ is treated as known in the fractional filtering. Both versions of LCM use $(\nu, \kappa) = (0.2, 0.6)$. The left- and right-hand-side panels have $d = 0.45$ and $d = 0.80$. The dotted vertical line highlights the value of σ_η where the empirical (unrestricted) estimate, $\widehat{d}_1$, is no longer significantly different from zero. Finally, we consider a sample size $n = 650$ and use 1000 replications.

Temporal Dependence and Rank					
	Returns_t	RV_{t-1}	DS_{t-1}	PE_{t-1}	TB_{t-1}
Mean	0.0055	0.0021	0.1346	0.0290	0.0468
Std. Dev.	0.0429	0.0047	0.0567	0.0042	0.0294
Skewness	-0.6833	11.0620	2.4834	-0.3293	0.5800
Kurtosis	5.5338	161.5564	12.5860	2.5694	3.7378
ACF(1)	0.0527	0.4287	0.9663	0.9955	0.9887
TELW	0.0631 (0.0516)	0.2853 (0.0516)	0.9405 (0.0516)	1.0290 (0.0516)	0.9188 (0.0516)
ELWM	0.0662 (0.0516)	0.2882 (0.0516)	0.8668 (0.0516)	1.1107 (0.0516)	0.9019 (0.0516)
$\mathcal{L}(\varrho = 0)$-unr	-	-0.8026	-0.8026	-0.8026	-0.8026
$\mathcal{L}(\varrho = 1)$-unr	-	-0.9305	-0.6227	-0.8088	-0.5697
$\widehat{r}$	-	1	0	1	0
$\mathcal{L}(\varrho = 0)$-res	-	-0.8026	-0.8026	-0.8026	-0.8026
$\mathcal{L}(\varrho = 1)$-res	-	-0.5473	-0.6079	-0.8088	-0.6252
$\widehat{r}$	-	0	0	1	0

Table 5: Descriptive statistics. This table displays statistics describing the unconditional and temporal dependence properties of returns and the four candidate predictors: RV, DS, PE and TB. Specifically, for the latter, we provide estimates of the first-order autocorrelation function (ACF), trimmed exact local Whittle (TELW) estimator of the fractional integration order (Andersen & Varneskov 2020) as well as exact local Whittle (ELWM) estimates with correction for the mean, or initial value, (Shimotsu 2010). The ELW estimators are implemented with bandwidth $\lfloor n^{0.7} \rfloor$ and, for TELW, trimming $\lfloor n^{0.1} \rfloor$ to reduce sensitivity to the mean. Moreover, LCM rank tests uses $(\kappa_G, \varkappa) = (0.4, 0.2)$. The rank tests are implemented using the individual TELW-estimated integration orders of the predictors (unr) and the integration order for PE in all tests (res), following the two-step procedure in Section 3.2. Finally, the sample of monthly observations spans March 1960 through March 2015 ($n = 661$).

RLCM Analysis of Return Predictions						
	RV_{t-1}			DS_{t-1}		
	OLS	IVX	RLCM	OLS	IVX	RLCM
$\widehat{\mathcal{B}}_c$	-1.1252	-1.1638	0.3257	0.0236	0.0200	0.1407
Wald	8.3130	8.2154	1.1619	0.3203	0.2326	5.7476
$\mathbb{P}$-Wald	0.0039	0.0042	0.2811	0.5714	0.6296	0.0165
	PE_{t-1}			TB_{t-1}		
	OLS	IVX	RLCM	OLS	IVX	RLCM
$\widehat{\mathcal{B}}_c$	-0.4956	-0.3987	-1.1421	-0.0280	-0.0277	-0.3953
Wald	1.3162	0.8460	12.4072	0.2218	0.2098	2.1507
$\mathbb{P}$-Wald	0.2513	0.3577	0.0004	0.6377	0.6469	0.1425

Table 6: Return predictions. This table provides coefficient estimates and significance tests based on Wald statistics (and associated $\mathbb{P}$-values) for three different methodologies; OLS, IVX and RLCM, where the fractional integration order of returns have been restricted to that of PE. The integration orders required for the fractional filtering procedure are provided by the TELW estimates in Table 5. The LCM procedure and feasible inference (cf., Appendix A.2) are implemented with $\nu = \nu_G = 0.2$ and $\kappa = \kappa_G = 0.6$. Inference for OLS and IVX employs Eicker-White standard errors. IVX is implemented with two instruments as in Section 4. Finally, the sample of monthly observations spans March 1960 through March 2015 ($n = 661$).

References

Andersen, T. G. & Benzoni, L. (2012), Stochastic volatility, *in* R. A. Meyers, ed., 'Encyclopedia of Complexity and Systems Science', Springer-Verlag. forthcoming.

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. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Andersen, T. G. & Varneskov, R. T. (2020), 'Consistent inference for predictive regressions in persistent economic systems', *Journal of Econometrics* **forthcoming**.

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

Bansal, R. & Yaron, A. (2004), 'Risks for the long run: A potential resolution of asset pricing puzzles', *Journal of Finance* **59**, 1481–1509.

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.

Bauer, D. & Maynard, A. (2012), 'Persistence-robust surplus-lag granger causality testing', *Journal of Econometrics* **169**, 293–300.

Boudoukh, J., Richardson, M. & Whitelaw, R. F. (2008), 'The myth of long-horizon predictability', *Review of Financial Studies* **21**(4), 1577–1605.

Breitung, J. & Demetrescu, M. (2015), 'Instrumental variable and variable addition based inference in predictive regressions', *Journal of Econometrics* **187**, 358–375.

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

Campbell, J. Y. (2018), *Financial Decisions and Markets: A Course in Asset Pricing*, Princeton University Press.

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. & 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.

Choi, I. (1993), 'Asymptotic normality of the least-squares estimates for higher order autoregressive integrated processes with some applications', *Econometric Theory* **9**, 263–282.

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.

Dolado, J. J. & Lütkepohl, H. (1996), 'Making wald tests work for cointegrated var systems', *Econometric Reviews* **15**, 369–386.

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

Fama, E. F. (1970), 'Efficient capital markets: A review of theory and empirical work', *Journal of Finance* **25**, 383–417.

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

Gabaix, X. (2012), 'Variable rare disasters: An exactly solved framework for ten puzzles in macro finance', *Quarterly Journal of Economics* **127**, 645–700.

Georgiev, I., Harvey, D. I., Leybourne, S. J. & Taylor, A. R. (2020), 'A bootstrap stationarity test for predictive regression invalidity', *Journal of Business & Economic Statistics* **forthcoming**.

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

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

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.

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

Kendall, M. (1954), 'Note on bias in the estimation of autocorrelation', *Biometrika* **41**, 403–404.

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

Lettau, M. & Ludvigson, S. (2001), 'Consumption, aggregate wealth, and expected stock returns', *Journal of Finance* **56**, 815–849.

Lettau, M. & Ludvigson, S. (2010), Measuring and modeling variation in risk-return trade-off, *in* Y. Aït-Sahalia & L. P. Hansen, eds, 'Handbook of Financial Econometrics', Elsevier Science B. V., North Holland, Amsterdam.

Lettau, M., Ludvigson, S. & Wachter, J. (2008), 'The declining equity premium: What role does macroeconomic risk play?', *Review of Financial Studies* **21**, 1653–1687.

Lin, Y. & Tu, Y. (2020), 'Robust inference for spurious regressions and cointegrations involving processes moderately deviated from a unit root', *Journal of Econometrics* **forthcoming**.

Liu, X., Yang, B., Cai, Z. & Peng, L. (2019), 'A unified test for predictability of asset returns regardless of properties of predicting variables', *Journal of Econometrics* **208**, 141–159.

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.

Marriott, F. & Pope, J. (1954), 'Bias in the estimation of autocorrelations', *Biometrika* **41**, 390–402.

Mikusheva, A. (2007), 'Uniform inference in autoregressive models', *Econometrica* **75**, 1411–1452.

Neuhierl, A. & Varneskov, R. T. (2020), 'Frequency dependent risk', *Journal of Financial Economics* **forthcoming**.

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

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. & Shimotsu, K. (2007), 'Determining the cointegrating rank in nonstationary fractional systems by the exact local whittle approach', *Journal of Econometrics* **141**, 574–596.

Ortu, F., Tamoni, A. & Tebaldi, C. (2013), 'Long-run risk and the persistence of consumption shocks', *Review of Financial Studies* **26**, 2876–2915.

Park, J. Y. & Phillips, P. C. B. (1989), 'Statistical inference in regressions with integrated processes: Part 2', *Econometric Theory* **5**, 95–131.

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

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

Phillips, P. C. B. (1987), 'Towards a unified asymptotic theory for autoregression', *Biometrika* **74**, 535–547.

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. B. & Magdalinos, T. (2007), 'Limit theory for moderate deviations from a unit root', *Journal of Econometrics* **136**, 115–130.

Phillips, P. C. B. & Ouliaris, S. (1988), 'Testing for cointegration using principal component methods', *Journal of Economic Dynamics and Control* **12**, 205–230.

Ren, Y., Tu, Y. & Yi, Y. (2019), 'Balanced predictive regressions', *Journal of Empirical Finance* **54**, 118–142.

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

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

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.

Robinson, P. M. & Yajima, Y. (2002), 'Determination of cointegrating rank in fractional systems', *Journal of Econometrics* **106**, 217–241.

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

Shenton, L. & Johnson, W. (1965), 'Moments of a serial correlation coefficient', *Journal of the Royal Statistical Society, Series B* **27**, 308–320.

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

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

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

Sims, C. A., Stock, J. H. & Watson, M. W. (1990), 'Inference in linear time series models with some unit roots', *Econometrica* **58**, 113–144.

Stambaugh, R. F. (1986), Bias in regressions with lagged stochastic regressors. Working Paper 156, CRSP, Graduate School of Business, University of Chicago.

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

Toda, H. & Yamamoto, T. (1995), 'Statistical inference in vector autoregressions with possibly integrated processes', *Journal of Econometrics* **66**, 225–250.

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. (2017), 'Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels', *Econometric Theory* **33**(6), 1457–1501.

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

Xu, K.-L. (2020), Testing for return predictability with co-moving predictors of unknown form. Unpublished manuscript, Department of Economics, Indiana University.

A Additional Theory and Feasible Inference

This section presents an alternative rank test and describes how to draw feasible inference.

A.1 Alternative Cointegration Rank Test

The eigenvalues δ_i and $\hat{\delta}_i$, $i = 1, \ldots, k+1$, defined in main text, may be utilized to design a cointegration rank test of the null hypothesis $\widetilde{\mathcal{H}}_0$ against $\widetilde{\mathcal{H}}_A$ based on the ratio statistics

$$\theta_{k+1} = \frac{\delta_{k+1}}{\sum_{i=1}^{k+1} \delta_i}, \quad \widehat{\theta}_{k+1} = \frac{\hat{\delta}_{k+1}}{\sum_{i=1}^{k+1} \hat{\delta}_i}, \quad \widehat{\Theta}_{k+1}^2 = \frac{\hat{\delta}_{k+1}^2 \sum_{i=1}^{k} \hat{\delta}_i^2 + \hat{\delta}_{k+1}^2 (\sum_{i=1}^{k} \hat{\delta}_i)^2}{(\sum_{i=1}^{k+1} \hat{\delta}_i)^4}, \tag{A.1}$$

whose asymptotic properties under $\widetilde{\mathcal{H}}_0$ follow from Theorem 1 and the delta method.

Theorem 4. *Under the conditions of Theorem 1 and $\widetilde{\mathcal{H}}_0$,*

$$m_G^{1/2} \left(\widehat{\vartheta}_{k+1} - \vartheta_{k+1} \right) / \widehat{\Theta}_{k+1} \xrightarrow{\mathbb{D}} N(0,1).$$

As noted by Phillips & Ouliaris (1988) and Robinson & Yajima (2002) in (fractional) cointegration contexts, Theorem 4 relies on $\widetilde{\mathcal{H}}_0$ and cannot be used to test against $\widetilde{\mathcal{H}}_A$, because the distribution becomes degenerate for $\delta_{k+1} \to 0$. However, for testing whether that the cointegration rank is unity, that is, against $\widetilde{\mathcal{H}}_A$, we may use the $100(1-\alpha)\%$ upper confidence interval for $\widehat{\theta}_{k+1}$,

$$\mathcal{CI}(\alpha, k+1) = \widehat{\theta}_{k+1} + \widehat{\Theta}_{k+1} Q(\alpha) / m_G^{1/2}, \tag{A.2}$$

with $Q(\alpha)$ being the $(1-\alpha)$th quantile of the standard Gaussian distribution, and compare it to a pre-specified threshold, as suggested by Phillips & Ouliaris (1988). Following their recommendation, and motivated by the numerical results in Nielsen & Shimotsu (2007), we have applied the test with a threshold $0.1/(k+1)$ in the simulation study in Section 4. However, the test is strictly dominated by the selection procedure based on (22) and (23). These results are left our for brevity.

A.2 Feasible Inference

When drawing feasible inference with the LCM approach, we must provide consistent estimators of the long-run covariance matrix $\boldsymbol{G}_{uu}$ and either $G_{\xi\xi}$ or $G_{\eta\eta}$, depending on whether we are drawing inference for models (ii)-(iii) or the cointegration model (iv). To this end, we need information from the residuals, after estimation of $\boldsymbol{\mathcal{B}}$. The main challenge is, again, that we observe $\hat{\boldsymbol{v}}_t^c$, not $\hat{\boldsymbol{v}}_t$ nor $\boldsymbol{v}_t$. As a result, the residuals η_t are latent, and we need to estimate them as,

$$\widehat{\eta}_t^c = (1-L)^{-\widehat{b}}\, \widehat{\eta}_t^{(b,c)}, \qquad \widehat{\eta}_t^{(b,c)} \; = \; \widehat{e}_t \; - \; \widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m)' \widehat{\boldsymbol{u}}_{t-1}^c, \tag{A.3}$$

where $\widehat{b}$ is some consistent estimator of b. Similarly, we can define $\widehat{\eta}_t$ and $\widehat{\eta}_t^{(b)}$ as in (A.3), but computed with (the unobservable) $\widehat{\boldsymbol{\mathcal{B}}}(\ell, m)$ and $\widehat{\boldsymbol{u}}_{t-1}$ and, thus, free of regressor endogeneity bias.

Despite the notation, it is important to note that the estimators $\widehat{\eta}_t^c$ and $\widehat{\eta}_t^{(b,c)}$ can be utilized to estimate both $G_{\xi\xi}$ in models (ii) and (iii) as well as $G_{\eta\eta}$ in (iv). Specifically, when drawing inference for the former, we estimate the variance with $\widehat{\eta}_t^c = \widehat{\eta}_t^{(b,c)}$, where the contribution from ξ_{t-1} will dominate that from $\eta_t^{(b)}$, $b = d_1 > 0$, asymptotically. In contrast, for model (iv), we use (A.3) the consistent estimate $\widehat{b} = \widehat{d}_1$, following Assumption M, since $\xi_{t-1} = 0$, $\forall t = 1, \dots n$. Once the estimate $\widehat{\eta}_t^c$ is computed for a given inference scenario, we, then, use the trimmed long-run covariance estimators, described in Section 3 for the purpose of rank testing, and compute the asymptotic variance as,

$$\widehat{\mathbf{AVAR}} = \widehat{\boldsymbol{G}}^c_{\widehat{u}\widehat{u}}(\ell_G, m_G)^{-1} \begin{cases} \widehat{G}^c_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G)/(2m), & \text{under models (ii) and (iii)}, \\ \widehat{G}^c_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G)\lambda_m^{2\widehat{b}}/\big(2(1+2\widehat{b})m\big), & \text{under model (iv)}. \end{cases} \tag{A.4}$$

with, again, $m_G = m_G(n)$ and $\ell_G = \ell_G(n)$. Specifically, we determine which of the inference scenarios to apply by means of the rank selection procedure in (22) and (23).

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

Assumption T-B. *Define $\underline{m}_n = m_d \wedge m_b \wedge m$ and $\bar{g}_n(m, m_b, m_d) = \frac{\ln(n)}{\sqrt{m_b}} \vee \frac{\ln(n)}{\sqrt{m_d}} \vee \frac{\lambda_m^b}{\sqrt{m}}$, then the following cross-restrictions are imposed on the trimming and bandwidth parameters:*

$$\frac{\sqrt{n}}{m_G} + \frac{m}{m_G} + \frac{m_G}{\ell_G\,\ell} + \bar{g}_n(m, m_b, m_d)\left(\frac{n}{m_G}\right)^b \to 0, \qquad as \quad n \to \infty.$$

Finally, additional conditions are imposed, when $n \to \infty$, depending on the model:

$$\begin{cases} (m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon} + (m/n)^{\underline{d}_x}/\ell^{1+\epsilon} \to 0, & \textit{under models (ii) and (iii)}, \\ (m_G/n)^{\underline{d}-b}/\ell_G^{1+\epsilon} \to 0, & \textit{under model (iv)}. \end{cases}$$

Assumption B is similar to Andersen & Varneskov (2020, Assumption B), imposing a mild consistency requirement on the estimator of the (fractional) integration order of the residuals. Moreover, Assumption T-B imposes additional (mild) conditions on the trimming and bandwidth parameters to eliminate the endogenous regressor bias when estimating the variance of the residuals. In addition to these, we invoke Assumption T-G for the covariance estimators. However, it is worth noting that the necessary conditions on the tuning parameters for consistency of the asymptotic variance estimator for feasible inference are milder than those stated in Assumption T-G. In particular, instead of the first three conditions, we require only $m_G \asymp n^{\kappa_G}$ and $\ell \asymp n^{\nu_G}$, with $0 < \nu_G < \kappa_G < \varrho \leq 1$ and $n/(m_G\ell_G^2) + n^2/(m_G\ell_G^2\underline{m}_n) \to 0$ as $n \to \infty$, the same as in Andersen & Varneskov (2020, Assumption T-G). Similarly, the equivalent condition four is a factor $1/\sqrt{m_G}$ smaller than the one stated. The reason is that we only need consistency for feasible inference, not a central limit theorem. However, we refrain from stating separate assumptions to distinguish the two cases.

Theorem 5. *Suppose Assumption B, T-B and the conditions of Theorems 1 and 3 hold, then*

$$\begin{cases} m\widehat{\mathbf{AVAR}} \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}^{-1}G_{\xi\xi}/2, & \textit{under models (ii) and (iii)}, \\ m\lambda_m^{-2b}\widehat{\mathbf{AVAR}} \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}^{-1}G_{\eta\eta}/(2(1+2b)), & \textit{under model (iv)}. \end{cases}$$

Feasible inference and testing for the LCM procedure, then, follows by applying Theorems 3 and 5 in conjunction with the continuous mapping theorem and Slutsky's theorem.

B Proofs

This section provides proofs of the main asymptotic results in the paper. Before proceeding, however, we introduce some notation. For a generic vector $\boldsymbol{V}$, let $\boldsymbol{V}(i)$ index the ith element, and, similarly, for a matrix $\boldsymbol{M}$, let $\boldsymbol{M}(i,q)$ denote its (i,q)th element. Moreover, $K \in (0,\infty)$ denotes a generic constant, which may take different values from line to line or from (in)equality to (in)equality. Sometimes the (stochastic) orders refer to scalars, sometimes to vectors and matrices. We refrain from making distinctions. Finally, we provide an auxiliary lemma that expands on Theorem 4 in Andersen & Varneskov (2020). We will henceforth refer to the latter as AV (2020) and, similarly, to their Online Appendix as AVOA (2020). Specifically, the lemma provides bounds for the differences,

$$\widehat{\boldsymbol{F}}^{c}_{\widehat{u}\widehat{u}}(\ell,m) - \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m), \quad \widehat{\boldsymbol{F}}^{c}_{\widehat{u}\widehat{e}}(\ell,m) - \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}(\ell,m), \tag{B.1}$$

$$\widehat{\boldsymbol{G}}^{c}_{\widehat{v}\widehat{v}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\widehat{v}\widehat{v}}(\ell_G,m_G), \quad \widehat{\boldsymbol{G}}^{c}_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G), \tag{B.2}$$

where, as described in Appendix A.2, $\widehat{\eta}^c_t$ constitutes an estimate of η_t that is employed when implementing feasible inference. In other words, we provide asymptotic bounds to describe the errors arising when using the fractionally filtered observations $\widehat{\boldsymbol{v}}^c_t$ rather than the unobservable $\widehat{\boldsymbol{v}}_t$ when calculating key measures and statistics, thus quantifying the impact of regressor endogeneity. The auxiliary lemma differs from AV (2020, Theorem 4) by allowing for cointegration, $b > 0$.

Lemma B.1. *Suppose Assumptions D1-D3, C, M, F, T-G, T hold. Moreover, suppose that the bandwidths satisfy $n^{1/2}/m \to 0$, $n^{1/2}/m_G \to 0$, then, for some arbitrarily small $\epsilon > 0$, it follows,*

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

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

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

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

(e) Suppose further Assumption B and T-B (instead of T-G) hold as well as $\underline{d} - b \ge 0$, then

$$\widehat{\boldsymbol{G}}^{c}_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G) \le \begin{cases} O_p((m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon}) + O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon}), & \text{for (ii)-(iii)}, \\ O_p\left((m_G/n)^{\underline{d}-b}/\ell_G^{1+\epsilon}\right), & \text{for (iv)}. \end{cases}$$

Proof. First, **(a)** and **(c)** follows directly from AV (2020, Theorems 4(a) and 4(c)), since the specification of the regressors in this paper readily follows their framework.[23]

[23] While AV (2020) state their results for $\underline{d}$ rather than $\underline{d}_x$ to maintain notational simplicity in their framework, it is clear

For **(b)**, let us first define $\widehat{e}_t = \widehat{e}_t^{(1)} + \widehat{e}_t^{(2)}$, where

$$\widehat{e}_t^{(1)} \equiv (1-L)^{\widehat{d}_1} a + \boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{Q}}(L)(1-L)^{\widehat{d}_1}\boldsymbol{x}_{t-1} + (1-L)^{\widehat{d}_1}\xi_{t-1}^{(-d_1)}, \quad \widehat{e}_t^{(2)} \equiv (1-L)^{\widehat{d}_1}\eta_t, \tag{B.3}$$

for which the component $\widehat{e}_t^{(1)}$ is equivalent to the case without cointegration considered by AV (2020, Theorem 4(b)) due to Assumptions D1-D3 and C. By applying the decomposition (B.3), we have

$$\widehat{\boldsymbol{F}}^c_{\widehat{u}\widehat{e}}(\ell,m) - \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}(\ell,m) = \widehat{\boldsymbol{F}}^c_{\widehat{c}\widehat{e}}(\ell,m) = \widehat{\boldsymbol{F}}^{(c,1)}_{\widehat{c}\widehat{e}}(\ell,m) + \widehat{\boldsymbol{F}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell,m) \tag{B.4}$$

where $\widehat{\boldsymbol{F}}^{(c,1)}_{\widehat{c}\widehat{e}}(\ell,m)$ and $\widehat{\boldsymbol{F}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell,m)$ are the TDAC between $\widehat{\boldsymbol{c}}_{t-1}$ and $\widehat{e}_t^{(1)}$, respectively, $\widehat{e}_t^{(2)}$. Now, by applying, AV (2020, Theorem 4(b)) and AVOA (2020, Lemma A.12(b)), we have,

$$\lambda_m^{-1}\widehat{\boldsymbol{F}}^{(c,1)}_{\widehat{c}\widehat{e}}(\ell,m) \le O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon}), \quad w_{\widehat{e}}^{(2)}(\lambda_j) = O_p(\lambda_j^{d_1}), \quad \boldsymbol{w}_{\widehat{c}}(\lambda_j,i) = O_p(\lambda_j^{d_i}), \tag{B.5}$$

for $i = 2,\dots,k+1$. Hence, since $0 < d_i \le d_1 + d_i$, $i = 2,\dots,k+1$, we may further write

$$\widehat{\boldsymbol{F}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell,m) \le \frac{2\pi}{n}\sum_{j=\ell}^{m} O_p(\lambda_j^{\underline{d}_x}) \le \frac{2\pi m^{1+\underline{d}_x}}{n^{1+\underline{d}_x}}\sum_{j=\ell}^{m} O_p\left(\left(\frac{j}{m}\right)^{\underline{d}_x}\frac{1}{j^{1+\epsilon}}\right) \le O_p\left(\left(\frac{m}{n}\right)^{1+\underline{d}_x}\frac{1}{\ell^{1+\epsilon}}\right), \tag{B.6}$$

for some arbitrarily small $\epsilon > 0$, using $|\sum_{j=\ell}^{m} O_p(j^{-p})| \le O_p(\ell^{-p})$ for some $p > 1$ by Varneskov (2017, Lemma C.4). The stated result follows by combining bounds for $\widehat{\boldsymbol{F}}^{(c,1)}_{\widehat{c}\widehat{e}}(\ell,m)$ and $\widehat{\boldsymbol{F}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell,m)$.

For **(d)**, by applying the same decomposition as for **(b)**, we have

$$\widehat{\boldsymbol{G}}^c_{\widehat{u}\widehat{e}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\widehat{u}\widehat{e}}(\ell_G,m_G) = \widehat{\boldsymbol{G}}^c_{\widehat{c}\widehat{e}}(\ell_G,m_G) = \widehat{\boldsymbol{G}}^{(c,1)}_{\widehat{c}\widehat{e}}(\ell_G,m_G) + \widehat{\boldsymbol{G}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell_G,m_G), \tag{B.7}$$

where, again, the DFT bounds in (B.5) apply to $w_{\widehat{e}}^{(2)}(\lambda_j)$ and $\boldsymbol{w}_{\widehat{c}}(\lambda_j,i)$. Moreover, by AV (2020, Theorem 4(c)), we have

$$\widehat{\boldsymbol{G}}^{(c,1)}_{\widehat{c}\widehat{e}}(\ell_G,m_G) \le O_p((m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon}). \tag{B.8}$$

Next, using, again, $0 < d_i \le d_1 + d_i$, $i = 2,\dots,k+1$, we may similarly write

$$\begin{aligned}\widehat{\boldsymbol{G}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell_G,m_G) &\le \frac{1}{m_G - \ell_G + 1}\sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{\underline{d}_x}) \\ &\le \frac{K m_G^{\underline{d}_x}}{n^{\underline{d}_x}}\sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{\underline{d}_x}\frac{1}{j^{1+\epsilon}}\right) \le O_p\left(\left(\frac{m_G}{n}\right)^{\underline{d}_x}\frac{1}{\ell_G^{1+\epsilon}}\right),\end{aligned} \tag{B.9}$$

using $m_G/(m_G - \ell_G + 1) \le K$ and Varneskov (2017, Lemma C.4). The stated result follows by combining asymptotic bounds for $\widehat{\boldsymbol{G}}^{(c,1)}_{\widehat{c}\widehat{e}}(\ell_G,m_G)$ and $\widehat{\boldsymbol{G}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell_G,m_G)$.

---

that their results apply to $\underline{d}_x$ as the parameter appears when applying the differencing operator to $\boldsymbol{u}_{t-1}$ and $\boldsymbol{c}_{t-1}$.

For **(e)**, let us first write $\widehat{\eta}_t^c \equiv (1-L)^{-\widehat{b}}\widehat{\eta}_t^{(b,c)}$, with

$$\widehat{\eta}_t^{(b,c)} = \widehat{e}_t - \widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m)'\widehat{\boldsymbol{u}}_{t-1}^c = \widehat{\eta}_t^{(b,1)} + \widehat{e}_t^{(2)} - \widehat{\tau}_{t-1}^{(1)} - \widehat{\tau}_{t-1}^{(2)} \tag{B.10}$$

using $\widehat{e}_t = \widehat{e}_t^{(1)} + \widehat{e}_t^{(2)}$, where we also let $\widehat{\eta}_t^{(b)} = \widehat{\eta}_t^{(b,1)} + \widehat{e}_t^{(2)}$,

$$\widehat{\eta}_t^{(b,1)} = \widehat{e}_t^{(1)} - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)'\widehat{\boldsymbol{u}}_{t-1}, \quad \widehat{\tau}_{t-1}^{(1)} = (\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m))'\widehat{\boldsymbol{u}}_{t-1}^c, \quad \widehat{\tau}_{t-1}^{(2)} = \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\widehat{\boldsymbol{c}}_{t-1}. \tag{B.11}$$

The main difference between this decomposition and the corresponding in AV (2020, Theorem 4) is the presence of $\widehat{e}_t^{(2)}$ and the fact that we may have $\widehat{b}, b \neq 0$. Hence, we need to distinguish between cases without cointegration $\widehat{b} = b = 0$, i.e., scenarios (ii) and (iii), as determined by the cointegration rank test, and scenario (iv), where $\widehat{b}$ satisfies Assumption B; see Appendix A.2.

The case without cointegration. Here, $\widehat{\eta}_t^c = \widehat{\eta}_t^{(0,c)}$, $\widehat{\eta}_t^{(0)} = \widehat{\eta}_t$ and let us make the decomposition,

$$\begin{aligned}\widehat{\boldsymbol{G}}^c_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G) &= \widehat{\boldsymbol{G}}^{(1,1)}_{\widehat{\tau}\widehat{\tau}}(\ell_G,m_G) + \widehat{\boldsymbol{G}}^{(2,2)}_{\widehat{\tau}\widehat{\tau}}(\ell_G,m_G) + 2\widehat{\boldsymbol{G}}^{(1,2)}_{\widehat{\tau}\widehat{\tau}}(\ell_G,m_G)\\ &\quad - 2\widehat{\boldsymbol{G}}^{(1)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G) - 2\widehat{\boldsymbol{G}}^{(2)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G),\end{aligned} \tag{B.12}$$

where the first three terms are the (trimmed) long-run variance and covariance estimates for $\widehat{\tau}_{t-1}^{(1)}$ and $\widehat{\tau}_{t-1}^{(2)}$, and the final two terms are the respective long-run covariances with $\widehat{\eta}_t$. Let us further write,

$$\widehat{\boldsymbol{G}}^{(i)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G) = \widehat{\boldsymbol{G}}^{(i,1)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G) + \widehat{\boldsymbol{G}}^{(i,2)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G), \quad i = 1,2, \tag{B.13}$$

to indicate the decomposition of $\widehat{\eta}_t$ into $\widehat{\eta}_t^{(0,1)} = \widehat{\eta}_t^{(1)}$ and $\widehat{e}_t^{(2)}$. Now, in this case, the asymptotic bounds for $\widehat{\boldsymbol{G}}^{(1,1)}_{\widehat{\tau}\widehat{\tau}}(\ell_G,m_G)$, $\widehat{\boldsymbol{G}}^{(2,2)}_{\widehat{\tau}\widehat{\tau}}(\ell_G,m_G)$, $\widehat{\boldsymbol{G}}^{(1,2)}_{\widehat{\tau}\widehat{\tau}}(\ell_G,m_G)$, $\widehat{\boldsymbol{G}}^{(1,1)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G)$, and $\widehat{\boldsymbol{G}}^{(1,2)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G)$ are derived in equations (A.26), (A.27), (A.29) and (A.30) of AVOA (2020) since $\widehat{\boldsymbol{G}}^{(2,2)}_{\widehat{\tau}\widehat{\tau}}(\ell_G,m_G) \leq O_p(\widehat{\boldsymbol{G}}^{(1,1)}_{\widehat{\tau}\widehat{\tau}}(\ell_G,m_G))$. Hence, to complete the proof, we need to establish corresponding asymptotic bounds for the terms $\widehat{\boldsymbol{G}}^{(2,1)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G)$ and $\widehat{\boldsymbol{G}}^{(2,2)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G)$, i.e., covariances with $\widehat{e}_t^{(2)}$. To this end, let us first use the bounds in (B.5), $\widehat{\boldsymbol{\mathcal{B}}}(\ell,m) = O_p(1)$, uniformly by AV (2020, Theorem 1), and $0 < d_i \leq d_1 + d_i$ to write,

$$\widehat{\boldsymbol{G}}^{(2,2)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G) \leq \frac{1}{m_G - \ell_G + 1}\sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{\underline{d}_x}) \leq O_p\left(\left(\frac{m_G}{n}\right)^{\underline{d}_x}\frac{1}{\ell_G^{1+\epsilon}}\right), \tag{B.14}$$

similarly to (B.9). Now, make the decomposition,

$$\widehat{\boldsymbol{G}}^{(2,1)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G) = (\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m))'\left(\widehat{\boldsymbol{G}}^{(2)}_{\widehat{u}\widehat{e}}(\ell_G,m_G) + \widehat{\boldsymbol{G}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell_G,m_G)\right) \tag{B.15}$$

where $\widehat{\boldsymbol{G}}^{(c,2)}_{\widehat{c}\widehat{e}}(\ell_G,m_G) \leq O_p((m_G/n)^{\underline{d}_x} 1/\ell_G^{1+\epsilon})$ by (B.9) and since

$$\boldsymbol{w}_u(\lambda_j,i) = O_p(1) + O_p\left(\frac{n^{1/2-d_i}}{j^{1-d_i}}\right) + O_p\left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right), \quad i = 2,\ldots,k+1, \tag{B.16}$$

by AVOA (2020, Lemma A.12(b)), we may write

$$\widehat{\boldsymbol{G}}^{(2)}_{\widehat{u}\widehat{e}}(\ell_G,m_G) \leq \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{d_1}) + \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{d_1+\underline{d}_x} n^{1/2}}{j}\right) + \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{d_1}\ln(n)n^{1/2}}{m_d^{1/2} j}\right)$$
$$\leq O_p(1) + O_p\left(\left(\frac{m_G}{n}\right)^{\underline{d}_x}\frac{n^{1/2}}{m_G^{1-\epsilon}\ell_G^{1+\epsilon}}\right) + O_p\left(\frac{n^{1/2}\ln(n)}{m_G^{1-\epsilon} m_d^{1/2}\ell_G^{1+\epsilon}}\right), \tag{B.17}$$

for some arbitrarily small $\epsilon > 0$, using $d_1 \geq 0$ and Varneskov (2017, Lemma C.4). Hence, by combining bounds, $n^{1/2}/m_G \to 0$, Lemmas B.1(a)-(b) in the absence of cointegration in conjunction with the continuous mapping theorem, we have $\widehat{\boldsymbol{G}}^{(2)}_{\widehat{u}\widehat{e}}(\ell_G,m_G) \leq O_p(1)$ and, thus,

$$\widehat{\boldsymbol{G}}^{(2,1)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G) \leq O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon}). \tag{B.18}$$

Consequently, by collecting bounds for all components in (B.12),

$$\widehat{\boldsymbol{G}}^{(2,1)}_{\widehat{\eta}\widehat{\tau}}(\ell_G,m_G) \leq O_p((m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon}) + O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon}), \tag{B.19}$$

thereby providing the requisite result when cointegration is absent.

The case with cointegration. First, recall $b = d_1 > 0$ and let us make the decomposition,

$$\widehat{\eta}_t^c \equiv (1-L)^{-\widehat{b}}\widehat{\eta}_t^{(b,c)} = (1-L)^{-\widehat{b}}\left(\widehat{\eta}_t^{(b)} - \widehat{\tau}^{(1)}_{t-1} - \widehat{\tau}^{(2)}_{t-1}\right) \equiv \widehat{\eta}_t - \widetilde{\tau}^{(1)}_{t-1} - \widetilde{\tau}^{(2)}_{t-1}, \tag{B.20}$$

noting that $\xi_{t-1} = 0$, for all $t = 1,\ldots,n$ in $\widehat{e}_t^{(1)}$ and, thus, in $\widehat{\eta}_t^{(b,1)}$. Hence, we may decompose the estimators with and without regressor endogeneity, similarly to (B.21),

$$\widehat{\boldsymbol{G}}^{c}_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G) - \widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}(\ell_G,m_G) = \widehat{\boldsymbol{G}}^{(1,1)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G,m_G) + \widehat{\boldsymbol{G}}^{(2,2)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G,m_G) + 2\widehat{\boldsymbol{G}}^{(1,2)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G,m_G)$$
$$- 2\widehat{\boldsymbol{G}}^{(1)}_{\widehat{\eta}\widetilde{\tau}}(\ell_G,m_G) - 2\widehat{\boldsymbol{G}}^{(2)}_{\widehat{\eta}\widetilde{\tau}}(\ell_G,m_G). \tag{B.21}$$

Next, we need the asymptotic bounds for $w_{\widehat{\eta}}(\lambda_j)$, $w^{(1)}_{\widetilde{\tau}}(\lambda_j)$ and $w^{(2)}_{\widetilde{\tau}}(\lambda_j)$. To this end, recall the bounding function $\bar{g}_n(m,m_b,m_d) = \frac{\ln(n)}{\sqrt{m_b}} \vee \frac{\ln(n)}{\sqrt{m_d}} \vee \frac{\lambda_m^b}{\sqrt{m}}$, then, by AVOA (2020, Lemma A.9(b)), we have

$$w_{\widehat{\eta}}(\lambda_j) = w_{\eta}(\lambda_j) + O_p\left(\left(\frac{j}{n}\right)^{d-b}\frac{n^{1/2}}{j}\right) + O_p\left(\bar{g}_n(m,m_b,m_d)\lambda_j^{-b}\right), \tag{B.22}$$

where $w_\eta(\lambda_j) = O_p(1)$. Moreover, by writing $\widetilde{\tau}^{(2)}_{t-1} = \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)'(1-L)^{-\widehat{b}}\widehat{\boldsymbol{c}}_{t-1} \equiv \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)'\widetilde{\boldsymbol{c}}_{t-1}$ as well as by defining $\theta_x = \underline{d}_x - b$, we may use $\widehat{\boldsymbol{\mathcal{B}}}(\ell,m) = O_p(1)$, uniformly by AV (2020, Theorem 1), in conjunction with AVOA (2020, Lemmas A.8 and A.9(a)) to deduce $w^{(2)}_{\widetilde{\tau}}(\lambda_j) = \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)'\boldsymbol{w}_{\widetilde{c}}(\lambda_j) = O_p(\boldsymbol{w}_{\widetilde{c}}(\lambda_j))$, with

$$\boldsymbol{w}_{\widetilde{c}}(\lambda_j) \leq O_p\left(\lambda_j^{\underline{d}_x-b}\right) + O_p\left(\lambda_j^{\underline{d}_x-b}\frac{\ln(n)}{j^{1/2}}\right) + O_p\left(n^{-(\underline{d}_x-b)-1}\right) = O_p\left(\lambda_j^{\underline{d}_x-b}\right), \tag{B.23}$$

as $j \to \infty$ when $n \to \infty$.[24] Hence, $w^{(2)}_{\widetilde{\tau}}(\lambda_j) \leq O_p(\lambda_j^{\underline{d}_x - b})$. Next, for $\widetilde{\tau}^{(1)}_{t-1}$, we may similarly write

$$\widetilde{\tau}^{(1)}_{t-1} = \left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\right)' \left((1-L)^{-\widehat{b}} (\widehat{\boldsymbol{u}}_{t-1} + \widehat{\boldsymbol{c}}_{t-1})\right) \equiv \left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\right)' (\widetilde{\boldsymbol{u}}_{t-1} + \widetilde{\boldsymbol{c}}_{t-1}),$$

and, thus, decompose its DFT as,

$$w^{(1)}_{\widetilde{\tau}}(\lambda_j) = \left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\right)' (\boldsymbol{w}_{\widetilde{u}}(\lambda_j) + \boldsymbol{w}_{\widetilde{c}}(\lambda_j)), \tag{B.24}$$

with $\boldsymbol{w}_{\widetilde{c}}(\lambda_j) \leq O_p(\lambda_j^{\underline{d}_x - b})$, as for $\widetilde{\tau}^{(2)}_{t-1}$. For $\boldsymbol{w}_{\widetilde{u}}(\lambda_j)$, let us further write,

$$\widetilde{\boldsymbol{u}}_{t-1} = (1-L)^{-\widehat{b}} (\widehat{\boldsymbol{u}}_{t-1} - \boldsymbol{u}_{t-1}) + (1-L)^{-\widehat{b}} \boldsymbol{u}_{t-1} \equiv \widetilde{\boldsymbol{u}}^{(1)}_{t-1} + \widetilde{\boldsymbol{u}}^{(2)}_{t-1}, \tag{B.25}$$

and, accordingly, decompose the DFT as $\boldsymbol{w}_{\widetilde{u}}(\lambda_j) = \boldsymbol{w}^{(1)}_{\widetilde{u}}(\lambda_j) + \boldsymbol{w}^{(2)}_{\widetilde{u}}(\lambda_j)$. First, for $\boldsymbol{w}^{(1)}_{\widetilde{u}}(\lambda_j)$, we use AVOA (Lemmas A.9(a) and A.10) with $\theta_x \geq 0$ to show,

$$\boldsymbol{w}^{(1)}_{\widetilde{u}}(\lambda_j) = O_p\left(\lambda_j^{-b} \frac{\ln(n)}{\sqrt{m_b}}\right) + O_p\left(\left(\frac{j}{n}\right)^{\underline{d}_x - b} \frac{n^{1/2}}{j}\right). \tag{B.26}$$

Similarly, by AVOA (Lemmas A.8 and A.9(a)), with $\underline{d} - b \geq 0$, $\boldsymbol{w}^{(2)}_{\widetilde{u}}(\lambda_j) = O_p(\lambda_j^{-b})$ and, thus,

$$\boldsymbol{w}_{\widetilde{u}}(\lambda_j) = O_p\left(\lambda_j^{-b}\right) + O_p\left(\left(\frac{j}{n}\right)^{\underline{d}_x - b} \frac{n^{1/2}}{j}\right), \tag{B.27}$$

by Assumption B. Hence, $w^{(1)}_{\widetilde{\tau}}(\lambda_j) = O_p(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)) \times O_p(\boldsymbol{w}_{\widetilde{u}}(\lambda_j))$, where, as for the case without cointegration, $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m) \leq O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon})$. Hence, it suffices to establish bounds for the terms in (B.21) with $\boldsymbol{w}_{\widetilde{u}}(\lambda_j)$ and apply the bound for $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)$.

Now, for the decomposition in (B.21), we use $w^{(2)}_{\widetilde{\tau}}(\lambda_j) \leq O_p(\lambda_j^{\underline{d}_x - b})$ and $b \leq \underline{d} \leq \underline{d}_x$ to write,

$$\widehat{\boldsymbol{G}}^{(2,2)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G, m_G) \leq \frac{K}{m_G} \sum_{\ell_G}^{m_G} O_p\left(\lambda_j^{2(\underline{d}_x - b)}\right) \leq O_p\left(\left(\frac{m_G}{n}\right)^{2(\underline{d} - b)} \frac{1}{\ell_G^{1+\epsilon}}\right) \tag{B.28}$$

for some arbitrarily small $\epsilon > 0$, using $m_G/(m_G - \ell_G + 1) \leq K$ and Varneskov (2017, Lemma C.4) to derive the final bound. Moreover, since we have the decomposition,

$$\widehat{\boldsymbol{G}}^{(1,1)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G, m_G) = O_p\left(\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\right)' \left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\right)\right) \times O_p\left(\widehat{\boldsymbol{G}}_{\widetilde{u}\widetilde{u}}(\ell_G, m_G)\right), \tag{B.29}$$

[24] Note that one term from AVOA (Lemma A.8) may be dropped since $\theta_x \geq 0$.

we may use the DFT bound in (B.27) to show

$$\widehat{\boldsymbol{G}}_{\widetilde{u}\widetilde{u}}(\ell_G, m_G) \leq \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{-2b}\right) + \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{-2b}\left(\frac{j}{n}\right)^{2\underline{d}}\frac{n}{j^2}\right)$$
$$+ \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{-2b}\left(\frac{j}{n}\right)^{2\underline{d}}\frac{n^{1/2}}{j}\right) \equiv \widetilde{\mathcal{E}}_1 + \widetilde{\mathcal{E}}_2 + \widetilde{\mathcal{E}}_3, \tag{B.30}$$

using, again, $m_G/(m_G - \ell_G + 1) \leq K$. For the terms in the decomposition, we have

$$\widetilde{\mathcal{E}}_1 \leq \frac{Kn^{2b}m_G}{m_G^{2b}}\sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{2(1-b)}\frac{1}{j^2}\right) \leq O_p\left(\left(\frac{n}{m_G}\right)^{2\underline{d}}\left(\frac{m_G}{n}\right)^{2(\underline{d}-b)}\frac{m_G}{\ell_G^2}\right),$$
$$\widetilde{\mathcal{E}}_2 \leq \frac{Km_G^{2(\underline{d}-b)-1}}{n^{2(\underline{d}-b)-1}}\sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{2(\underline{d}-b)}\frac{1}{j^2}\right) \leq O_p\left(\left(\frac{n}{m_G}\right)\left(\frac{m_G}{n}\right)^{2(\underline{d}-b)}\frac{1}{\ell_G^2}\right),$$
$$\widetilde{\mathcal{E}}_3 \leq \frac{Kn^{1/2+1}}{m_G}\left(\frac{m_G}{n}\right)^{1+\underline{d}-2b}\sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{1+\underline{d}-2b}\frac{1}{j^2}\right) \leq O_p\left(\frac{n^{1/2+\underline{d}}}{m_G^{\underline{d}}}\left(\frac{m_G}{n}\right)^{2(\underline{d}-b)}\frac{1}{\ell_G^2}\right),$$

using, again, Varneskov (2017, Lemma C.4). Hence, by collecting results, we may write,

$$\widehat{\boldsymbol{G}}_{\widetilde{u}\widetilde{u}}(\ell_G, m_G) \leq O_p\left(\left(\frac{m_G}{n}\right)^{2(\underline{d}-b)}\frac{1}{\ell_G^{1+\epsilon}}\left(\left(\frac{n}{m_G}\right)^{2\underline{d}}\frac{m_G}{\ell_G^{1-\epsilon}} + \left(\frac{n}{m_G}\right)\frac{1}{\ell_G^{1-\epsilon}} + \left(\frac{n}{m_G}\right)^{\underline{d}}\frac{n^{1/2}}{\ell_G^{1-\epsilon}}\right)\right) \tag{B.31}$$

Furthermore, since $O_p((\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m))'(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m))) \leq O_p((m/n)^{2\underline{d}_x}/\ell^{2(1+\epsilon)})$, it suffices to show that the second term inside the parenthesis in (B.31) multiplied by $((m/n)^{2\underline{d}_x}/\ell^{2(1+\epsilon)})$ is $o(1)$ to establish $\widehat{\boldsymbol{G}}^{(1,1)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G, m_G) \leq o_p((m_G/n)^{2(\underline{d}-b)}/\ell_G^{1+\epsilon})$. For the first of these terms, we have

$$\left(\frac{n}{m_G}\right)^{2\underline{d}}\frac{m_G}{\ell_G^{1-\epsilon}}\left(\frac{m}{n}\right)^{2\underline{d}_x}\frac{1}{\ell^{2(1+\epsilon)}} \leq \left(\frac{m}{m_G}\right)^{\underline{d}}\frac{m_G}{\ell_G^{1-\epsilon}\ell^{2(1+\epsilon)}} \to 0,$$

by Assumption T-B, as $\epsilon > 0$ is arbitrarily small. Similarly, for the second and third term,

$$\left(\frac{n}{m_G}\right)\frac{1}{\ell_G^{1-\epsilon}}\left(\frac{m}{n}\right)^{2\underline{d}_x}\frac{1}{\ell^{2(1+\epsilon)}} \leq \left(\frac{n}{m_G}\right)\frac{1}{\ell_G^{1-\epsilon}\ell^{2(1+\epsilon)}} \to 0,$$
$$\left(\frac{n}{m_G}\right)^{\underline{d}}\frac{n^{1/2}}{\ell_G^{1-\epsilon}}\left(\frac{m}{n}\right)^{2\underline{d}_x}\frac{1}{\ell^{2(1+\epsilon)}} \leq \left(\frac{m}{m_G}\right)^{\underline{d}_x}\left(\frac{m}{n}\right)^{\underline{d}_x}\frac{n^{1/2}}{\ell_G^{1-\epsilon}\ell^{2(1+\epsilon)}} \to 0,$$

by invoking $n^{1/2}/m_G \to 0$ and Assumption T-B. Hence, this implies

$$\widehat{\boldsymbol{G}}^{(1,1)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G, m_G) \leq o_p\left((m_G/n)^{2(\underline{d}-b)}/\ell_G^{1+\epsilon}\right) \tag{B.32}$$

and, by the Cauchy-Schwarz inequality,

$$|\widehat{\boldsymbol{G}}^{(1,2)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G,m_G)| \le \sqrt{\widehat{\boldsymbol{G}}^{(1,1)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G,m_G)\widehat{\boldsymbol{G}}^{(2,2)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G,m_G)} \le o_p\left((m_G/n)^{2(\underline{d}-b)}/\ell_G^{1+\epsilon}\right). \tag{B.33}$$

Next, for $\widehat{\boldsymbol{G}}^{(2)}_{\widetilde{\eta}\widetilde{\tau}}(\ell_G,m_G)$, we may use (B.22), $w^{(2)}_{\widetilde{\tau}}(\lambda_j) \le O_p(\lambda_j^{\underline{d}_x-b})$ and $b \le \underline{d} \le \underline{d}_x$ to write

$$\widehat{\boldsymbol{G}}^{(2)}_{\widetilde{\eta}\widetilde{\tau}}(\ell_G,m_G) \le \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{\underline{d}_x-b}\right) + \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{2(\underline{d}-b)}\frac{n^{1/2}}{j}\right)$$
$$+\frac{K\bar{g}_n(m,m_b,m_d)}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{\underline{d}-2b}\right) \equiv \widetilde{\mathcal{A}}_1+\widetilde{\mathcal{A}}_2+\widetilde{\mathcal{A}}_3. \tag{B.34}$$

By applying the same arguments as for $\widetilde{\mathcal{E}}_1$, $\widetilde{\mathcal{E}}_2$ and $\widetilde{\mathcal{E}}_3$, we have, with $\bar{g}_n(\cdot) \equiv \bar{g}_n(m,m_b,m_d)$,

$$\widetilde{\mathcal{A}}_1 \le O_p\left(\left(\frac{m_G}{n}\right)^{\underline{d}-b}\frac{1}{\ell_G^{1+\epsilon}}\right),$$
$$\widetilde{\mathcal{A}}_2 \le \frac{Kn^{1/2}}{m_G^{1-\epsilon}}\left(\frac{m_G}{n}\right)^{2(\underline{d}-b)}\sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{2(\underline{d}-b)}\frac{1}{j^{1+\epsilon}}\right) \le O_p\left(\frac{n^{1/2}}{m_G^{1-\epsilon}}\left(\frac{m_G}{n}\right)^{2(\underline{d}-b)}\frac{1}{\ell_G^{1+\epsilon}}\right),$$
$$\widetilde{\mathcal{A}}_3 \le K\bar{g}_n(\cdot)\left(\frac{n}{m_G}\right)^{\underline{d}}\sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{2(\underline{d}-b)}\frac{1}{j^{1+\epsilon}}\right) \le O_p\left(\bar{g}_n(\cdot)\left(\frac{n}{m_G}\right)^{b}\left(\frac{m_G}{n}\right)^{\underline{d}-b}\frac{1}{\ell_G^{1+\epsilon}}\right),$$

implying that, by Assumption T-B, $\widetilde{\mathcal{A}}_2+\widetilde{\mathcal{A}}_3 \le o_p(\widetilde{\mathcal{A}}_1)$ and, thus,

$$\widehat{\boldsymbol{G}}^{(2)}_{\widetilde{\eta}\widetilde{\tau}}(\ell_G,m_G) \le O_p\left((m_G/n)^{\underline{d}-b}/\ell_G^{1+\epsilon}\right). \tag{B.35}$$

For the final term, $\widehat{\boldsymbol{G}}^{(1)}_{\widetilde{\eta}\widetilde{\tau}}(\ell_G,m_G)$, we have

$$\widehat{\boldsymbol{G}}^{(1)}_{\widetilde{\eta}\widetilde{\tau}}(\ell_G,m_G) = O_p\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m)-\widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\right)\times O_p\left(\widehat{\boldsymbol{G}}_{\widetilde{u}\widetilde{\eta}}(\ell_G,m_G)\right).$$

Hence, we may use (B.22), (B.27) and $b \le \underline{d} \le \underline{d}_x$ to write,

$$\widehat{\boldsymbol{G}}_{\widetilde{u}\widetilde{\eta}}(\ell_G,m_G) \le \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{-b}\right) + \frac{K\,n^{1/2}}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{\underline{d}-b}}{j}\right) + \frac{K\,n^{1/2}}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{\underline{d}-2b}}{j}\right)$$
$$+\frac{K\,n}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{2(\underline{d}-b)}}{j^2}\right) + \frac{K\bar{g}_n(\cdot)}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{-2b}\right) + \frac{K\,n^{1/2}\bar{g}_n(\cdot)}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{\underline{d}-2b}}{j}\right) \equiv \sum_{g=1}^{6}\widetilde{\mathcal{B}}_g.$$

First, note that we readily have $\widetilde{\mathcal{B}}_2+\widetilde{\mathcal{B}}_6 \le O_p(\widetilde{\mathcal{B}}_3)$. For the remaining terms, we may apply the same arguments used to establish bounds for $\{\widetilde{\mathcal{A}}_i, \widetilde{\mathcal{E}}_i\}$, $i = 1, 2, 3$, to show,

$$\begin{aligned}
\widetilde{\mathcal{B}}_1 &\leq \frac{K\,n^b}{m_G^{b-\epsilon}} \sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{1-b} \frac{1}{j^{1+\epsilon}}\right) \leq O_p\left(\left(\frac{n}{m_G}\right)^{\underline{d}} \left(\frac{m_G}{n}\right)^{\underline{d}-b} \frac{m_G^{\epsilon}}{\ell_G^{1+\epsilon}}\right), \\
\widetilde{\mathcal{B}}_4 &= O_p\left(\widetilde{\mathcal{E}}_2\right) \leq O_p\left(\left(\frac{n}{m_G}\right)\left(\frac{m_G}{n}\right)^{2(\underline{d}-b)} \frac{1}{\ell_G^2}\right), \\
\widetilde{\mathcal{B}}_5 &= O_p\left(\bar{g}_n(\cdot)\widetilde{\mathcal{E}}_1\right) \leq O_p\left(\left(\frac{n}{m_G}\right)^{2\underline{d}} \left(\frac{m_G}{n}\right)^{2(\underline{d}-b)} \frac{\bar{g}_n(\cdot) m_G}{\ell_G^2}\right), \\
\widetilde{\mathcal{B}}_3 &\leq \frac{K\,n^{1/2+\underline{d}}}{m_G} \left(\frac{m_G}{n}\right)^{2(\underline{d}-b)} \sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{2(\underline{d}-b)} \frac{1}{j^{1+\underline{d}}}\right) \leq O_p\left(\frac{n^{1/2+\underline{d}}}{m_G}\left(\frac{m_G}{n}\right)^{2(\underline{d}-b)} \frac{1}{\ell_G^{1+\underline{d}}}\right),
\end{aligned}$$

which, needs to be scaled with $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m) = O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon})$ to determine the asymptotic order of $\widehat{\boldsymbol{G}}^{(1)}_{\widehat{\eta}\widetilde{\tau}}(\ell_G, m_G)$. Specifically, we will use the trimming and bandwidth conditions in Assumption T-B to show that $\widehat{\boldsymbol{G}}^{(1)}_{\widehat{\eta}\widetilde{\tau}}(\ell_G, m_G) \leq o_p((m_G/n)^{\underline{d}}/\ell_G^{1+\epsilon})$, similarly to the arguments used to bound the terms in $\widehat{G}^{(1,1)}_{\widetilde{\tau}\widetilde{\tau}}(\ell_G, m_G)$ above. First, for $\widetilde{\mathcal{B}}_1$ and $\widetilde{\mathcal{B}}_4$, this follow by

$$\left(\frac{n}{m_G}\right)^{\underline{d}} \left(\frac{m}{n}\right)^{\underline{d}_x} \frac{m_G^{\epsilon}}{\ell^{1+\epsilon}} \leq \left(\frac{m}{m_G}\right)^{\underline{d}} \frac{m_G^{\epsilon}}{\ell^{1+\epsilon}} \to 0, \quad \left(\frac{n}{m_G}\right)\left(\frac{m_G}{n}\right)^{\underline{d}-b}\left(\frac{m}{n}\right)^{\underline{d}_x} \frac{1}{\ell_G^{1-\epsilon}\ell^{1+\epsilon}} \to 0,$$

respectively, using Assumption T-B and $\epsilon > 0$ being arbitrarily small. To see this, note that the conditions $n^{1/2}/m_G \to 0$ and $m_G/(\ell_G\,\ell) \to 0$ implies $(n/m_G)/(\ell_G\,\ell) \to 0$. Similarly, for $\widetilde{\mathcal{B}}_3$ and $\widetilde{\mathcal{B}}_5$,

$$\begin{aligned}
\frac{n^{1/2+\underline{d}}}{m_G}\left(\frac{m_G}{n}\right)^{\underline{d}-b}\left(\frac{m}{n}\right)^{\underline{d}_x}\frac{1}{\ell_G^{\underline{d}-\epsilon}}\frac{1}{\ell^{1+\epsilon}} &\leq \left(\frac{m_G}{n}\right)^{\underline{d}-b}\left(\frac{m}{m_G}\right)^{\underline{d}}\frac{n^{1/2}}{m_G}\left(\frac{m_G^{\underline{d}}}{\ell_G^{\underline{d}-\epsilon}\ell^{1+\epsilon}}\right) \to 0, \\
\left(\frac{n}{m_G}\right)^{2\underline{d}}\left(\frac{m_G}{n}\right)^{\underline{d}-b}\frac{m_G}{\ell_G^{1-\epsilon}}\left(\frac{m}{n}\right)^{\underline{d}_x}\frac{\bar{g}_n(\cdot)}{\ell^{1+\epsilon}} &\leq \left(\frac{m}{m_G}\right)^{\underline{d}}\bar{g}_n(\cdot)\left(\frac{n}{m_G}\right)^{b}\frac{m_G}{\ell_G^{1-\epsilon}\ell^{1+\epsilon}} \to 0,
\end{aligned}$$

respectively, using Assumption T-B and $\underline{d} \leq 1$. Hence, $\widehat{\boldsymbol{G}}^{(1)}_{\widehat{\eta}\widetilde{\tau}}(\ell_G, m_G) \leq o_p((m_G/n)^{\underline{d}}/\ell_G^{1+\epsilon})$ and, by collecting results from all components of the decomposition (B.21), we have

$$\widehat{\boldsymbol{G}}^c_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) \leq O_p\left((m_G/n)^{\underline{d}-b}/\ell_G^{1+\epsilon}\right), \tag{B.36}$$

thereby providing the result in the presence of cointegration, concluding the proof. □

B.1 Proof of Theorem 1

First, recall that $\widehat{\boldsymbol{v}}_t = (\widehat{e}_t, \widehat{\boldsymbol{u}}'_{t-1})'$, then, by invoking Lemmas B.1(c)-(d), we have,

$$\widehat{\boldsymbol{G}}^c_{\widehat{v}\widehat{v}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\widehat{v}\widehat{v}}(\ell_G, m_G) \leq O_p\left((m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon}\right), \tag{B.37}$$

for some arbitrarily small $\epsilon > 0$. Hence, we may continue by working with the corresponding estimate without regressor endogeneity, $\widehat{\boldsymbol{v}}_t$. Next, define $\widehat{\boldsymbol{A}}(L) \equiv \widehat{\boldsymbol{D}}(L)\boldsymbol{D}(L)^{-1}$ and $\boldsymbol{a}_t \equiv \boldsymbol{D}(L)\boldsymbol{z}_t$ such that we have $\widehat{\boldsymbol{v}}_t = \widehat{\boldsymbol{A}}(L)\boldsymbol{a}_t$. Moreover, we may write $\boldsymbol{a}_t = \boldsymbol{\mu}_t + \boldsymbol{v}_t$, where $\boldsymbol{\mu}_t \equiv \boldsymbol{D}(L)\boldsymbol{\mu}\mathbf{1}_{\{t\geq 1\}}$ and $\boldsymbol{v}_t = (e_t, \boldsymbol{u}_{t-1}')'$ with $e_t = \varphi_{t-1} + \eta_t^{(d_1)}$ and $\varphi_{t-1} = \boldsymbol{\mathcal{B}}'\boldsymbol{u}_{t-1} + \xi_{t-1}$. Finally, let us define $\boldsymbol{\psi}_{t-1} = (\varphi_{t-1}, \boldsymbol{u}_{t-1}')'$ and write,

$$\begin{aligned}\widehat{\boldsymbol{G}}_{\widehat{v}\widehat{v}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\psi\psi}(1, m_G) &= \Big(\widehat{\boldsymbol{G}}_{\psi\psi}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\psi\psi}(1, m_G)\Big) + \Big(\widehat{\boldsymbol{G}}_{vv}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\psi\psi}(\ell_G, m_G)\Big) \\ &\quad + \Big(\widehat{\boldsymbol{G}}_{\widehat{v}\widehat{v}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{vv}(\ell_G, m_G)\Big) \equiv \boldsymbol{\mathcal{U}}_1^{(G)} + \boldsymbol{\mathcal{U}}_2^{(G)} + \boldsymbol{\mathcal{U}}_3^{(G)},\end{aligned} \tag{B.38}$$

Then, the following lemma provides asymptotic bounds for $\boldsymbol{\mathcal{U}}_1^{(G)}$, $\boldsymbol{\mathcal{U}}_2^{(G)}$ and $\boldsymbol{\mathcal{U}}_3^{(G)}$ as well as a central limit theorem for $\widehat{\boldsymbol{G}}_{\psi\psi}(1, m_G)$. Hence, the stated limit theory follows by applying Assumption T-G to eliminate the sampling and trimming errors in conjunction with Slutsky's theorem. ☐

Lemma B.2. *Under the conditions of Theorem 1, the following uniform bounds hold:*

(a) $m_G^{1/2}\boldsymbol{\mathcal{U}}_1^{(G)} = O_p\left(\ell_G/\sqrt{m_G}\right)$.

(b) $m_G^{1/2}\boldsymbol{\mathcal{U}}_2^{(G)} \leq O_p\left((\ell_G/\sqrt{m_G})(\ell_G/n)^{d_1}\right) + O_p\left((m_G/n)^{d_1}/\sqrt{m_G}\right)$.

(c) For some arbitrarily small $\epsilon > 0$,

$$m_G^{1/2}\boldsymbol{\mathcal{U}}_3^{(G)} \leq O_p\left(\frac{n}{\sqrt{m_G}\ell_G^2}\left(\frac{m_G}{n}\right)^{2\underline{d}_x}\right) + O_p\left(\frac{n^{1/2}}{\sqrt{m_G}}\left(\frac{m_G}{n}\right)^{\underline{d}}\frac{m_G^{\epsilon}}{\ell_G^{1+\epsilon}}\right) + O_p\left(\ln(n)(m_G/m_d)^{1/2}\right).$$

(d) Let $m_G^{1+2\varpi}/n^{\varpi}$ *for* $m_G \asymp n^{\kappa_G}$ *and* $\ell \asymp n^{\nu_G}$*, with* $0 < \nu_G < \kappa_G < \varrho \leq 1$*, then*

$$\begin{aligned}& m_G^{1/2}\mathrm{vec}\Big(\widehat{\boldsymbol{G}}_{\psi\psi}(1, m_G) - \boldsymbol{G}_{\psi\psi}\Big) \\ &\qquad \xrightarrow{\mathbb{D}} N\left(\mathbf{0}, \left(\boldsymbol{G}_{\psi\psi}\otimes\boldsymbol{G}_{\psi\psi} + \left(\boldsymbol{G}_{\psi\psi}\otimes\boldsymbol{G}_{\psi\psi}^{(1)}, \ldots, \boldsymbol{G}_{\psi\psi}\otimes\boldsymbol{G}_{\psi\psi}^{(k+1)}\right)\right)/2\right).\end{aligned}$$

Proof. **For (a).** First, by the cancellation of terms in the summation, $\boldsymbol{\mathcal{U}}_1^{(G)} = -\widehat{\boldsymbol{G}}_{\psi\psi}(1, \ell_G - 1)$. The result, then, follows by Christensen & Varneskov (2017, Lemma 6).

For (b). First, define the $(k+1)\times 1$ vector $\check{\boldsymbol{c}}_t \equiv (\eta_t^{(d_1)}, \mathbf{0}_k')'$, such that $\boldsymbol{v}_t - \boldsymbol{\psi}_t = \check{\boldsymbol{c}}_t$, and make the decomposition,

$$\boldsymbol{\mathcal{U}}_2^{(G)} = \widehat{\boldsymbol{G}}_{\check{c}\check{c}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\psi\check{c}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\check{c}\psi}(\ell_G, m_G) \tag{B.39}$$

Next, as for the bounds in (B.5), we invoke AVOA (2020, Lemma A.12(b)) to show that $w_\eta^{(d_1)}(\lambda_j) = O_p(\lambda_j^{d_1})$, when $j \asymp n^{\jmath}$, $\forall \jmath > 0$. Hence, uniformly,

$$\widehat{\boldsymbol{G}}_{\check{c}\check{c}}(\ell_G, m_G) = \frac{1}{m_G - \ell_G + 1}\sum_{j=\ell_G}^{m_G} O_p\left(\lambda_j^{2d_1}\right) \leq \left(\frac{m_G}{n}\right)^{2d_1}\frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{\ell_G}{m_G}\right)^{2d_1}\right). \tag{B.40}$$

Moreover, by applying Shimotsu & Phillips (2005, Lemma 5.4(a)), we have

$$\frac{1}{m_G}\sum_{j=\ell_G}^{m_G}\left(\frac{\ell_G}{m_G}\right)^{2d_1} = \int_{\frac{\ell_G}{m_G}}^{1} x^{2d_1}dx + O\left(m_G^{-1}\right) = O\left((\ell_G/m_G)^{1+2d_1}\right) + O\left(m_G^{-1}\right), \tag{B.41}$$

and, as a result, $\widehat{\boldsymbol{G}}_{\check{c}\check{c}}(\ell_G, m_G) \leq O_p((\ell_G/m_G)(\ell_G/n)^{2d_1}) + O_p((m_G/n)^{2d_1}/m_G)$. Now, by combining Assumptions D1-D3, the same arguments and the Cauchy-Schwarz inequality, $\widehat{\boldsymbol{G}}_{\check{c}\psi}(\ell_G, m_G) \leq O_p((\ell_G/m_G)(\ell_G/n)^{d_1}) + O_p((m_G/n)^{d_1}/m_G)$, providing the result as $\ell_G/n + m_G/n \to 0$ and $d_1 > 0$.

For (c). First, let us further decompose the error term as,

$$\boldsymbol{\mathcal{U}}_3^{(G)} = \left(\widehat{\boldsymbol{G}}_{aa}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{vv}(\ell_G, m_G)\right) + \left(\widehat{\boldsymbol{G}}_{\widehat{v}\widehat{v}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{aa}(\ell_G, m_G)\right) \equiv \boldsymbol{\mathcal{U}}_{31}^{(G)} + \boldsymbol{\mathcal{U}}_{32}^{(G)}.$$

Moreover, write $\boldsymbol{a}_t = \boldsymbol{\psi}_t + \boldsymbol{\mu}_t + \check{\boldsymbol{c}}_t \equiv \boldsymbol{b}_t + \check{\boldsymbol{c}}_t$ and $\widehat{\boldsymbol{v}}_t = \widehat{\boldsymbol{A}}(L)(\boldsymbol{b}_t + \check{\boldsymbol{c}}_t) \equiv \tilde{\boldsymbol{v}}_t^{(1)} + \tilde{\boldsymbol{v}}_t^{(2)}$. Now, as,

$$\boldsymbol{\mathcal{U}}_{31}^{(G)} = \widehat{\boldsymbol{G}}_{\mu\mu}(\ell_G, m_G) + \left(\widehat{\boldsymbol{G}}_{\psi\mu}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\mu\psi}(\ell_G, m_G)\right) + \left(\widehat{\boldsymbol{G}}_{\check{c}\mu}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\mu\check{c}}(\ell_G, m_G)\right),$$

we may apply the same arguments as for AVOA (2020, Lemma A.4(a)) (cf., the error term $\boldsymbol{\mathcal{G}}_2$) to provide the following stochastic bounds,

$$\begin{aligned}\boldsymbol{\mathcal{U}}_{311}^{(G)} &\equiv \widehat{\boldsymbol{G}}_{\mu\mu}(\ell_G, m_G) + \left(\widehat{\boldsymbol{G}}_{\psi\mu}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\mu\psi}(\ell_G, m_G)\right) \\ &\leq O_p\left(\frac{n}{m_G\ell_G^2}\left(\frac{m_G}{n}\right)^{2\underline{d}_x}\right) + O_p\left(\frac{n^{1/2}}{m_G}\left(\frac{m_G}{n}\right)^{\underline{d}_x}\frac{m_G^{\epsilon}}{\ell_G^{1+\epsilon}}\right),\end{aligned} \tag{B.42}$$

for some arbitrarily small $\epsilon > 0$. Moreover, by $w_\eta^{(d_1)}(\lambda_j) = O_p(\lambda_j^{d_1})$ and Shimotsu (2010, Lemma B.2),

$$\widehat{\boldsymbol{G}}_{\mu\check{c}}(\ell_G, m_G) \leq \frac{K}{m_G}\sum_{j=\ell_G}^{m_G} O_p\left(n^{1/2}j^{-1}\right) \times O_p\left(\lambda_j^{d_1}\right) \leq \frac{K m_G^{\epsilon} n^{1/2}}{m_G}\left(\frac{m_G}{n}\right)^{d_1}\sum_{j=\ell_G}^{m_G} O_p(j^{-1-\epsilon}), \tag{B.43}$$

which, by Varneskov (2017, Lemma C.4), is uniformly $O_p(n^{1/2}m_G^{\epsilon-1}(m_G/n)^{d_1}\ell_G^{-1-\epsilon})$, implying

$$\boldsymbol{\mathcal{U}}_{31}^{(G)} \leq O_p\left(\frac{n}{m_G\ell_G^2}\left(\frac{m_G}{n}\right)^{2\underline{d}_x}\right) + O_p\left(\frac{n^{1/2}}{m_G}\left(\frac{m_G}{n}\right)^{\underline{d}}\frac{m_G^{\epsilon}}{\ell_G^{1+\epsilon}}\right). \tag{B.44}$$

Next, for the second term $\boldsymbol{\mathcal{U}}_{32}^{(G)}$, we may write,

$$\begin{aligned}\widehat{\boldsymbol{G}}_{aa}(\ell_G, m_G) &= \widehat{\boldsymbol{G}}_{bb}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\check{c}\check{c}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{b\check{c}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\check{c}b}(\ell_G, m_G), \\ \widehat{\boldsymbol{G}}_{\widehat{v}\widehat{v}}(\ell_G, m_G) &= \widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(1)}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(2)}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(12)}(\ell_G, m_G) + \widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(21)}(\ell_G, m_G).\end{aligned}$$

Moreover, by applying AVOA (2020, Lemmas A.8-A.9(a), Equations (A.60) and (A.65)), we have

$$\boldsymbol{w}_{\tilde{v}}^{(1)}(\lambda_j,i)=\boldsymbol{w}_b(\lambda_j,i)+O_p\left(\frac{\ln(n)}{m_d^{1/2}}\right)+O_p\left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right),\quad \boldsymbol{w}_b(\lambda_j,i)\equiv\boldsymbol{w}_u(\lambda_j,i)+\boldsymbol{w}_\mu(\lambda_j,i),$$
$$\boldsymbol{w}_{\tilde{v}}^{(2)}(\lambda_j)=\boldsymbol{w}_{\check{c}}(\lambda_j)+O_p\left(\frac{\lambda_j^{d_1}\ln(n)}{j^{1/2}}\right)+O_p\left(\frac{\lambda_j^{d_1}\ln(n)}{m_d^{1/2}}\right),\quad \boldsymbol{w}_{\check{c}}(\lambda_j)=O_p\left(\lambda_j^{d_1}\right),$$

when $j\asymp n^{\jmath}$, $\jmath>0$, for $i=1,\ldots,k+1$. Now, the difference $\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(1)}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{bb}(\ell_G,m_G)$ has already been considered in the proof of AVOA (2020, Lemma A.4(a)) (cf. the term $\boldsymbol{\mathcal{G}}_3$). Hence, by letting,

$$\bar{f}_G(\ell_G,m_G,n)=1\vee\frac{n^{1/2}m_G^{\epsilon}}{m_G\ell_G^{1+\epsilon}}\vee\frac{n}{m_G\ell_G^2},\quad\text{with}\quad\bar{f}_G(\ell_G,m_G,n)\to 1,\tag{B.45}$$

as $n\to\infty$ by Assumptions T and T-G (condition two), we have, by their arguments,

$$\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(1)}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{bb}(\ell_G,m_G)\leq O_p\left(\frac{\ln(n)^2}{m_d}\bar{f}_G(\ell_G,m_G,n)\right)+O_p\left(\frac{\ln(n)}{\sqrt{m_d}}\sqrt{\bar{f}_G(\ell_G,m_G,n)}\right),\tag{B.46}$$

and, thus, $\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(1)}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{bb}(\ell_G,m_G)\leq O_p(\ln(n)/\sqrt{m_d})$. Next, by applying the periodogram approximation error decomposition for $\boldsymbol{w}_{\tilde{v}}^{(2)}(\lambda_j)$, we have

$$\begin{aligned}\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(2)}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{\check{c}\check{c}}(\ell_G,m_G)&\leq\frac{K}{m_G}\sum_{j=\ell_G}^{m_G}O_p\left(\frac{\lambda_j^{2d_1}\ln(n)^2}{j}\right)+O_p\left(\frac{\lambda_j^{2d_1}\ln(n)^2}{j^{1/2}m_d^{1/2}}\right)+O_p\left(\frac{\lambda_j^{2d_1}\ln(n)^2}{m_d}\right)\\&\leq O_p\left(\left(\frac{m_G}{n}\right)^{2d_1}\left(\frac{m_G^{\epsilon}\ln(n)^2}{m_G\ell_G^{1+\epsilon}}+\frac{m_G^{\epsilon}\ln(n)^2}{\ell_G^{1+\epsilon}\sqrt{m_Gm_d}}+\frac{\ln(n)^2}{m_d}\left(\left(\frac{\ell_G}{m_G}\right)^{1+2d_1}+\frac{1}{m_G}\right)\right)\right)\end{aligned}$$

for some arbitrarily small $\epsilon>0$, using (B.41) and the same arguments as for (B.42) and (B.43). Hence, by Assumptions T and T-G, $\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(2)}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{\check{c}\check{c}}(\ell_G,m_G)\leq o_p(\boldsymbol{\mathcal{U}}_{311}^{(G)})$. Similarly,

$$\begin{aligned}&\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(12)}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{b\check{c}}(\ell_G,m_G)\leq\frac{K}{m_G}\sum_{j=\ell_G}^{m_G}\left(O_p\left(\frac{\ln(n)^2\lambda_j^{d_1}}{m_d^{1/2}j^{1/2}}\right)\right.\\&\left.+O_p\left(\frac{\ln(n)^2\lambda_j^{d_1}}{m_d}\right)+O_p\left(\frac{\ln(n)^2\lambda_j^{d_1}n^{1/2}}{m_d^{1/2}j^{3/2}}\right)+O_p\left(\frac{\ln(n)^2\lambda_j^{d_1}n^{1/2}}{m_dj}\right)\right)\\&\leq O_p\left(\left(\frac{m_G}{n}\right)^{d_1}\left(\frac{m_G^{\epsilon}\ln(n)^2}{\ell_G^{1+\epsilon}\sqrt{m_Gm_d}}+\frac{\ln(n)^2}{m_d}\left(\left(\frac{\ell_G}{m_G}\right)^{1+d_1}+\frac{1}{m_G}\right)+\frac{\ln(n)^2n^{1/2}}{m_d^{1/2}m_G\ell_G^{3/2}}+\frac{n^{1/2}\ln(n)^2}{\ell_G^{1+\epsilon}m_G^{1-\epsilon}m_d}\right)\right),\end{aligned}$$

which, by Assumptions T and T-G, similarly has $\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(12)}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{b\check{c}}(\ell_G,m_G)\leq o_p(\boldsymbol{\mathcal{U}}_{311}^{(G)})$, and the equivalent result for $\widehat{\boldsymbol{G}}_{b\check{c}}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(21)}(\ell_G,m_G)$ follows by symmetry. The final bound, thus, follows by collecting results for $\boldsymbol{\mathcal{U}}_{311}^{(G)}$ and $\widehat{\boldsymbol{G}}_{\tilde{v}\tilde{v}}^{(1)}(\ell_G,m_G)-\widehat{\boldsymbol{G}}_{bb}(\ell_G,m_G)$.

For (d). The central limit theory follows by Nielsen & Shimotsu (2007, Lemma 5). □

B.2 Proof of Theorem 2

The result follows by combining Theorem 1 and Robinson & Yajima (2002, Theorem 4). □

B.3 Proof of Theorem 3

First, recall that $\widehat{\boldsymbol{v}}_t = (\widehat{e}_t, \widehat{\boldsymbol{u}}'_{t-1})'$, then, by invoking Lemmas B.1(a)-(b) and the continuous mapping theorem,

$$\sqrt{m}\lambda_m^{-b}\Big(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell,m)\Big) \leq O_p\left((m/n)^{d_x - b}\sqrt{m}/\ell^{1+\epsilon}\right), \tag{B.47}$$

for some arbitrarily small $\epsilon > 0$. Hence, we may continue by working with the corresponding estimate without regressor endogeneity, $\widehat{\boldsymbol{v}}_t$. As for the proof of Theorem 1, define $\widehat{\boldsymbol{A}}(L) \equiv \widehat{\boldsymbol{D}}(L)\boldsymbol{D}(L)^{-1}$ and $\boldsymbol{a}_t \equiv \boldsymbol{D}(L)\boldsymbol{z}_t$ such that $\widehat{\boldsymbol{v}}_t = \widehat{\boldsymbol{A}}(L)\boldsymbol{a}_t$, and further write $\boldsymbol{a}_t = \boldsymbol{\mu}_t + \boldsymbol{v}_t$, where $\boldsymbol{\mu}_t \equiv \boldsymbol{D}(L)\boldsymbol{\mu}\mathbf{1}_{\{t\geq 1\}}$ and, again, $\boldsymbol{v}_t = (e_t, \boldsymbol{u}'_{t-1})'$ with $e_t = \varphi_{t-1} + \eta_t^{(d_1)}$, $d_1 = b$, and $\varphi_{t-1} = \boldsymbol{\mathcal{B}}'\boldsymbol{u}_{t-1} + \xi_{t-1}$. Finally, define $\mu_t^{(e)}$ as the first element of the vector $\boldsymbol{\mu}_t$ and $\boldsymbol{\mu}_t^{(u)}$ as the remaining $k \times 1$ subvector. Then, as in the corresponding proof of AV (2020, Theorem 1), we can write by addition and subtraction,

$$\widehat{\boldsymbol{\mathcal{B}}}(\ell,m) - \boldsymbol{\mathcal{B}} = \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{u\eta}^{(b)}(1,m) + \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{u\xi}(1,m) - \boldsymbol{\mathcal{C}}_1 + \boldsymbol{\mathcal{C}}_2 + \boldsymbol{\mathcal{C}}_3^{(b)} + \boldsymbol{\mathcal{C}}_4, \tag{B.48}$$

where the four error terms, $\boldsymbol{\mathcal{C}}_1$, $\boldsymbol{\mathcal{C}}_2$, $\boldsymbol{\mathcal{C}}_3^{(b)}$, and $\boldsymbol{\mathcal{C}}_4$ are defined as

$$\begin{aligned}
\boldsymbol{\mathcal{C}}_1 &\equiv \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{\widehat{u}\mu}^{(u)}(\ell,m)\boldsymbol{\mathcal{B}}, \qquad \boldsymbol{\mathcal{C}}_2 \equiv \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{\widehat{u}\mu}^{(e)}(\ell,m),\\
\boldsymbol{\mathcal{C}}_3^{(b)} &\equiv \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\left(\widehat{\boldsymbol{F}}_{\widehat{u}\eta}^{(b)}(\ell,m) - \widehat{\boldsymbol{F}}_{u\eta}^{(b)}(\ell,m) + \boldsymbol{\mathcal{D}}_1^{(b)}\right), \qquad \boldsymbol{\mathcal{D}}_1^{(b)} \equiv \widehat{\boldsymbol{F}}_{u\eta}^{(b)}(\ell,m) - \widehat{\boldsymbol{F}}_{u\eta}^{(b)}(1,m),\\
\boldsymbol{\mathcal{C}}_4 &\equiv \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\left(\widehat{\boldsymbol{F}}_{\widehat{u}\xi}(\ell,m) - \widehat{\boldsymbol{F}}_{u\xi}(\ell,m) + \boldsymbol{\mathcal{D}}_2\right), \qquad \boldsymbol{\mathcal{D}}_2 \equiv \widehat{\boldsymbol{F}}_{u\xi}(\ell,m) - \widehat{\boldsymbol{F}}_{u\xi}(1,m),
\end{aligned}$$

with the superscripts indicating $\boldsymbol{\mu}_t^{(u)}$ and $\mu_t^{(e)}$, respectively. Whereas the asymptotic properties of the terms $\boldsymbol{\mathcal{C}}_1$, $\boldsymbol{\mathcal{C}}_2$, $\boldsymbol{\mathcal{C}}_3^{(b)}$ and $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{u\eta}^{(b)}(1,m)$ are the same irrespective of the models (ii)-(iii) and model (iv), the properties $\boldsymbol{\mathcal{C}}_4$ and $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{u\xi}(1,m)$ depend on the inference regime.

Inference for model (iv): Since $\xi_{t-1} = 0$, $\forall t = 1,\ldots,n$, we have $\boldsymbol{\mathcal{C}}_4 = \mathbf{0}$ and $\widehat{\boldsymbol{F}}_{u\xi}(1,m) = \mathbf{0}$. Next, by applying AVOA (2020, Lemma A.2), we have,

$$\sqrt{m}\lambda_m^{-b}\big(\boldsymbol{\mathcal{C}}_1 + \boldsymbol{\mathcal{C}}_2 + \boldsymbol{\mathcal{C}}_3^{(b)}\big) = o_p(1), \quad \lambda_m^{-1}\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}. \tag{B.49}$$

The result, then, follows by applying AVOA (2020, Lemma A.3) to $\sqrt{m}\lambda_m^{-1-b}\widehat{\boldsymbol{F}}_{u\eta}^{(b)}(1,m)$ in conjunction with (B.49), the continuous mapping theorem and Slutsky's theorem.

Inference for models (ii) and (iii): Since $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{u\eta}^{(b)}(1,m) = O_p(\lambda_m^b m^{-1/2})$, with $0 < b = d_1 \leq 1$, and we may use AVOA (2020, Lemma A.2) to show $\sqrt{m}\,\boldsymbol{\mathcal{C}}_4 = o_p(1)$, despite ξ_{t-1} being non-trivial, the central limit theory follows by applying AVOA (2020, Lemma A.3) to $\sqrt{m}\lambda_m^{-1}\widehat{\boldsymbol{F}}_{u\xi}(1,m)$ in conjunction with (B.49), the continuous mapping theorem and Slutsky's theorem.

The mutual consistency condition follows by the corresponding in AV (2020, Theorem 1) since it is derived for the worst case bound $\underline{d} = b = 0$ and, thus, applies to both inference scenarios. □

B.4 Proof of Theorem 5

The result follows by combining Lemmas B.1(c) and (e) with AVOA (2020, Lemma A.4). □