

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, Revised December 2021

We wish to thank the Co-Editor Guido Kuersteiner, two anonymous referees as well as seminary participants at the 2018 conference in honor of Peter C.B. Phillips at Yale University, the 13th annual SoFiE conference, Durham University Business School and Singapore Management University for helpful comments. 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Torben G. Andersen and Rasmus T. Varneskov. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.,

Consistent Local Spectrum (LCM) Inference for Predictive Return Regressions
Torben G. Andersen and Rasmus T. Varneskov
NBER Working Paper No. 28569
March 2021, Revised December 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 develop a consistent Local speCtruM (LCM) estimation procedure, that delivers asymptotic Gaussian inference. Furthermore, we provide a new LCM-based estimator of the conditional mean persistence, that leverages biased regression slopes as well as new LCM-based tests for significance of (a subset of) the predictors, which are valid even without estimating the return persistence. Simulations illustrate the theoretical arguments. Finally, in an empirical application to monthly S&P 500 return predictions, we provide evidence for a fractionally integrated conditional mean component. Moreover, using our new LCM procedure and tools, we document significant predictive power for key state variables such as the default spread and treasury interest rates.

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) suggest that returns are predictable by persistent state variables, such as the mean and volatility of consumption growth or the time-varying disaster recovery rate; see Neuhierl & Varneskov (2021). Nonetheless, the reliability of the empirical findings and the design of appropriate econometric methodology remain 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,

$$y_t = \mathcal{B} x_{t-1} + v_t, \quad t = 1, \dots, n, \quad (1)$$

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

¹These assumptions simplify the exposition. 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 dependent return

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

If one invokes standard assumptions, including weak dependence and stationarity of the returns and regressor, it is straightforward to test for return predictability via the ordinary least squares (OLS) estimator $\hat{\mathcal{B}}_{\text{OLS}} = \sum_{t=1}^n y_t x_{t-1} / \sum_{t=1}^n x_{t-1}^2$. The null hypothesis of no predictability implies $\mathcal{B} = 0$, and a regular t -test for significance will apply. 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,

$$x_t = \phi_n x_{t-1} + w_t, \quad t = 1, \dots, n, \quad (2)$$

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

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, $\sigma_{vw} \neq 0$. Marriott & Pope (1954) show that this endogeneity bias asymptotically ($n \rightarrow \infty$), to first order, equals $-(\sigma_{vw}/\sigma_{ww}^2)(1 + 3\phi)/n$, 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 σ_{vw} 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 are typically substantially larger than the innovations in the regressor, inflating $(\sigma_{vw}/\sigma_{ww}^2)$, the bias may be substantial. This motivates Stambaugh (1986) to implement a bias-correction, which is applied frequently in the subsequent literature.

Whether this endogeneity correction ensures satisfactory inference hinges on the quality of the asymptotic approximation to the distribution for the regression coefficient, $\hat{\mathcal{B}}_{\text{OLS}}$. 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.

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.

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

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

⁴They 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 does not necessarily capture the “true” conditional mean dynamics of the returns, and the associated predictive relation should, at best, be viewed as providing an “imperfect” or a noisy indicator for the conditional mean, which can be interpreted as an omitted regressor problem. 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,

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

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 $\hat{\mathcal{B}}_{\text{OLS}}$ 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 inference problem above 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 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 the 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

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

section is that prior knowledge about the nature of the persistence of the regressor is not assumed, 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,

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

When β_z is chosen below, but near, unity, \tilde{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 \tilde{z}_t serving as instrument, $\hat{\mathcal{B}}_{\text{IVX}} = \sum_{t=1}^n y_t \tilde{z}_{t-1} / \sum_{t=1}^n x_{t-1} \tilde{z}_{t-1}$. In the unit root and LUR scenarios, the estimation error for OLS, $\sum_{t=1}^n v_t x_{t-1} / \sum_{t=1}^n x_{t-1}^2$ 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, $n^{(1+\beta_z)/2}(\hat{\mathcal{B}}_{\text{IVX}} - \mathcal{B}) \xrightarrow{\mathbb{D}} MN(0, \sigma_{\text{IVX}}^2)$. 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

⁶To see this, note that $\tilde{z}_t = z_t - (C_\phi/n^{\delta_\phi}) \sum_{s=1}^t \phi_{nz}^{t-s} x_{s-1}$, where $z_t = \phi_{nz} z_{t-1} + w_t$, implying \tilde{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.

ratios, macroeconomic variables, lagged returns, and realized volatility measures, it is a very strong 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

⁷In 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.

⁸The 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.

Wald tests. Likewise, Liu, Yang, Cai & Peng (2019) consider univariate predictive regressions, where 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 (2021a), 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,

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

where $u_{t-1} \in I(0)$ is weakly dependent, and $v_t \in I(d_y - b)$ with $0 \leq b \leq d_y$ captures the possibility of cointegration (when $b > 0$).⁹ 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 (2021a) 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

⁹As usual, we use the notation $I(d)$ to signify that a variable is integrated of exact order d .

with various financial and macroeconomic variables. This feature corresponds to having different localization coefficients in the LUR setting (3). Andersen & Varneskov (2021*a*) 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 applying a 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”.¹⁰

In this paper, we extend the framework in Andersen & Varneskov (2021*a*) to allow for imperfect regressors (in the spirit of Pastor & Stambaugh (2009)) that may exhibit general forms of endogeneity, which is similar to treating a omitted regressor problem, with the latter allowed to be persistent. 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, the LCM procedure relies critically on consistent estimation of the fractional integration orders of the variables, which is particularly difficult for return regressions, because 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 developing: (i) a new feasible inference procedure, which holds irrespective of whether the regressors are “imperfect” or “perfect”; an LCM-bias (LCMB) approach to persistence estimation for returns, which leverages biased regression slopes at lower frequencies and converges at a sufficiently fast rate to invoke MBLS in a second step; and (iii) a preliminary significance test for (a subset of) the regressors, which does not have asymptotic impact on our feasible inference.

We label the combination of (i)-(iii) the LCMB-augmented inference procedure and establish its theoretical properties 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, as well as substantial bias in the coefficient estimates. Importantly, and in contrast, we show that our LCMB-augmented procedure displays a desirable combination of correct size, non-trivial power and bias robustness in general settings.

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 LCMB-augmented procedure, we find key state variables, such as the default spread and treasury interest rates, 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

¹⁰These issues are not treated in our companion paper Andersen & Varneskov (2021*b*) either, which consider testing for parameter instability in persistent predictive relations, adopting the setting of Andersen & Varneskov (2021*a*).

asymptotic limit theory and feasible inference, conditional on being able to estimate the return persistence. Section 4 introduces our LCMB approach to persistence estimation and preliminary significance tests. Section 5 contains the simulation study, and Section 6 provides the empirical analysis of return predictions. Finally, Section 7 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 (2021a) 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 by the former.

2.1 Predictive System and Assumptions

Suppose we observe a $(k + 1) \times 1$ vector $\mathbf{Z}_t = (y_t, \mathbf{x}'_{t-1})'$ at times $t = 1, \dots, n$, where y_t is the asset returns and \mathbf{x}_{t-1} is a vector of candidate predictors, which are stipulated to have a multi-component structure,

$$\mathbf{x}_{t-1} = \mathbf{x}_{t-1} + \mathbf{c}_{t-1}, \quad \mathbf{x}_t \perp\!\!\!\perp \mathbf{c}_s, \quad \text{for all } t, s, \quad (6)$$

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

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

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

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

where $\Gamma(\cdot)$ is the gamma function.¹¹ In this setting, in which all variables may exhibit high degrees of persistence, the predictive relation between y_t and the observable regressors \mathbf{x}_{t-1} will be defined through the weakly dependent components of the unobservable persistent signals \mathbf{x}_{t-1} . Specifically, we assume,

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

¹¹Formal assumptions on the components of the system are stated below.

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

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

where $\mathbf{Q}(L) = \mathbf{D}_x(L)(1 - L)^{-d_1}$, with $\mathbf{D}_x(L)$ being the $k \times k$ lower-right submatrix of the fractional filtering matrix $\mathbf{D}(L)$, $a = \mu_y - \mathbf{B}' \mathbf{Q}(L) \boldsymbol{\mu}_x$ for $\boldsymbol{\mu} = (\mu_y, \boldsymbol{\mu}_x')'$ as well as the innovations,

$$v_t = (1 - L)^{b-d_1} \eta_t \in I(d_1 - b) \quad \text{and} \quad \xi_{t-1}^{(-d_1)} = (1 - L)^{-d_1} \xi_{t-1} \in I(d_1).$$

Unfortunately, as discussed in detail below, this regression is latent, with three nonstandard layers disguising the predictive relation. First, we observe \mathbf{x}_{t-1} , not the persistent signals \mathbf{x}_{t-1} nor the measurement errors \mathbf{c}_{t-1} . Second, the error ξ_{t-1} cloaks the conditional mean, similarly to an omitted regressor. Third, the persistence of the variables, measured by $\mathbf{D}(L)$, is unknown a priori.

Despite these challenges, and assuming a latent predictive relation, it is important to realize that the system (6)-(10) encompasses most multivariate fractionally integrated systems in the literature. To see this, suppose $\mathbf{c}_{t-1} = \mathbf{0}$ 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, captured by \mathbf{x}_{t-1} , 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, \mathbf{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. Moreover, the regression model is balanced, regardless of the forecasting prowess of the regressors, that is, $y_t \in I(d_1)$ under both $\mathcal{H}_0 : \mathbf{B} = \mathbf{0}$ and $\mathcal{H}_A : \mathbf{B} \neq \mathbf{0}$. 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, \mathcal{H}_A , the fractional filter adjusts the persistence of the “latent” signals, \mathbf{x}_{t-1} , to ensure regression balance. This allows d_1 to differ from the persistence of the observable candidate predictors, $\{d_2, \dots, d_{k+1}\}$. However, under \mathcal{H}_A , we implicitly think of the predictive relation as being between y_t and a “persistence transformed” signal, $\tilde{\mathbf{x}}_{t-1} \equiv \mathbf{Q}(L) \mathbf{x}_{t-1}$ rather than \mathbf{x}_{t-1} . Of course, in the special case, where $\mathbf{Q}(L) = \mathbf{I}_k$, a k -dimensional identity matrix, this adjustment is negligible.

In the next section, we discuss these points, provide examples and draw parallels to the extant literature, particularly Pastor & Stambaugh (2009) and Andersen & Varneskov (2021a). Before proceeding, however, we impose some formal structure on the system. The regularity conditions mirror those imposed by Andersen & Varneskov (2021a) 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 distinct 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 $\mathbf{q}_t = (\mathbf{u}'_{t-1}, \eta_t)'$ and $\boldsymbol{\zeta}_{t-1} = (\mathbf{c}'_{t-1}, \xi_{t-1})'$ rather than \mathbf{v}_t , when exploring the asymptotic properties for the LCM procedure below.

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

Assumption D2. *\mathbf{q}_t is a linear process, $\mathbf{q}_t = \sum_{j=0}^{\infty} \mathbf{A}_j \boldsymbol{\epsilon}_{t-j}$, with coefficient matrices $\sum_{j=0}^{\infty} j^{1/2} \|\mathbf{A}_j\| < \infty$, the innovations satisfy, almost surely, $\mathbb{E}[\boldsymbol{\epsilon}_t | \mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t | \mathcal{F}_{t-1}] = \mathbf{I}_{k+1}$, 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 , $\mathbb{P}[\|\mathbf{q}_t\| > c] \leq C\mathbb{P}[\zeta > c]$. Finally, the periodogram of $\boldsymbol{\epsilon}_t$ is denoted by $\mathbf{J}(\lambda)$.*

Assumption D3. *For $\mathbf{A}(\lambda, i)$, the i -th row of $\mathbf{A}(\lambda) = \sum_{j=0}^{\infty} \mathbf{A}_j e^{ij\lambda}$, its partial derivative satisfies $\|\partial \mathbf{A}(\lambda, i) / \partial \lambda\| = O(\lambda^{-1} \|\mathbf{A}(\lambda, i)\|)$ as $\lambda \rightarrow 0^+$, for $i = 1, \dots, k+1$.*

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 conditions equivalent to those imposed by Assumption D1-D3 on \mathbf{q}_t . These assumptions are formally stated in Appendix A.1 for brevity. Moreover, the following, additional, conditions are required:*

- (a) *The co-spectrum between $\boldsymbol{\zeta}_{t-1}$ and η_t satisfies $\mathbf{f}_{\zeta\eta}(\lambda) \sim \mathbf{G}_{\zeta\eta}$, as $\lambda \rightarrow 0^+$, where the finite and non-random vector $\mathbf{G}_{\zeta\eta}$ may have non-zero entries.*
- (b) *$\mathbf{u}_t \perp \boldsymbol{\zeta}_s$ for all $t, s \geq 1$.*
- (c) *If the i -th element of $\boldsymbol{\zeta}_{t-1}$ is trivial, i.e., if $\boldsymbol{\zeta}_{t-1}(i) = 0$ for all $t \geq 1$, then the co-spectrum condition $\mathbf{G}_{\zeta\eta}(i) = \mathbf{G}_{\zeta\zeta}(i, g) = \mathbf{G}_{\zeta\zeta}(g, i) = 0$ for $g = 1, \dots, k+1$, is naturally also required.*

Assumption M. *Let $0 \leq d_1 \leq 1$ and $0 \leq d_i \leq 2$ for all $i = 2, \dots, k+1$. 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 $\mathbf{f}_{qq}(\lambda)$ as $\lambda \rightarrow 0^+$, which depends on the smoothness parameter $\varpi \in (0, 2]$. In addition, D1 requires locally full rank of \mathbf{u}_{t-1} and it being locally exogenous to η_t as $\lambda \rightarrow 0^+$, but not global exogeneity. Finally, condition D2 specifies linearity, martingale and moment conditions for \mathbf{q}_t , allowing for general multivariate dependence among the variables, thus accommodating flexible lead-lag and predictive structures.¹²

¹²Note that Assumption D2 implicitly restricts elements of the lower $k \times (k+1)$ submatrix of \mathbf{A}_0 to be equal to zero since the weakly dependent predictor components, \mathbf{u}_{t-1} , do not depend on the lead-one innovation sequence, $\boldsymbol{\epsilon}_t$. Moreover, the summability condition in D2 is slightly stronger than the square summability condition in Andersen & Varneskov (2021a), which is necessary to derive the asymptotic properties for our new LCM-bias estimator in Section 4.

Whereas D1 allows the latent predictive signals, \mathbf{x}_{t-1} , to exhibit mild endogeneity (as $\lambda \rightarrow c > 0$) through \mathbf{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) \mathbf{c}_{t-1} and the innovations η_t , and these may, furthermore, 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 observable variables can be explosive, $d_i > 1$. However, as explained above, if the persistence of the regressors deviate from d_1 , this will require a transformation such that their unobservable persistent signals $\tilde{\mathbf{x}}_{t-1} \equiv \mathcal{Q}(L)\mathbf{x}_{t-1}$ comply with balance of the predictive relation (10). In general, 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 involving multiple financial and macroeconomic variables.

Finally, we impose $b = d_1$, which implies $v_t = \eta_t$ and, consequently, that the return prediction model exhibits (fractional) cointegration, if $\xi_{t-1} = 0$ for all t . 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 \mathbf{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 $\mathbf{c}_{t-1}(i) = 0, \forall t \geq 1$, 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 6 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 characterized 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 genuine empirical foundation.*

Remark 2. *The endogenous, weakly dependent measurement errors $\mathbf{c}_{t-1} \in I(0)$ are of strictly lower order than \mathbf{x}_{t-1} , since $\underline{d}_x > 0$. However, this does not imply that they are irrelevant for the asymptotic analysis. They will, in fact, distort the inference, unless we impose strict bounds on the tuning parameters for the LCM estimator. In fact, our limit theory developed below dictates a scaling of these*

by a function of the sample size, n . However, when $\underline{d}_x > 0$ is small, the wedge between the signal and the noise is small and, since $\mathbf{f}_{\zeta\eta}(\lambda) \sim \mathbf{G}_{\zeta\eta}$ by Assumption C, the endogeneity bias will dominate the Gaussian limit theory, unless appropriately accounted for. We discuss sufficient conditions for eliminating the endogeneity bias, as we develop the theoretical analysis.

2.2 Return Regressions: Dynamics and Implications

The predictive system (6)-(10) has several distinct features. First, the interaction between persistence and regression balance warrants additional discussion. Specifically, the mapping between the latent persistent components of the observable predictors, \mathbf{x}_{t-1} , and the corresponding modified signal, $\tilde{\mathbf{x}}_{t-1}$ ensures that the predictor signal is transformed to match the persistence of the (latent) conditional mean. To appreciate the mechanics of the filter $\mathbf{Q}(L)$, consider a scenario where the conditional mean component has $d_1 = 0.8$. If we observe a single candidate regressor, whose predictive signal x_{t-1} has $d_x = 1.8$, a regular linear specification is incompatible with the conditional mean dynamics. However, our framework accommodates a well-defined predictive relation between y_t and the dampened (first-differenced) signal $\tilde{x}_{t-1} = \mathbf{Q}(L)x_{t-1} = (1 - L)x_{t-1}$ with memory $d_{\tilde{x}} = d_x - 1 = 0.8$. In other words, we may think of $\mathcal{X}_{t-1} = x_{t-1} + c_{t-1}$ as the observable – raw – predictor, for which a predictive relation may exist between y_t and $\tilde{\mathcal{X}}_{t-1} = (1 - L)\mathcal{X}_{t-1}$ via \tilde{x}_{t-1} .¹³ Importantly, the transformation of the signal persistence required to be compatible with the conditional mean dynamics need not be known a priori, but is instead implemented automatically as an integral part of the LCM procedure.

Second, under the alternative hypothesis, \mathcal{H}_A , where the regressors, or a subset thereof, exhibit significant predictive power for y_t , these 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 component, but fail to fully span its variation. For example, this arises when a significant regressor has been omitted from \mathcal{X}_{t-1} . In contrast, if the predictors are “perfect”, we have $\varphi_{t-1} = \mathbf{B}'\mathbf{u}_{t-1}$.

Third, the system accommodates endogenous regressors through, \mathbf{c}_{t-1} , which is independent of the persistent signal, \mathbf{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 rather than a persistent first-order autoregressive (AR) system with half-lives 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 accommodate a second component in both factors, which is less persistent, but may exhibit non-trivial correlation with the return innovations. These components are not informative

¹³More generally, our setting accommodates different and fractional persistence among the signals \mathbf{x}_{t-1} , which, as a result, necessitates different adjustments via $\mathbf{Q}(L)$. However, common for all is that they must match the persistence of the conditional mean, i.e., have fractional integration order d_1 after transformation. Note that also includes cases with $d_x < d_1$, whose persistence needs to be “enhanced”. For example, if $d_x = 0.2$ and $d_1 = 0.6$, the signal must be aggregated with weights given by the fractional memory differential $d_1 - d_x$.

about the conditional equity premium, but they facilitate richer system dynamics.¹⁴

Fourth, the model facilitates non-trivial correlation between unspanned component of the conditional mean, ξ_{t-1} , and the observable explanatory variables (again, through \mathbf{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 for asset returns to possess a weakly dependent component η_t , which may be “highly volatile” relative to the persistent conditional mean, thereby generating a “low” signal-to-noise ratio in the return regression and rendering predictability hard to detect empirically. This feature is consistent with a comprehensive literature that finds limited return serial correlation, yet predictive power from highly persistent financial and macroeconomic variables; see, e.g., Welch & Goyal (2008), Lettau & Ludvigson (2010) and the 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 3. *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,*

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

where $0 < \phi < 1$, the eigenvalues of \mathbf{A} are inside the unit circle, and the innovation vector $(\eta_t, w_t, \mathbf{u}_t)'$ 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 $\varphi_{t-1} \neq a_\varphi + \mathbf{B}' \mathbf{x}_{t-1}$. Moreover, if the predictors are imperfect, this generates unspanned return persistence, as captured in our setup via the inclusion of $\xi_{t-1}^{(-d_1)}$ in equation (10). Finally, their key identifying assumption for \mathbf{B} is $0 < \phi < 1$, allowing the persistent conditional mean to be disentangled from the noise. If this assumption fails, they require 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) $\mathbf{B} = \mathbf{0}$ and ξ_{t-1} is trivial, $\forall t = 1, \dots, n$; returns are not predictable.
- (ii) $\mathbf{B} = \mathbf{0}$ and ξ_{t-1} non-trivial, $\forall t = 1, \dots, n$; returns are not predictable by \mathbf{x}_{t-1} .
- (iii) $\mathbf{B} \neq \mathbf{0}$ and ξ_{t-1} non-trivial, $\forall t = 1, \dots, n$; returns are predictable, and \mathbf{x}_{t-1} is “imperfect”.
- (iv) $\mathbf{B} \neq \mathbf{0}$ and ξ_{t-1} is trivial, $\forall t = 1, \dots, n$; returns are predictable, and \mathbf{x}_{t-1} is “perfect”.

¹⁴A multi-component structure of the conditional mean of consumption growth is consistent with the dynamic decomposition in, e.g., Ortú, 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.

Hypotheses (i) and (ii) imply that \mathbf{X}_{t-1} possesses no predictive power for returns, but they have distinct dynamic implications; namely, returns are $I(0)$ and $I(d_1)$, respectively. Moreover, the first hypothesis stipulates that returns are not predictable by *any* strongly dependent regressor, whereas the second allows for predictability through a strongly persistent regressor, but a “wrong” or “incomplete” set of predictors is being examined. The extensive empirical and theoretical literature on return predictability suggests that, in many settings, we should focus on the null hypothesis (ii) rather than (i), especially if examining a set of predictors sequentially in single-regressor models, where issues with omitted regressors loom large. Hypotheses (iii) and (iv) also carry distinct dynamic implications. Both imply $y_t \in I(d_1)$, but (iii) has regression errors 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 regimes for persistent variables, for which standard OLS delivers spurious inference; see, e.g., Granger & Newbold (1974), Phillips (1987), and Tsay & Chung (2000). Of course, if we knew that y_t and $\mathbf{X}_{t-1} = \mathbf{x}_{t-1}$ form a fractionally cointegrated system – i.e., the signals are significant, observable and “perfect” – we may apply inference procedures explicitly designed for such systems with strongly persistent variables, see, e.g., Robinson & Marinucci (2003), Robinson & Hualde (2003), Christensen & Nielsen (2006) and Johansen & Nielsen (2012). However, generally, we do not know, a priori, which of the hypotheses capture the given inference scenario, i.e., whether the regressors are endogenous and/or the predictors are “perfect”, and we must estimate the persistence of \mathbf{z}_t , which is hampered by the “low” signal-to-noise ratio for the returns.

As discussed in Remark 3 and Section 1, 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. Compared with Andersen & Varneskov (2021a), we allow for “imperfect” predictors and the joint presence of endogeneity and cointegration.¹⁵ Hence, all subsequent results are new.

2.3 The Local Spectrum Approach

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

$$\mathbf{f}_{\mathcal{X}\mathcal{X}}(\lambda) \simeq \mathbf{\Lambda}_{xx}^{-1} \mathbf{G}_{uu} \overline{\mathbf{\Lambda}}_{xx}^{-1} + \mathbf{G}_{cc}, \quad (11)$$

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

¹⁵ Andersen & Varneskov (2021a) study the asymptotic properties of LCM approach in a general predictive setting. However, when examining the effect of regressor endogeneity on the inference, they rule out cointegration.

for λ close to zero, where $\bar{\Lambda}_{yy}$ and $\bar{\Lambda}_{xx}$ are the complex conjugates of Λ_{yy} and Λ_{xx} , defined as,

$$\Lambda_{yy} = (1 - e^{i\lambda})^{d_1}, \quad \Lambda_{xx} = \text{diag} \left[(1 - e^{i\lambda})^{d_2}, \dots, (1 - e^{i\lambda})^{d_{k+1}} \right].$$

These decompositions are intuitive. First, $\mathbf{f}_{\mathcal{X}\mathcal{X}}(\lambda)$ shares the multi-component structure of the observable regressors \mathbf{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, $\mathbf{f}_{\mathcal{X}y}(\lambda)$ contains information about the forecasting prowess of the regressors through \mathbf{B} , and this effect, captured in the first term, dominates the remaining ones at lower frequency ordinates, i.e., as $\lambda \rightarrow 0^+$. Third, $\mathbf{G}_{c\xi} \bar{\Lambda}_{yy}^{-1}$ and $\mathbf{G}_{c\eta}$ capture endogeneity-induced bias terms. They stem from the potentially non-trivial co-spectra between the regressor measurement errors, \mathbf{c}_{t-1} , and then ξ_{t-1} and η_t , respectively, denoting the conditional mean (imperfection) errors and return innovations. Since the conditional mean component may be persistent (when $d_1 > 0$), the first bias term may diverge, as $\lambda \rightarrow 0^+$, but at a slower rate than the first term, since $\underline{d}_x > 0$. Finally, the co-spectra $\mathbf{f}_{x\xi}^{(-d_1)}(\lambda)$ and $\mathbf{f}_{x\eta}(\lambda)$ introduce sampling errors for estimators of \mathbf{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 at rates depending on the integration orders of the predictors and asset returns. In contrast, the co-spectral densities for the unobserved weakly dependent components of the predictive system, \mathbf{u}_{t-1} and e_t , are,

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

which both are asymptotically bounded, yet convey equivalent information about \mathbf{B} . This suggests that inference based on the stationary components, \mathbf{u}_{t-1} and e_t may circumvent issues regarding balance, degeneracy of point estimates and spurious inference, motivating Andersen & Varneskov (2021a) to introduce the LCM procedure, consisting of two main steps. First, the procedure *fractionally filters* the observed variables $\mathbf{Z}_t = (y_t, \mathbf{X}'_{t-1})'$ to obtain an estimate of $\mathbf{v}_t = (e_t, \mathbf{u}'_{t-1})'$. 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. To retain flexibility by allowing for different estimators of the fractional integration orders, we abstain from dedicating a specific estimator and, instead, assume to have one available, \hat{d}_i for $i = 1, \dots, k+1$, 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 $i = 1, \dots, k+1$ elements of \mathbf{z}_t , we assume to have an estimator with the property,

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

Assumption F is very mild, essentially only requiring the existence of an estimator which, under appropriate assumptions on equation (7), is consistent. However, since we accommodate both (asymptotic)

totically) 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 (2021a), 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, $\widehat{\mathbf{D}}(L)$, the estimates for \mathbf{v}_t are,

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

where $\widehat{\mathbf{u}}_{t-1}^c = \widehat{\mathbf{u}}_{t-1} + \widehat{\mathbf{c}}_{t-1}$, with $\widehat{\mathbf{u}}_{t-1} = \widehat{\mathbf{D}}_x(L) \mathbf{x}_{t-1}$ and $\widehat{\mathbf{c}}_{t-1} = \widehat{\mathbf{D}}_x(L) \mathbf{c}_{t-1}$.¹⁶ Similarly, we define $\widehat{\mathbf{v}}_t \equiv (\widehat{e}_t, \widehat{\mathbf{u}}_{t-1}')'$, which is the equivalent (albeit, unobservable) estimate of \mathbf{v}_t , without an endogenous component in the regressors. Using frequency domain techniques, we may then extract asymptotically similar information from $\widehat{\mathbf{v}}_t$ and $\widehat{\mathbf{v}}_t^c$. 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, $\widehat{\mathbf{D}}(L) \boldsymbol{\mu} \mathbf{1}_{\{t \geq 1\}}$, in a unified manner during second stage estimation.

Step 2: Medium band least squares estimation. We estimate and draw inference about $\boldsymbol{\beta}$ using a frequency domain least squares estimator and $\widehat{\mathbf{v}}_t^c$. To define the former, we let \mathbf{h}_t and \mathbf{k}_t be generic (and compatible) vector time series, $\lambda_j = 2\pi j/n$ denote the Fourier frequencies, and write

$$\mathbf{w}_h(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \mathbf{h}_t e^{it\lambda_j}, \quad \overline{\mathbf{w}}_k(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \mathbf{k}_t e^{-it\lambda_j}, \quad \mathbf{I}_{hk}(\lambda_j) = \mathbf{w}_h(\lambda_j) \overline{\mathbf{w}}_k(\lambda_j), \quad (15)$$

for discrete Fourier transforms (DFTs) and their corresponding cross-periodogram, respectively. Moreover, we define the trimmed discretely averaged co-periodogram (TDAC), using the real part of the cross-periodogram, indicated by $\Re(\mathbf{I}_{hk}(\lambda_j))$, as,

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

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

$$\widehat{\boldsymbol{\beta}}_c(\ell, m) = \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\widehat{\mathbf{u}}}^c(\ell, m)^{-1} \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\widehat{e}}^c(\ell, m), \quad (17)$$

for which $\ell, m \rightarrow \infty$ and $\ell/m + m/n \rightarrow 0$, as $n \rightarrow \infty$. 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 $m/n \rightarrow 0$, equation (17) turns out to be first-order

¹⁶Consistent with the Type II nature of the fractional model in (7), the operator $\widehat{\mathbf{D}}(L)$ uses all available observations.

equivalent to,

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

that is, the corresponding estimator based on $\widehat{\mathbf{v}}_t$. In other words, trimming and a local bandwidth suffice to annihilate biases resulting from 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 \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 challenging 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 this conjecture in both our simulation study and the empirical analysis below. Consequently, the next section develops (feasible) inference theory for LCM that holds uniformly for the four model hypotheses (i)-(iv), conditional on Assumption F being satisfied, thus tackling the issues surrounding imperfect regressors. Subsequently, in Section 4, we provide a new LCM-bias approach to estimating d_1 , that relies on regression slopes and, in the process, develop a pre-test for regressor significance, which circumvents the need to estimate d_1 and has no asymptotic impact on the inference.

3 Limit Theory for LCM when \widehat{d}_1 is Consistent

This section derives asymptotic limit theory for LCM, assuming Assumption F applies. That is, we consider a scenario, where a well-behaved consistent estimator for the persistence of the conditional mean component in asset returns is available. The theory extends Andersen & Varneskov (2021a) by allowing for “imperfect” predictors and the simultaneous presence of endogeneity and cointegration. Moreover, we introduce a new feasible inference procedure that works uniformly across all competing hypothesis (i)-(iv), without prior knowledge of which scenario applies to the inference problem.

3.1 Asymptotic Central Limit Theory

Our development of the LCM inference procedure below focuses mostly on the scenarios (ii)-(iv), but applies equally well to model (i). The main distinction is, then, between the imperfect regressor case, models (ii)-(iii), versus the perfect regressor case, model (iv). Moreover, although Assumption M imposes $b = d_1$, we state the conditions on the tuning parameters in terms of b to facilitate comparison

with Andersen & Varneskov (2021a).

Finally, we let “ \asymp ” signify that the ratio of two terms converges to a positive constant and then summarize the main constraints on the trimming and bandwidth parameters as follows.

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 $n \rightarrow \infty$,*

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

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 information loss from frequency trimming; 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.¹⁷

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 obtained when $\underline{d}_x - b = 0$. Intuitively, we require stronger trimming to alleviate the endogeneity bias and obtain the same asymptotic efficiency in the presence of endogenous regressors, if the excess persistence of the system is small. As detailed in Remark 2, this occurs when the wedge between the asymptotic orders of the persistent signals and endogenous measurement errors is small.

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

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

¹⁷We 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.

When consistent estimation of d_1 is feasible, Theorem 1 proves that the LCM procedure is asymptotically Gaussian for the predictive models with imperfect regressors, (ii) and (iii), as well as the cointegration model, (iv). The asymptotic distribution theory differs, however. 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). In contrast, if the regressors are “perfect”, ξ_{t-1} is absent, 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 the efficiency of the MBLS estimator, in analogy with super consistency properties.

Despite the different limit theory for models (ii)-(iii) versus (iv), it remains Gaussian, 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. Similar uniformity does not apply for OLS, the narrow band least squares (NBLS) estimator, or maximum likelihood inference in fractionally cointegrated VAR models, where inference exhibits varying forms of non-Gaussianity in non-stationary cases; see Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Johansen & Nielsen (2012).¹⁸ Likewise, 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 1 are due to the fractional filtering, rendering the approach, after eliminating various errors and biases through trimming, reminiscent of the ELW inference in Shimotsu & Phillips (2005).

Moreover, the limiting theory of the LCM procedure is correctly centered, and thus free from biases induced by persistent and endogenous regressors, 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 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. Importantly, this can be achieved without prior knowledge of the scenario, that describes the system and without pre-testing. We introduce such a feasible procedure in the following section.

Remark 4. *We impose $0 < d_1 \leq 1$ in Theorem 1, but we can accommodate $d_1 = 0$, with appropriate changes to the asymptotic variance for both models (ii)-(iii) and (iv). In particular, for the former, we must replace $G_{\xi\xi}$ with $G_{\xi\xi} + G_{\eta\eta}$. Moreover, since cointegration no longer features in model (iv), given $b = d_1 = 0$, the result is readily obtained by setting $b = 0$ in the limit theory, implying a lower convergence rate. Similar comments apply to all subsequent results.*

¹⁸Such 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 theories for these alternatives rely on the presence of cointegration. Finally, as demonstrated by Andersen & Varneskov (2021a, 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.

Remark 5. *The condition $\underline{d}-b \geq 0$ is equivalent to the “balanced cointegration” requirement in Andersen & Varneskov (2021a, Eq. (8)), implying that the fractional cointegration cannot be stronger than the persistence of the regressors. This condition is not truly required, but it simplifies the tuning parameter restrictions imposed on ℓ and m in Assumption T considerably.*

3.2 Feasible Inference

This section provides a feasible inference procedure, which is uniformly valid across scenarios (ii)-(iv) where, as shown by Theorem 1, we obtain a different asymptotic limit theory depending on whether the system features cointegration in model (iv) or not. Feasible inference requires consistent estimators of the long-run covariance matrix \mathbf{G}_{uu} and either $G_{\xi\xi}$ for models (ii)-(iii) or $G_{\eta\eta}/(1+2b)$ for model (iv). Similarly, we must account for the different rates of converge in the distinct inference regimes. The main challenges are, first, that we observe $\hat{\mathbf{v}}_t^c$ instead of either $\hat{\mathbf{v}}_t$ or \mathbf{v}_t . Second, the residual components ξ_{t-1} and $\eta_t^{(b)}$ are latent, and we must estimate them, after estimation of \mathbf{B} , as,

$$\hat{\eta}_t^{(b,c)} = \hat{e}_t - \hat{\mathbf{B}}_c(\ell, m)' \hat{\mathbf{u}}_{t-1}^c. \quad (19)$$

Importantly, and unlike Andersen & Varneskov (2021a), we do not rely on a separate, consistent estimator of b , but rather utilize the different asymptotic orders in the two sources of error. Specifically, if the regressors are “imperfect,” as in models (ii)-(iii), the contribution from ξ_{t-1} will dominate that from $\eta_t^{(b)}$, $b = d_1 > 0$, asymptotically and we may, thus, use $\hat{\eta}_t^{(b,c)}$ to estimate $G_{\xi\xi}$. In contrast, for model (iv), we recover information about the long-run variance of $\eta_t^{(b)}$, since $\xi_{t-1} = 0, \forall t$.

Once the error process $\hat{\eta}_t^{(b,c)}$ is estimated, we use a class of trimmed long-run covariance estimators to obtain the asymptotic variance components. In particular, for a generic vector \mathbf{h}_t , we adopt,

$$\hat{\mathbf{G}}_{hh}(\ell_G, m_G) = \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \Re(\mathbf{I}_{hh}(\lambda_j)), \quad (20)$$

with 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 (2021a), but with the regression errors $\hat{\eta}_t^{(b,c)}$ defined differently, and it is akin to those in Christensen & Varneskov (2017). If we restrict $\ell_G = 1$, 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 $\eta_t^{(b)}$, when we seek to recover information about ξ_{t-1} in models (ii)-(iii).

Specifically, the components of the respective asymptotic variances in Theorem 1 for models without cointegration (ii)-(iii) and the corresponding model with cointegration (iv) are computed using equation (20) with either the fractionally filtered regressors, $\hat{\mathbf{u}}_{t-1}^c$, or the estimated residuals, $\hat{\eta}_t^{(b,c)}$. These

estimates, denoted by $\widehat{\mathbf{G}}_{\widehat{u}\widehat{\eta}}^c(\ell_G, m_G)$ and $\widehat{G}_{\widehat{\eta}\widehat{\eta}}^{(b,c)}(\ell_G, m_G)$, are, then, combined to form,

$$\widehat{\mathbf{AVAR}}(\ell_G, m_G) = \widehat{\mathbf{G}}_{\widehat{u}\widehat{\eta}}^c(\ell_G, m_G)^{-1} \widehat{G}_{\widehat{\eta}\widehat{\eta}}^{(b,c)}(\ell_G, m_G)/(2m), \quad (21)$$

which, importantly, is identical for all models (ii)-(iv). We need to impose (mild) conditions on its tuning parameters, similarly to those for the LCM coefficient estimator in Assumption T.

Assumption T-G. *Let the bandwidth $m_G \asymp n^{\kappa_G}$ and $\ell_G \asymp n^{\nu_G}$, with $0 < \nu_G < \kappa_G < \varrho \leq 1$, and define $\underline{m}_n = m_d \wedge m_b \wedge m$. Then, $m = m_G$ and the following cross-restrictions are imposed,*

$$\frac{n}{m_G \ell_G^2} + \frac{n^2}{m_G \ell_G^2 \underline{m}_n} + \left(\frac{n}{m_G} \right)^b \frac{1}{\sqrt{m_d}} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Whereas the first two regularity conditions mirror those imposed by Andersen & Varneskov (2021a, Assumption T-G), the restriction $m_G = m$ and the third condition are unique to our asymptotic variance estimator. They are necessary for our feasible inference in model (iv), as we must estimate the long-run variance $G_{\eta\eta}/(1+2b)$ and recover information about the scale λ_m^{2b} , arising from the faster rate of convergence of LCM in this scenario. In principle, we only need the restriction $m_G = m$ for $\widehat{G}_{\widehat{\eta}\widehat{\eta}}^{(b,c)}(\ell_G, m_G)$, but we maintain it for both components to limit the number of tuning parameters.

Theorem 2. *Suppose Assumption T-G and the conditions of Theorem 1 hold, then*

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

Importantly, the estimator converges to the correct asymptotic variance in both inference regimes. Hence, feasible inference and testing for the LCM procedure follows by applying Theorems 1 and 2 in conjunction with the continuous mapping theorem and Slutsky's theorem:

Corollary 1. *Suppose the conditions of Theorems 1 and 2 hold, then*

$$\widehat{\mathbf{AVAR}}(\ell_G, m_G)^{-1/2} \left(\widehat{\mathbf{B}}_c(\ell, m) - \mathbf{B} \right) \xrightarrow{\mathbb{D}} N(\mathbf{0}, \mathbf{I}_k), \quad \text{under models (ii)-(iv).}$$

Our feasible inference procedure is uniformly valid across scenarios (ii)-(iv), without prior knowledge of the which scenario describes the system and without pre-testing, as long as Assumption F holds and we can estimate d_1 consistently. Hence, it is not subject to the model selection and pre-testing critique of Leeb & Pötscher (2005), despite carrying out estimation in two steps.

Remark 6. *We provide a model selection procedure in Appendix A.2 to determine, asymptotically, whether the regressors are imperfect or perfect, that is, whether models (i)-(iii) or (iv) describes the predictive system. While this may be of separate interest, we reiterate that our feasible inference is uniformly valid across the model (ii)-(iv).*

4 An LCM-Bias Approach to Estimating d_1

This section introduces a new LCM-bias (LCMB) approach to estimation of the fractional integration (persistence) parameter for asset returns. We emphasize, however, that it applies more generally to problems where a persistent conditional mean is cloaked by a high noise-to-signal ratio. The approach is inspired by the following observations. Assumption F is arguably violated, when applying standard univariate time series techniques to returns, generating estimates $\hat{d}_1 \simeq 0$. At the same time, a large literature finds predictability using very persistent state variables, suggesting that the mean returns contain a highly persistent component. As an alternative, we propose a new *multivariate* approach, where we “borrow” an estimate of the fractional integration order from the regressors and impose it on the returns. Importantly, this estimate need not be correct. Instead, we take advantage of LCM and the frequency domain analysis to quantify the *bias* from having imposed the (possibly, wrong) integration order and, subsequently, utilize this bias to quantify how incorrect our initial guess is. This allows us to back out an estimate of d_1 . Hence, our procedure relies on LCM coefficient estimates, rather than univariate techniques. In addition, we provide new LCM-based significance tests for biased coefficient estimates, which are asymptotically valid, even without estimating the return persistence.

4.1 A Slightly Modified Framework

In order to develop our multivariate LCMB estimator for d_1 , we impose some additional mild structure on the estimation problem by modifying Assumptions M and F:

Assumption M- d_1 . *In addition to Assumption M, suppose a subset of the regressors, denoted by \mathcal{S}_x , with dimension $1 \leq k_x \leq k$ has $0 \leq d_i \leq 1$, for $i \in \mathcal{S}_x$. Moreover, define $\gamma_x = \max_{i \in \mathcal{S}_x}(d_i)$ and suppose that $b = d_1$, $0 < b < 1$ and $b \leq \gamma_x$.*

Assumption F- d_1 . *Suppose Assumption F holds for elements $i = 2, \dots, k+1$ of \mathbf{Z}_t , i.e., the regressors. Furthermore, we have an estimator for which $\hat{\gamma}_x - \gamma_x = O_p(1/\sqrt{m_d})$.*

The additional restrictions in Assumption M- d_1 are rather innocuous; they are satisfied when a “sensible” set of regressors have been chosen, with persistence spanning a realistic range for the conditional mean return. Moreover, Assumption F- d_1 only requires that we are able to estimate the persistence of the regressors and a function there-off; namely, the *maximal* persistence on the unit interval for the subset \mathcal{S}_x . We shall subsequently utilize $\hat{\gamma}_x$ as the initial “guess” for the return persistence but, importantly, do not require that $\gamma_x = d_1$. Instead, we define,

$$\psi = \gamma_x - d_1, \quad \text{such that} \quad 0 \leq \psi \leq \gamma_x. \quad (22)$$

Thus, ψ quantifies the distance between our initial guess and the true persistence of the conditional mean return. The memory restrictions imply $0 < \underline{d}_x \leq \gamma_x \leq 1$ which, in conjunction with $0 < b < 1$, is critical for obtaining the Gaussian central limit theory. We conjecture the condition $b \leq \gamma_x$ may be relaxed, but this may impact the limit theory, as discussed in the following remark.

Remark 7. The condition $\psi \geq 0$ greatly simplifies the asymptotic analysis, especially the tuning parameter restrictions. However, inspired by the analyses of NBLS and MBLS in stationary fractional cointegration settings by Christensen & Nielsen (2006) and Christensen & Varneskov (2017), we conjecture that a Gaussian limit theory can be obtained, if our initial persistence guess $\hat{\gamma}_x$ implies $\psi > -1/2$. Similarly, the study of NBLS in non-stationary environments by, e.g., Robinson & Marinucci (2001, 2003) suggests that having $-1 \leq \psi \leq -1/2$ will induce a non-Gaussian limit theory. Thus, we may be able to entertain other initial guesses for the return persistence. We do not investigate such choices theoretically here, but we explore alternative selections through simulation in Section 5.

4.2 Multivariate LCMB Limit Theory

The main idea behind the LCMB approach is to compare the relative bias of two LCM coefficient estimates with different bandwidths using $\hat{\gamma}_x$ as an initial benchmark value for the return persistence. To this end, we define a constant $\varkappa > 1$, such that $\tilde{m} = m/\varkappa > \ell$. The bandwidth \tilde{m} diverges at the same rate as m , but letting $\varkappa > 1$ enables us to asymptotically identify ψ . We further define,

$$\hat{\mathcal{B}}_c(\ell, m, \hat{\gamma}_x) = \hat{\mathbf{F}}_{\tilde{u}\tilde{u}}^c(\ell, m)^{-1} \hat{\mathbf{F}}_{\tilde{u}\varepsilon}^c(\ell, m), \quad \text{where } \check{e}_t = (1 - L)^{\hat{\gamma}_x} y_t, \quad (23)$$

as the LCM estimator implemented using $\hat{\gamma}_x$ as the value for d_1 , and equivalently write $\hat{\mathcal{B}}_c(\ell, \tilde{m}, \hat{\gamma}_x)$. Besides the additional structure imposed on the estimation problem in Assumptions M- d_1 and F- d_1 , we also require some strengthening of the regularity conditions on the tuning parameters.

Assumption T- d_1 . Suppose the conditions of Assumption T hold as well as

$$\begin{aligned} \frac{m^{1+2(\varpi-2)}}{n^{2(\varpi-b)}} + \frac{\ell^{1+\varpi+\psi}}{m^{1/2+b+\psi} n^{\varpi-b}} + \frac{1}{\ell} \left(\left(\frac{n}{m} \right)^{1/2+b} + \left(\frac{n}{m} \right)^{1/2+\gamma_x-d_x} \right) \\ + \frac{m^{1/2}}{\ell^2} \left(\frac{n}{m} \right) \left(\left(\frac{m}{n} \right)^{d_x-b} + \left(\frac{m}{n} \right)^{2d_x-\gamma_x} \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

The regularity conditions in Assumption T- d_1 tighten the corresponding restrictions one through four in Assumption T, especially if b is large. To see this, note that condition one in Assumption T- d_1 and five in Assumption T imply $b/(1/2 + b) < \kappa < 2(\varpi - b)(1 + 2(\varpi - b))$, thus inducing a tight bound around the bandwidth selection $\kappa = 2/3$, if $\varpi = 2$ and b is close to one. This explains why we require $b < 1$, and why we impose $2b < \varpi$ for Assumptions D1, T and T- d_1 to be consistent with Theorem 3 below. Hence, if b is high, yet below one, the conditions suggest to select $0.65 < \kappa < 0.75$. In contrast, conditions two and four of Assumption T- d_1 provide no meaningful strengthening of the trimming restrictions. However, condition three may be binding. For example, if we set $\kappa = 0.70$, the condition implies $0.30(1/2 + b) < \nu$, which may bind for b close to one.

Finally, we write $c(\psi) = \cos(\pi\psi/2)/(1 + \psi)$ to characterize the LCM bias and, similarly,

$$\mathbf{V}(\eta, \gamma_x) = \mathbf{G}_{uu}^{-1} \frac{G_{\eta\eta}}{2(1 + 2\gamma_x)}, \quad \Phi(\varkappa, \gamma_x) = \begin{pmatrix} 1 & \varkappa^{\gamma_x - 1/2} \\ \varkappa^{\gamma_x - 1/2} & \varkappa^{2\gamma_x - 1} \end{pmatrix}, \quad (24)$$

with $\mathbf{V}(\xi, \psi)$ and $\Phi(\varkappa, \psi)$ defined analogously, to decompose the asymptotic variance.

Theorem 3. *Suppose Assumptions D1-D3, C, M-d₁, F-d₁ and T-d₁ hold along with the conditions $0 < b < 1$, $b \leq \underline{d}$, $n^{1/2}/m \rightarrow 0$, and $\max(0, (1 - 3\kappa/2)/(1 + \kappa/2), 2b) < \varpi \leq 2$, then,*

(a) *In the imperfect regressor models (ii) and (iii), we have,*

$$\sqrt{m} \lambda_m^{-\psi} \begin{pmatrix} \widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x) - \lambda_m^\psi c(\psi) \mathbf{B} \\ \widehat{\mathbf{B}}_c(\ell, \widetilde{m}, \widehat{\gamma}_x) - \lambda_{\widetilde{m}}^\psi c(\psi) \mathbf{B} \end{pmatrix} \xrightarrow{\mathbb{D}} N(\mathbf{0}, \mathbf{V}(\xi, \psi) \otimes \Phi(\varkappa, \psi)).$$

(b) *In the cointegration model (iv), it holds that,*

$$\sqrt{m} \lambda_m^{-\gamma_x} \begin{pmatrix} \widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x) - \lambda_m^{\gamma_x} c(\gamma_x) \mathbf{B} \\ \widehat{\mathbf{B}}_c(\ell, \widetilde{m}, \widehat{\gamma}_x) - \lambda_{\widetilde{m}}^{\gamma_x} c(\gamma_x) \mathbf{B} \end{pmatrix} \xrightarrow{\mathbb{D}} N(\mathbf{0}, \mathbf{V}(\eta, \gamma_x) \otimes \Phi(\varkappa, \gamma_x)).$$

Theorem 3 provides some interesting findings. First, the asymptotic properties of the LCM coefficient estimates are similar to those described following Theorem 1; the distribution is asymptotically Gaussian irrespective of regressor endogeneity, imperfection and persistence. However, the convergence rate now depends on ψ for models (ii)-(iii) and γ_x for the cointegration model (iv). Second, the estimates are inconsistent, unless $\psi = 0$, that is, unless our initial guess for d_1 is correct. Indeed, if $\psi = 0$, we recover a multivariate version of Theorem 1. Third, the relative efficiency of $\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x)$ and $\widehat{\mathbf{B}}_c(\ell, \widetilde{m}, \widehat{\gamma}_x)$ depends on whether the system is stationary in the respective modeling regimes. For example, under model (iv), the asymptotic variance of $\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x)$ is smaller than the corresponding one for $\widehat{\mathbf{B}}_c(\ell, \widetilde{m}, \widehat{\gamma}_x)$, when $\gamma_x > 1/2$, and vice versa for $\gamma_x < 1/2$. This feature is reminiscent of NBLS being more efficient than OLS, when estimating non-stationary cointegration models, which is due to the signal being stronger at lower frequencies; see, e.g., Robinson & Marinucci (2001).

Finally, and importantly, the LCM coefficient estimates shrink towards their limits at different rates, depending on the bandwidths m and \widetilde{m} . This is the feature, which allows us to recover ψ and subsequently d_1 . Specifically, this type of asymptotic degeneracy suggests that we may extract such estimates from the *ratio* of biased LCM coefficients. To this end, let \mathcal{L} be a $k \times 1$ vector, which is used to linearly transform $\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x)$ and, thus, asymptotically $\lambda_m^\psi c(\psi) \mathbf{B}$ to a scalar. We will, then, first assume that $\mathcal{L}' \mathbf{B} \neq 0$ to construct our ratio estimator for ψ . However, this assumption may be tested prior to implementation, and we propose a pre-test in the next subsection, that has no impact on the final (conservative) convergence rate of the estimator. This pre-test is a valid significance test, implemented using our benchmark γ_x for the return persistence. Hence, our procedure requires at least one of the regressors to have predictive power for returns, i.e., if all selected regressors are insignificant,

we cannot detect return persistence via the LCMB approach. This will, however, asymptotically, be revealed by the pre-test. Since numerous asset pricing theories imply that the conditional mean return is driven by, e.g., valuation ratios, discount rates, volatility, etc., these should be included among the regressors to ensure that $\mathcal{L}'\mathcal{B} \neq 0$, indeed, is satisfied, with high probability.

Now, assuming $\mathcal{L}'\mathcal{B} \neq 0$ holds, our ratio estimator of ψ is defined as,

$$\hat{\psi}(\mathcal{L}) = \frac{\ln \mathcal{R}(\mathcal{L})}{\ln(\varkappa)}, \quad \text{where} \quad \mathcal{R}(\mathcal{L}) = \frac{\mathcal{L}'\hat{\mathcal{B}}_c(\ell, m, \hat{\gamma}_x)}{\mathcal{L}'\hat{\mathcal{B}}_c(\ell, \tilde{m}, \hat{\gamma}_x)}, \quad (25)$$

motivated by the structure of the asymptotically degenerate coefficient estimates. Further letting,

$$\mathcal{S}(\mathcal{L}, \eta, \gamma_x) = \mathcal{L}'\mathcal{V}(\eta, \gamma_x)\mathcal{L}, \quad \Theta(\varkappa, \psi, \gamma_x) = (1 - \varkappa^{\psi + \gamma_x - 1/2})^2 / (c(\psi)\mathcal{L}'\mathcal{B})^2,$$

and defining $\mathcal{S}(\mathcal{L}, \xi, \psi)$ and $\Theta(\varkappa, \psi, \psi)$ analogously, we may thus establish a limit theory for $\hat{\psi}(\mathcal{L})$.

Theorem 4. *Suppose the conditions of Theorem 3 and $\mathcal{L}'\mathcal{B} \neq 0$ hold. Then,*

$$\begin{cases} \sqrt{m}(\hat{\psi}(\mathcal{L}) - \psi) \xrightarrow{\mathbb{D}} N\left(0, \frac{\mathcal{S}(\mathcal{L}, \xi, \psi)\Theta(\varkappa, \psi, \psi)}{\ln(\varkappa)^2}\right), & \text{in models (ii)-(iii), when } \psi \neq 1/4, \\ \sqrt{m}\lambda_m^{-b}(\hat{\psi}(\mathcal{L}) - \psi) \xrightarrow{\mathbb{D}} N\left(0, \frac{\mathcal{S}(\mathcal{L}, \eta, \gamma_x)\Theta(\varkappa, \psi, \gamma_x)}{\ln(\varkappa)^2}\right), & \text{in model (iv), when } \psi + \gamma_x \neq 1/2. \end{cases}$$

Theorem 4 shows that our LCMB estimator for $\psi = \gamma_x - d_1$ inherits its desirable asymptotic properties from the LCM procedure, including its rate of convergence. Specifically, for the imperfect regressor models (ii)-(iii), we obtain the semiparametric rate $\hat{\psi}(\mathcal{L}) - \psi = O_p(1/\sqrt{m})$, when $\psi \neq 1/4$, and $\hat{\psi}(\mathcal{L}) - \psi = o_p(1/\sqrt{m})$, when $\psi = 1/4$. Moreover, for the cointegration model (iv), the rate is improved to $\hat{\psi}(\mathcal{L}) - \psi = O_p(1/\sqrt{m}\lambda_m^b)$, when $\psi + \gamma_x \neq 1/2$, and $\hat{\psi}(\mathcal{L}) - \psi = o_p(1/\sqrt{m}\lambda_m^b)$, if $\psi + \gamma_x = 1/2$, in analogy with super consistency. Since we have $\hat{\gamma}_x - \gamma_x = O_p(1/\sqrt{m_d})$ by Assumption F- d_1 , Theorem 4 shows that $\hat{d}_1(\psi) \equiv \hat{\gamma}_x - \hat{\psi}(\mathcal{L})$ obtains a conservative rate $\hat{d}_1(\psi) - d_1 \leq O_p(1/\sqrt{m})$. Critically, if our LCMB persistence estimator is implemented with a preliminary bandwidth $m_p \asymp n^{\kappa_p}$, where $\kappa < \kappa_p < \varrho \leq 1$, it follows that our original Assumption F is satisfied, our central limit theory in Theorem 1 holds, and that we may use the feasible inference procedure from Corollary 1 for hypothesis testing on \mathcal{B} . The one remaining caveat is pre-testing the assumption $\mathcal{L}'\mathcal{B} \neq 0$.

4.3 Preliminary Significance Testing

As a final step, this section provides an asymptotically valid significance test for $\mathcal{L}'\mathcal{B} \neq 0$, when the LCM procedure is implemented using $\hat{\gamma}_x$ as benchmark value for d_1 . Importantly, this test does not require knowledge of the return persistence, and the rate of convergence is fast enough to retain a conservative convergence rate $\hat{d}_1(\psi) - d_1 \leq O_p(1/\sqrt{m})$. Therefore, Assumption F is satisfied, if a preliminary bandwidth with $\kappa < \kappa_p$ is adopted. Hence, we are able to implement the second (MBLS) step of the LCM procedure following our pre-testing as well as LCMB estimation of d_1 , without being

subject to the Leeb & Pötscher (2005) model selection critique. As before, we define,

$$\check{\eta}_t^{(b,c)} = \check{e}_t - \widehat{\mathbf{B}}_c(\ell, m, \hat{\gamma}_x)' \hat{\mathbf{u}}_{t-1}^c, \quad \widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}(\ell_G, m_G) = \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \Re \left(\mathbf{I}_{\check{\eta}\check{\eta}}^{(b,c)}(\lambda_j) \right), \quad (26)$$

as well as the asymptotic variance estimator,

$$\widehat{\mathbf{AVAR}}(\ell_G, m_G, \hat{\gamma}_x) = \widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^c(\ell_G, m_G)^{-1} \widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b,c)}(\ell_G, m_G)/(2m). \quad (27)$$

While the intuition and motivation behind the estimator (27) is similar to the one in equation (21), their asymptotic properties differ, as evidenced by the following theorem.

Assumption T-G- d_1 . Suppose Assumption T-G and $(n/m_G)^\psi/\sqrt{m_G} \rightarrow 0$ hold.

Theorem 5. Suppose Assumption T-G- d_1 and the conditions of Theorem 3 hold, then

$$m \lambda_m^{-2\psi} \widehat{\mathbf{AVAR}}(\ell_G, m_G, \hat{\gamma}_x) \xrightarrow{\mathbb{P}} \begin{cases} \mathbf{V}(\xi, \psi), & \text{under model (ii),} \\ \mathbf{V}(\xi, \psi) + \mathbf{G}_{uu}^{-1} \left(\frac{1}{1+2\psi} - c(\psi)^2 \right) \mathbf{B}' \mathbf{G}_{uu} \mathbf{B}, & \text{under model (iii),} \\ \mathbf{G}_{uu}^{-1} \left(\frac{1}{1+2\psi} - c(\psi)^2 \right) \mathbf{B}' \mathbf{G}_{uu} \mathbf{B}, & \text{under model (iv).} \end{cases}$$

The result for model (ii) in Theorem 5 is the same as in Theorem 2, but the remaining parts are different. Specifically, when $\mathbf{B} \neq \mathbf{0}$ under the alternative hypothesis, \mathcal{H}_A , the inconsistency of the LCMB coefficient estimator $\widehat{\mathbf{B}}_c(\ell, m, \hat{\gamma}_x)$ spills over to the residual estimate, $\check{\eta}_t^{(b,c)}$, with a rate inherited from the degeneracy rate of the coefficient estimate in Theorem 3. In model (iii), this bias is of the same asymptotic order as $\xi_{t-1}^{(\gamma_x - d_1)}$ after fractional filtering, explaining the decomposition of terms. In model (iv), it dominates $\eta_t^{(\gamma_x)}$ and, thus, generates the first-order limit. This implies that our asymptotic variance estimator is inconsistent in models (iii) and (iv). However, it is still useful for testing $\mathcal{L}'\mathbf{B} \neq 0$ and, thus, the significance of \mathbf{B} more generally. To see this, define,

$$\widehat{\mathfrak{T}}(\mathcal{L}) = \frac{\mathcal{L}' \widehat{\mathbf{B}}_c(\ell, m, \hat{\gamma}_x)}{\sqrt{\mathcal{L}' \widehat{\mathbf{AVAR}}(\ell_G, m_G, \hat{\gamma}_x) \mathcal{L}}}, \quad \text{where } \widehat{\mathfrak{T}}(\mathcal{L}) \xrightarrow{\mathbb{D}} N(0, 1), \quad (28)$$

under model (ii), by Theorems 3(a) and 5, the continuous mapping theorem and Slutsky's theorem. Moreover, $\widehat{\mathfrak{T}}(\mathcal{L}) = O_p(\sqrt{m})$ in both models (iii) and (iv), as the λ_m^ψ scaling cancels. Hence, the test $\widehat{\mathfrak{T}}(\mathcal{L})$ has asymptotic power to detect the significance of the regressors without any actual estimate for the return persistence. However, this robustness likely comes with a loss of finite sample power, as Theorem 5 shows $\widehat{\mathbf{AVAR}}(\ell_G, m_G, \hat{\gamma}_x)$ is inflated for model (iii), and the rate of divergence is strictly slower than the rate $\sqrt{m}\lambda_m^{-b}$, achieved by LCM in Theorem 1, for model (iv). Nonetheless, importantly, using $\widehat{\mathfrak{T}}(\mathcal{L})$ for the pre-test, if indeed $\mathcal{L}'\mathbf{B} \neq 0$, we preserve a conservative convergence rate,

$$\widehat{d}_1(\psi, \mathbb{P}) - d_1 = O_p(1/\sqrt{m}), \quad \text{for } \widehat{d}_1(\psi, \mathbb{P}) = \widehat{d}_1(\psi) \times \mathbb{P}(\widehat{\mathfrak{T}}(\mathcal{L}) \neq 0), \quad (29)$$

To summarize, we recommend the following recipe for implementing the LCM procedure for return regressions. First, carry out a significance test $\widehat{\mathfrak{T}}(\mathcal{L})$ for the regressor(s) using a preliminary guess $\widehat{\gamma}_x$ for the return persistence and a bandwidth m_p satisfying $m/m_p \rightarrow 0$, as $n \rightarrow \infty$. If we fail to find any significant regressor(s), we cannot estimate d_1 . Thus, we terminate the procedure, and conclude that our finding is consistent with models (i) or (ii), because $\widehat{\mathfrak{T}}(\mathcal{L})$ is asymptotically valid, even in the absence of an estimate for d_1 . In contrast, if we conclude that $\widehat{\mathfrak{T}}(\mathcal{L}) \neq 0$, we go on to implement the LCMB estimator for d_1 . Using the resulting estimate, we then carry out the LCM analysis in Section 3, as all requisite assumptions are satisfied. We refer to this as the LCMB-augmented procedure.

Remark 8. *While the LCMB persistence estimator in (29) exhibits a sufficiently fast rate of convergence to satisfy the consistency requirements in Assumption F, even following pre-testing, it is important to note that we cannot readily use Theorem 3 to draw inference on d_1 , if we carry out pre-testing. Specifically, since $\widehat{\mathfrak{T}}(\mathcal{L}) = O_p(\sqrt{m})$, the pre-testing errors for $\mathbb{P}(\widehat{\mathfrak{T}}(\mathcal{L}) \neq 0)$ are either of the same order as the sampling errors for $\widehat{d}_1(\psi)$ in Theorem 4(a), or it dominates the errors in Theorem 4(b). Therefore, this inference problem is subject to the Leeb & Pötscher (2005) critique. In contrast, when choosing $\kappa < \kappa_p$ for our second-stage MBLs estimator and feasible inference, the errors from the filtering estimates are ameliorated by suitable selection of the bandwidth and trimming function.*

5 Monte Carlo Evidence

This section illustrates some of the key inferential problems surrounding return regressions in a transparent numerical setting. In particular, 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 properties of predictability tests relying on either OLS, IVX or LCM inference. In particular, test size is assessed within an imperfect predictor specification. Moreover, we examine the bias and RMSE of the coefficient estimates for the three methods. Finally, we study the finite-sample properties of the LCM, with and without applying the LCMB procedure for estimation of the return persistence.

5.1 Simulation Setting

We study inference problems for return regressions in a setting reminiscent of the one in Hong (1996) and Shao (2009), albeit allowing the variables to exhibit non-stationary fractional integration. To render the analysis manageable, we assume \mathcal{B} and \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 ensures the signal of the persistent regressor is directly observable and excludes endogeneity. Then, we generate fractional ARMA(0,0) processes as,

$$(1-L)^{d_1}(y_t - \mu_y) = \varphi_{t-1} + \eta_t^{(d_1)}, \quad \varphi_{t-1} = \mathcal{B}u_{t-1} + \mathcal{B}_\xi \xi_{t-1}, \quad (1-L)^{d_2}(x_{t-1} - \mu_x) = u_{t-1}, \quad (30)$$

and $\eta_t^{(d_1)} = (1-L)^{d_1}\eta_t$, where $\zeta_t \sim \text{i.i.d. } N(0, \sigma_\zeta^2)$ 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 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_1, d_2) , $(\mathcal{B}, \mathcal{B}_\xi)$ and σ_η . We consider four distinct long-memory scenarios,

$$(d_1, d_2) \in \{(0.00, 0.45), (0.45, 0.45), (0.00, 0.80), (0.80, 0.80)\},$$

with different dynamic implications. In two cases, the data generating processes (DGPs) produce a balanced regression setting, with the regressor exhibiting the same persistence as the (latent) conditional mean return. In the first of these, we have an asymptotically stationary mean and in the other a non-stationary, yet mean-reverting, conditional mean process. In addition, we analyze two alternative DGPs, where the conditional mean is considerably less persistent ($d_1 = 0$) than the observed regressor. They are included to assess the impact of “overdifferencing,” when wrongly imposing an initial integration order, $\hat{\gamma}_x$, for the LCMB approach as well as to examine alternative specifications, when returns are weakly dependent. The magnitudes of the integration orders are inspired by our empirical study in the next section, with $d = 0.45$ corresponding to our realized volatility predictor and $d = 0.80$ being close to the estimated persistence for the conditional mean return using the LCMB approach (we obtain $\hat{d}_1(\psi, \mathbb{P}) \simeq 0.7676$). As $\sigma_\eta \in [0, 25]$ varies, the degree of noise in the predictive relation changes, possibly rendering the persistence of the conditional mean undetectable in finite samples.¹⁹

Initially, we entertain univariate predictions using x_{t-1} , but fix $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$ in equation (30), implying that asset returns has 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, $\hat{d}_1 \simeq 0$, 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 inference. Furthermore, we consider several different implementations of the LCM procedure: One based on “standard” univariate time series estimates of d_1 as in Andersen & Varneskov (2021a); an “oracle” version where (d_1, d_2) is treated as known; and four versions of the LCMB-augmented procedure with initial persistence values $\hat{\gamma}_x = \{0, 1/2, 1, \hat{d}_2\}$, thereby encompassing the approach advocated in Section 4 and fixed alternatives, as discussed in Remark 7. In addition, we study the case $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$, corresponding to a “perfect” predictor scenario, to assess the bias and RMSE properties of the respective estimators and the power of the LCM-based significance tests.²⁰ We emphasize that, in the latter scenario, OLS, IVX and the standard LCM estimators 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 with 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

¹⁹We have run similar experiments with ARMA(1, 0) short-run dynamics and/or $(d_1, d_2) = (0.45, 0.80)$. The corresponding results, when allowing for mild autoregressive dynamics as well as alternative long memory configurations, are almost identical to those presented in Figures 1-8 and are, thus, omitted for brevity.

²⁰We also considered DGPs with $\mathcal{B} = 1.2$ and $\mathcal{B}_\xi = 1.2$, which deliver similar, and therefore omitted, results.

instrument, an additional deterministic instrument $\sin((t-1)\pi/n)$, $t = 1, \dots, n$, 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.²¹ Moreover, we implement the standard LCM, oracle LCM and first-stage LCMB estimates of d_1 using trimming and bandwidth parameters $(\nu, \kappa) = (0.20, 0.70)$. These are similar to the ones considered in Andersen & Varneskov (2021a) and reflect the dynamic properties of returns and, especially, the persistent predictor variables in Section 6. Specifically, the bandwidth is chosen locally ($m/n \rightarrow 0$) to avoid placing excessive weight on the higher-frequency errors stemming from η_t , and we select $\kappa = 0.70$ due to the tight bounds derived for LCMB estimator, when $d_1 = b$ is high. Likewise, the trimming parameter reflects condition three in Assumption T with $d_1 = 0.80$, inspired by our estimate from the empirical analysis. Furthermore, for the standard LCM coefficient estimator and the LCMB approach, we obtain initial estimates of d_1 and d_2 using the TELW estimator of Andersen & Varneskov (2021a), with corresponding trimming and bandwidth parameters $\ell_d = \lfloor n^{0.3} \rfloor$ and $m_d = \lfloor n^{0.75} \rfloor$. The LCMB persistence estimate is, then, constructed using $\varkappa = 4/3$, thereby letting \tilde{m} be 75% of m , as,

$$\max \left(\hat{d}_1(\psi, \mathbb{P}), \hat{d}_1(\text{TELW}) \right),$$

that is, with a preliminary significance test for \mathcal{X}_{t-1} using Theorems 4 and 5 which, as detailed in Section 4.3, satisfy Assumption F with rate \sqrt{m} . Hence, when using the LCMB approach to extract information about d_1 , we implement the second-stage MBLS regression with $(\nu, \kappa) = (0.20, 0.69)$ for comparable asymptotic efficiency and to satisfy the requisite bandwidth conditions. We further note that, importantly, the results are qualitatively robust to varying the tuning parameters $\nu \pm 0.10$ and $\kappa \pm 0.05$ for both the standard LCM procedure and our LCMB-augmented approach. For the bandwidth, the tight bound is due to the LCMB persistence estimate in the first stage; the oracle version of LCM is robust to this selection. The significance tests for LCM and the LCMB-augmented procedure are carried out using the feasible inference in Section 3.2, where the consistent spectrum estimator of the asymptotic variance is implemented with $\nu_G = \nu$ and $\kappa_G = \kappa$. Our main analysis examines a sample size $n = 650$, mimicking the one for the empirical analysis ($n = 661$), but we also examine the limiting properties of our LCMB-augmented procedure using $n = 2000$. Finally, we consider a 5% nominal test size and rely on 1000 replications.

5.2 Simulation Results

Figure 1 shows the estimated integration order of returns as a function of σ_η for the baseline sample size $n = 650$ using either the TELW estimator (top panels) or the LCMB approach (middle panels). Moreover, we provide corresponding LCMB estimates in the bottom panel for the larger sample size, $n = 2000$. Several features are noteworthy. One, the estimated persistence decreases as a function of

²¹We have also carried out OLS-based testing for return predictability using Newey & West (1987) standard errors. The results are almost identical to those presented and, thus, left out for ease of exposition.

σ_η , eventually implying failure to reject $d_1 = 0$. This occurs, not surprisingly, more rapidly for the weaker signal, $d_1 = 0.45$, compared to the stronger signal, $d_1 = 0.80$, confirming that a persistent mean return component may be difficult to identify using standard univariate time series techniques. Two, the LCMB approach delivers sizable finite-sample improvements over TELW for a wide range of σ_η , except when adopting $\hat{\gamma}_x = 1$ as the initial persistence value. Three, whereas the persistence estimates from the TELW estimator barely changes, when increasing the sample size, the LCMB approach improves noticeably and is able to recover important information about d_1 for a considerably larger range of σ_η , corroborating the advantages of adopting a multivariate, frequency domain regression approach to persistence estimation within high noise-to-signal ratio settings.

Given that standard univariate time series techniques cannot identify the fractional integration order of the returns for a wide range of σ_η , we now examine the size implications. Figure 2 provides results for OLS, IVX, standard LCM and oracle LCM in its left panels and for the LCMB-augmented procedures in the right panels. Notably, OLS and IVX are oversized for a wide range of σ_η , if the conditional mean return is persistent, $d_1 = \{0.45, 0.80\}$, even when the estimated fractional integration order, incorrectly, suggests that the return series is $I(0)$. This is akin to the 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). We thus find that similar problems arise for return regressions with “imperfect” predictors in fractionally integrated settings for empirically relevant scenarios. For example, the bottom left panel features a regressor with a lower persistence than obtained, in Section 6, for common return predictors like the default spread, price-earnings ratio, and 3-month T-bill rate. Moreover, serious size distortions also arise for IVX, although this procedure, otherwise, is equipped to handle local-to-unity regressors. In contrast, the size properties of *all* LCM-based tests are close to the nominal 5% level.

Figures 3 and 4 depict the bias properties of the estimators under the null hypothesis $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$ and alternative $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$. None of the estimators exhibit any notable bias under the null. However, under the alternative $\mathcal{H}_A : \mathcal{B} \neq 0$, we observe that OLS and IVX are severely downward biased, when $d_1 = 0$, that is, when the conditional mean return is less persistent than the regressors. Conversely, the standard LCM estimator is upward biased when $d_1 = \{0.45, 0.80\}$. These patterns are intuitive. For OLS and IVX, the asymptotic order of the denominator (i.e., the predictor variance) is larger than numerator (the predictor-return covariance), causing the regression slope to be biased towards zero. We label this a *persistence bias*, which also features in our unreported results for the long memory configuration $(d_1, d_2) = (0.45, 0.80)$, as it is generated by the wedge between d_1 and d_2 , not their absolute magnitudes. Similarly, for the standard LCM estimator, the downward bias in the TELW estimates (eventually, obtaining $\hat{d}_1 \simeq 0$) causes the fractional filtering of y_t to be insufficient, leaving the numerator of higher asymptotic order than the denominator, where $\hat{\mathbf{u}}_{t-1}^c$ has been filtered “correctly”, and, thus, blows up the estimate, since $\mathcal{B} > 0$. This pattern is not visible for the oracle implementation of LCM. Hence, we refer to this as a *fractional filtering bias*.

Turning to the LCMB-augmented procedures, we obtain mixed results depending on the selection

of $\hat{\gamma}_x$. Choosing $\hat{\gamma}_x = 0$, the subsequent coefficient estimates are (almost) unbiased, when $d_1 = 0$, but produces only small improvements over the standard LCM estimator, when $d_1 = \{0.45, 0.80\}$. In contrast, if choosing $\hat{\gamma}_x = 1$, the patterns are reverse, providing strong bias reduction when $d_1 = \{0.45, 0.80\}$, but being downward biased for $d_1 = 0$. Hence, the selection $\hat{\gamma}_x = 0$ makes the LCMB-augmented procedure reminiscent of the standard LCM estimator, and $\hat{\gamma}_x = 1$ generates similarities with OLS and IVX. The advocated selection $\gamma_x = \hat{d}_2$, in turn, performs well in all cases; it is almost unbiased for $d_1 = 0$ and generally displays a reduced bias for $d_1 = \{0.45, 0.80\}$.

Finally, we compare the standard LCM estimator and our new LCMB-augmented procedure in terms of size and power in Figure 5 as well as bias and root-mean-squared error (RMSE) in Figure 6, with the LCMB selection $\gamma_x = \hat{d}_2$ throughout. Moreover, the corresponding results for the larger sample size, $n = 2000$, are provided in Figures 7 and 8. Whereas the size and power results are similar for the two procedures, major differences arise between their respective bias and RMSE properties, when $d_1 = \{0.45, 0.80\}$, as the LCMB-augmented procedure improves the finite sample performance for a wide range of σ_η . Perhaps more important, Figure 8 demonstrates that the former improves with sample size, further ameliorating its bias and lowering the RMSE, whereas the properties of the standard LCM procedure worsen, due to its fractional filtering bias. Hence, overall, the LCMB-augmented procedure displays a desirable combination of correct size, non-trivial power and bias robustness in finite samples. In contrast, OLS and IVX are either oversized, when the conditional mean is persistent, $d_1 = \{0.45, 0.80\}$, or suffers from a persistence bias when d_1 is smaller than d_2 .

6 Empirical Illustration: Forecasting Equity Market Returns

This section explores predictive regressions for monthly S&P 500 equity index returns involving popular macro-finance variables using OLS, IVX and LCMB-augmented procedures. Specifically, we examine the predictive content of the regressors in Bansal et al. (2014) and Campbell et al. (2018).

6.1 Data Description

We use monthly log-return and realized volatility (RV) series for the S&P 500 index over March 1960-March 2015, yielding $n = 661$ observations. As in, e.g., Andersen & Bollerslev (1998), Barndorff-Nielsen & Shephard (2002), and Andersen, Bollerslev, Diebold & Labys (2003), RV is constructed as the square-root of the cumulative daily squared returns each month. Following Campbell et al. (2018), we then include the default spread (DS), three-month U.S. Treasury bills (TB), and price-earnings ratio (PE) as additional regressors. They have all been deemed successful predictors of equity returns; see, e.g., Lettau & Ludvigson (2010) and Campbell (2018, Chapters 5.3-5.4). The DS equals the difference between the log-percentage yields on Moody's BAA and AAA bonds; TB is log-transformed; and PE is 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 from the Federal Reserve Bank of St. Louis (FRED), while the PE data stem from the website associated with Shiller (2000).

6.2 LCMB Analysis of Return Predictability

First, we estimate the fractional integration order of the returns and the four state variables; RV, DS, TB and PE. We adopt the TELW estimator of Andersen & Varneskov (2021a) and the exact local Whittle (ELW) estimator with a correction for the mean, or initial value, of Shimotsu & Phillips (2005) and Shimotsu (2010).²² The results, reported in Table 1, 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 preclude returns from having a hard-to-identify “latent” persistent conditional mean component.

Second, we proceed by implementing the LCMB approach to estimate d_1 and undertake preliminary significance testing. Specifically, we adopt the procedure in Section 4 with $\hat{\gamma}_x = \max_{i=2,\dots,5} = \hat{d}_i(\text{TELW}) = 1.0356$, since the latter is insignificantly different from one, and tuning parameters $\nu = \nu_G = 0.20$, $\kappa_p = \kappa_G = 0.70$ and $\varkappa = 4/3$, similarly to the simulation study. Moreover, since we want to avoid imposing ex-ante significance of the form $\mathcal{L}'\mathcal{B} \neq 0$, we carry out a preliminary multivariate LCM regression with $\hat{\gamma}_x$ as our initial guess for the conditional mean persistence. The results, reported in the top part of Table 2, show that DS is the only individually significant predictor among the regressors in the pre-test. Hence, we subsequently implement the LCMB estimator in (29) using univariate LCM regressions with DS, obtaining an estimate $\hat{d}_1(\psi, \mathbb{P}) = 0.7676$. This finding corroborates the hypothesis that asset returns contain a fractionally integrated conditional mean component, which we cannot identify using standard univariate time series techniques. The estimate implies a conditional mean process that is a strongly persistent, non-stationary, yet mean-reverting.

Next, we carry out a second-stage, multivariate MBLS analysis with $\kappa = \kappa_G = 0.69$ and the same trimming parameters and LCMB persistence estimate, $\hat{d}_1(\psi, \mathbb{P})$ as above. The results are included along with corresponding for OLS and IVX results, using, again, Eicker-White inference, in the bottom part of Table 2. There are several interesting observations. First, neither OLS nor IVX indicate *any* significant predictability, and their coefficient estimates, standard errors and Wald tests are similar. Second, using the LCMB-augmented procedure, we find similar (insignificant) coefficient estimates for RV and PE. However, we also find that DS and TB are significant at a 1% and 5% level, respectively, and the regressors are jointly significant at the 1% level. The positive coefficient sign for the former is consistent with a risk-return trade-off, and the negative for the latter aligns with return-valuation theory (Campbell 2018). Fourth, these results are confirmed by univariate regressions in Table 3, where the LCMB-augmented procedure deliver estimates that are similar in terms of sign, magnitude and significance, demonstrating robustness towards imperfect, or omitted, regressor bias.

These findings are considerably more incisive than usual for predictive return regressions, especially at short horizons, e.g., Welch & Goyal (2008), Lettau & Ludvigson (2010), Campbell (2018, Chapter 5) and references therein. We attribute this to the advantages of our LCMB-augmented procedure. First, uncovering the persistence of the conditional mean via our multivariate regression approach, we

²²The 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.

may adequately filter returns, reducing the impact of the “large” contemporaneous innovations, thus mitigating their asymptotic and finite sample effects. Second, by letting $\ell/m + m/n \rightarrow 0$ as $n \rightarrow \infty$, we further reduce the impact of 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 LCM procedure is robust to endogenous innovations, which tend to generate severe biases, e.g., Stambaugh (1999), Pastor & Stambaugh (2009) and Phillips & Lee (2013). These LCM features all mitigate critical attenuation biases, consistent with the larger coefficient estimates for DS and TB from the LCMB-augmented procedure than from OLS and IVX in Tables 2 and 3. In contrast, Andersen & Varneskov (2021a) find (standard) LCM 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 distinction across these applications is the much stronger signal-to-noise ratio for the RV measures, which alleviates concerns regarding unidentifiable integrated components in the return volatility series.

7 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 show that the LCM procedure is consistent and delivers asymptotically Gaussian inference, if we can estimate the persistence of the conditional mean return, d_1 . Furthermore, we provide a new LMCB estimator of d_1 that leverages biased slopes from (multivariate) regressions as well as new LCM-based tests for significance of (a subset of) the predictors, which are valid even without estimating the return persistence. By combining these parts into the LCMB-augmented procedure, we are able to estimate the return persistence, coefficient estimates and draw feasible inference in general settings. Simulations are used to illustrate the theoretical arguments as well as demonstrate favorable finite sample properties of the LCMB-augmented 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 LCMB-augmented procedure, we find strong predictive power for key economic state variables such as the default spread and treasury interest rates.

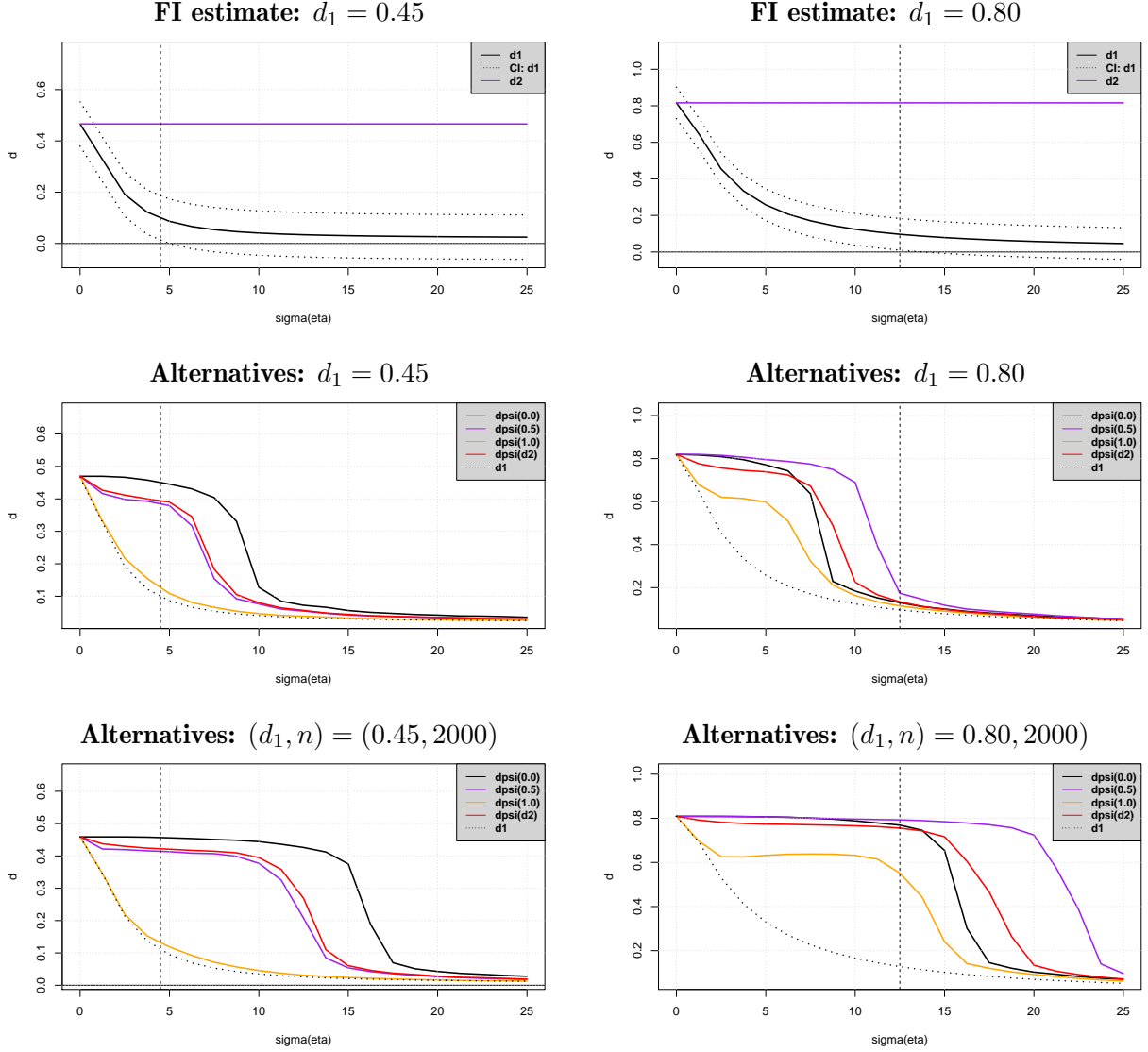


Figure 1: Fractional integration estimation. The top panels provide estimates of d_1 and d_2 as a function of the standard deviation of the weakly dependent component of returns, σ_η . Moreover, 95% confidence intervals are provided for \hat{d}_1 . The estimates are constructed using the TELW estimator with tuning parameters $\ell_d = \lfloor n^{0.3} \rfloor$ and $m_d = \lfloor n^{0.75} \rfloor$. The dotted vertical line highlights the value of σ_η where the empirical (unrestricted) estimate, \hat{d}_1 , is no longer significantly different from zero. The middle panels provide corresponding estimates based on the LCMB approach, leveraging biased coefficient estimates and the ratio $\hat{\psi}(L)$, as described in Sections 4-5, and uses a trimming parameter $\nu = \lfloor n^{0.20} \rfloor$ as well as a preliminary bandwidth $m_p = \lfloor n^{0.70} \rfloor$. The LCMB procedure is implemented with one of the persistence guesses $\hat{\gamma}_x = \{0.0, 0.5, 1.0, \hat{d}_2\}$ and indicated by “dpsi” in the figures. The top and middle panels consider a sample size $n = 650$. The bottom panel is similar to the middle panel, but consider a larger sample size $n = 2000$. Finally, the simulations is implemented with 1000 replications.

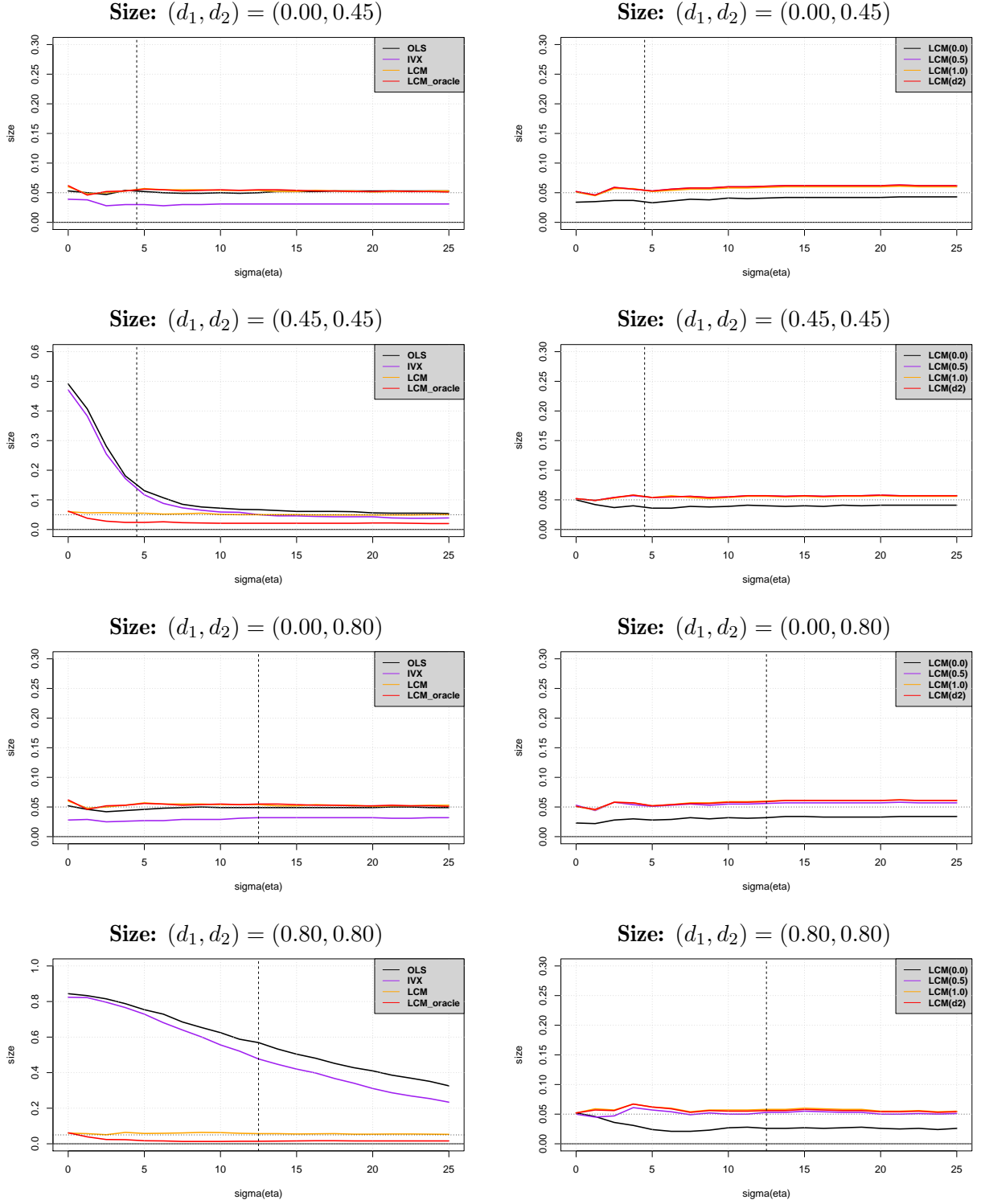


Figure 2: Size properties. The left panels provide the size of OLS, IVX, the LCM estimator of Andersen & Varneskov (2021a) and an oracle version of LCM, where (d_1, d_2) is treated as known. The right panels provide results for LCM estimators that use different LCMB estimates of d_1 in the fractional filtering, as described in Sections 4-5. The LCMB estimates are implemented with $\hat{\gamma}_x = \{0.0, 0.5, 1.0, \hat{d}_2\}$, as indicated by the parenthesis; see Section 5 and Figure 1 for details. Finally, we consider a sample size $n = 650$, a 5% nominal test size and use 1000 replications.

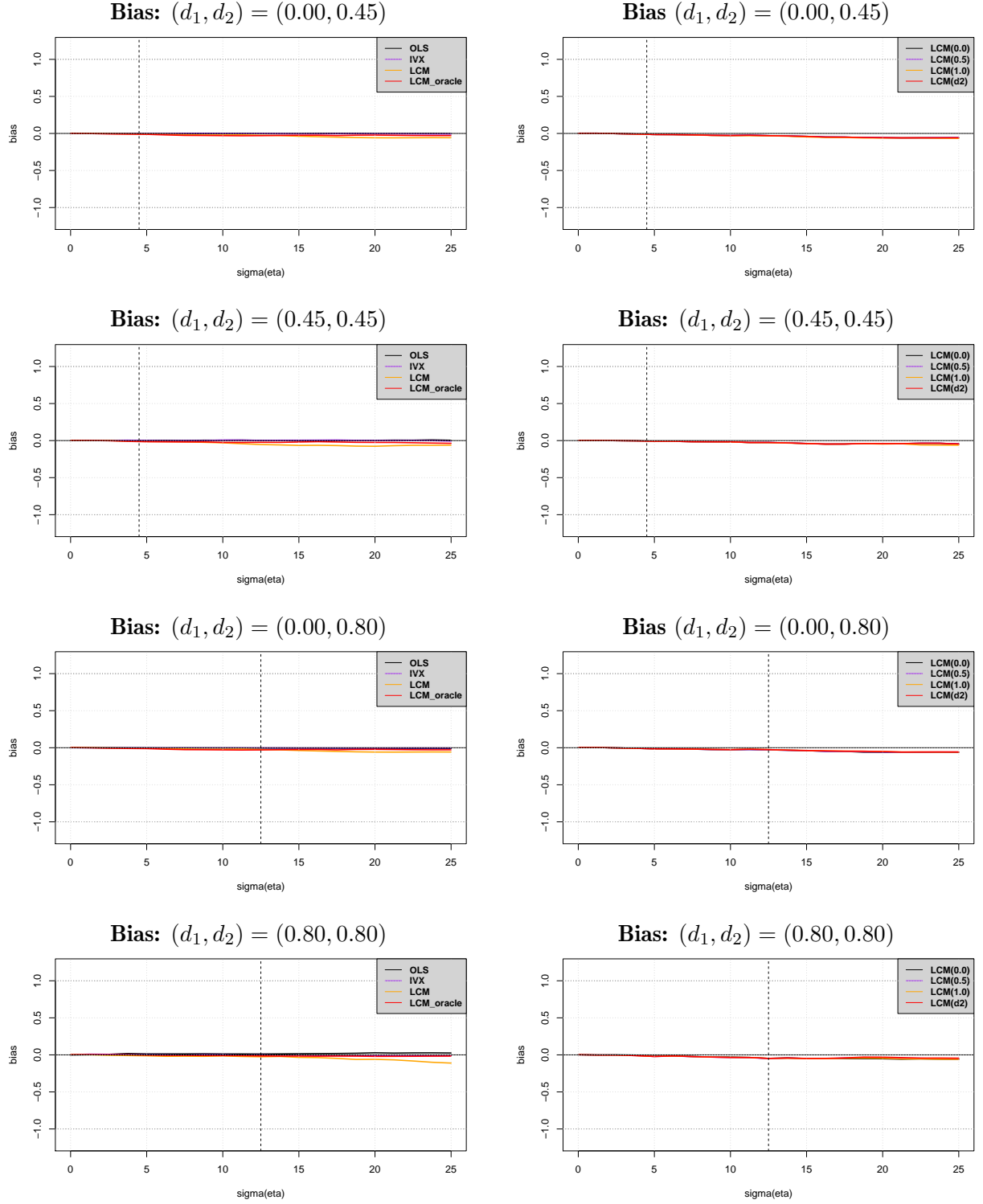


Figure 3: Bias properties: Null hypothesis. The left panels provide the bias of OLS, IVX, the LCM estimator of Andersen & Varneskov (2021a) and an oracle version of LCM, where (d_1, d_2) is treated as known. The right panels provide results for LCM estimators that use different LCMB estimates of d_1 in the fractional filtering, as described in Sections 4-5. The LCMB estimates are implemented with $\hat{\gamma}_x = \{0.0, 0.5, 1.0, \hat{d}_2\}$, as indicated by the parenthesis; see Section 5 and Figure 1 for details. Finally, we consider a sample size $n = 650$ and use 1000 replications.

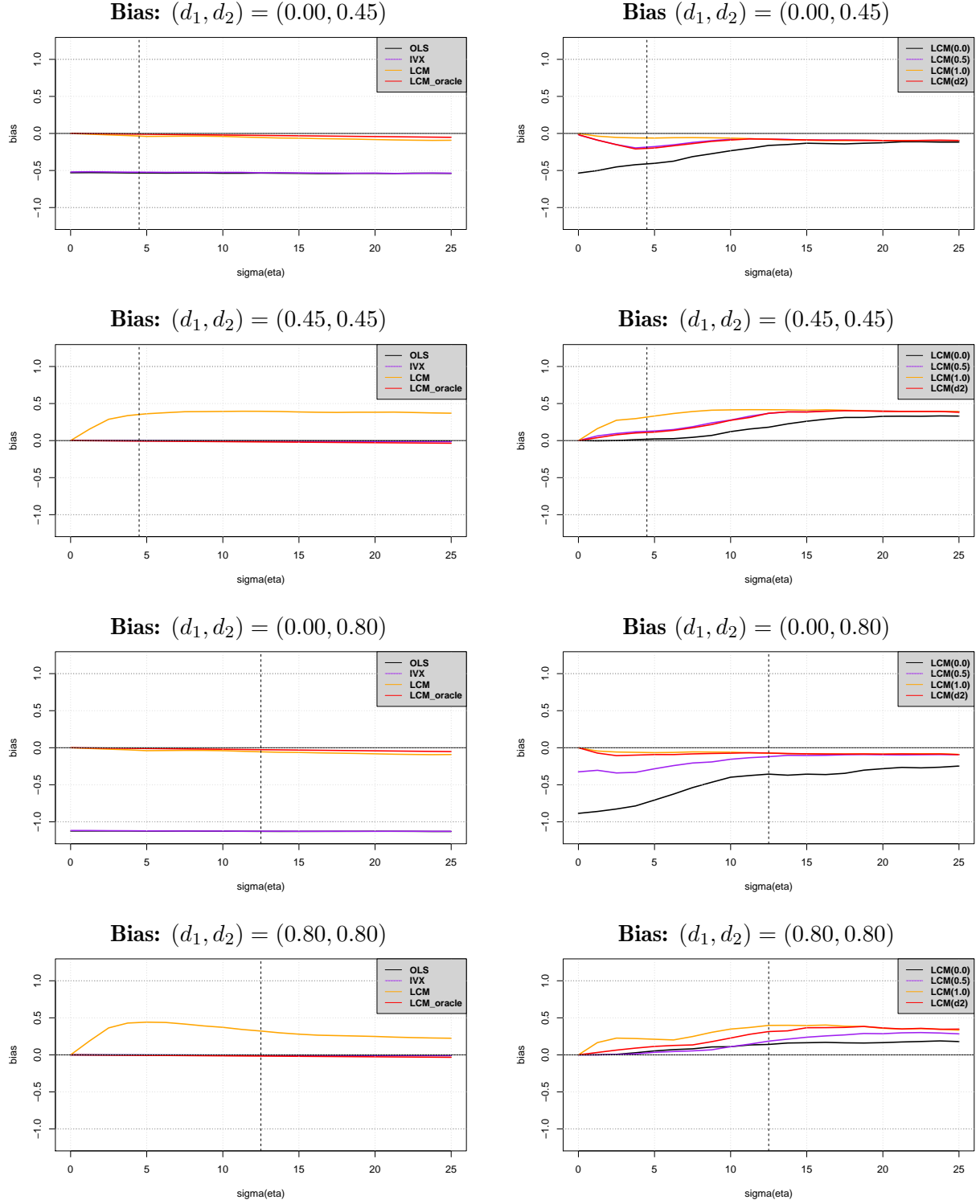


Figure 4: Bias properties: Alternative hypothesis. The left panels provide the bias of OLS, IVX, the LCM estimator of Andersen & Varneskov (2021a) and an oracle version of LCM, where (d_1, d_2) is treated as known. The right panels provide results for LCM estimators that use different LCMB estimates of d_1 in the fractional filtering, as described in Sections 4-5. The LCMB estimates are implemented with $\hat{\gamma}_x = \{0.0, 0.5, 1.0, \hat{d}_2\}$, as indicated by the parenthesis; see Section 5 and Figure 1 for details. Finally, we consider a sample size $n = 650$ and use 1000 replications.

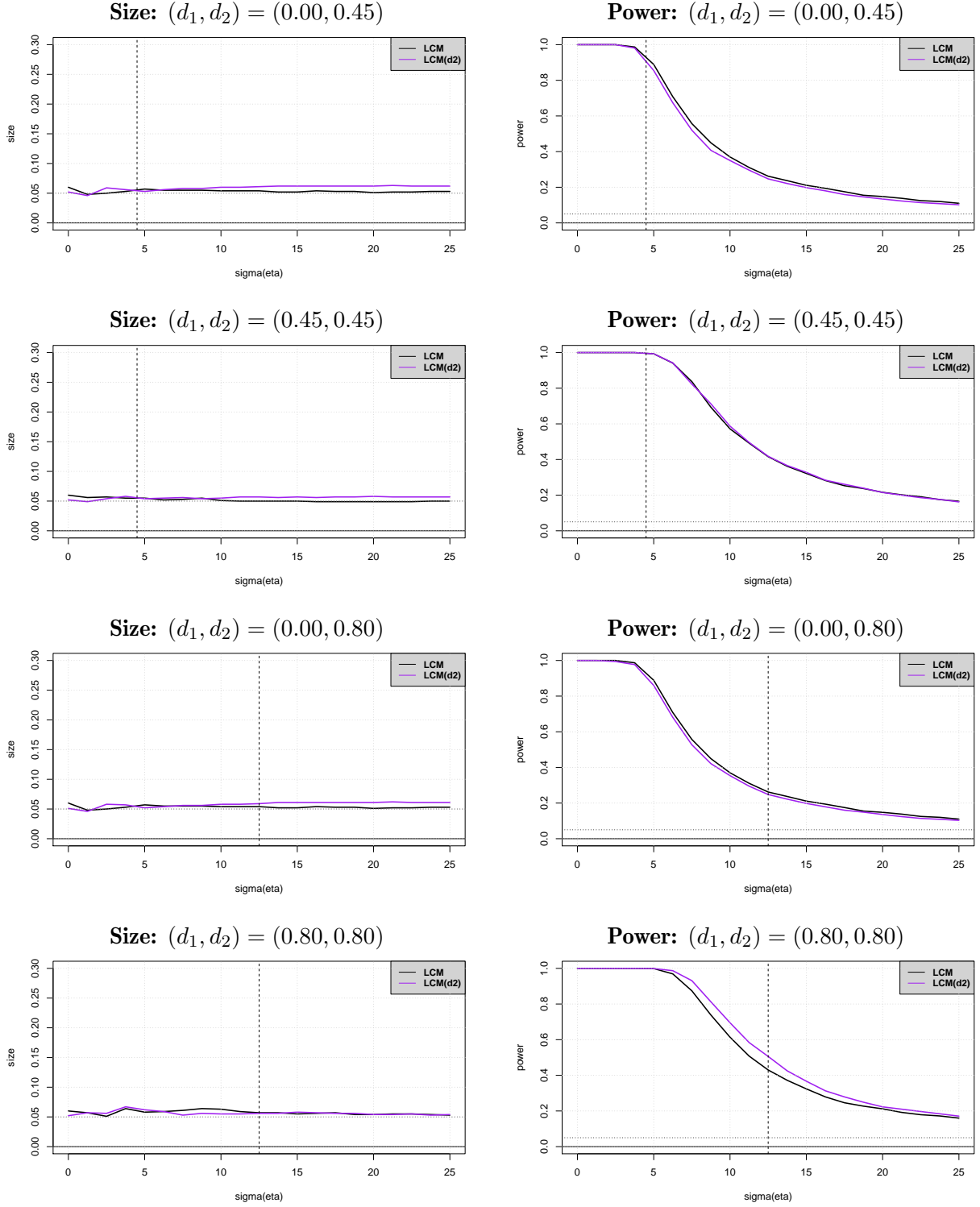


Figure 5: Size and power properties: Impact of LCMB. The left panels provide the size of LCM using either the LCMB approach with $\hat{\gamma}_x = \hat{d}_2$ to estimate d_1 or the standard implementation in Andersen & Varneskov (2021a). The right panels provide corresponding power results. See Sections 4-5 for implementation details. Size is for the parameterization $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$, and power is for $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$. Finally, we consider a sample size $n = 650$, a 5% nominal test size and use 1000 replications.

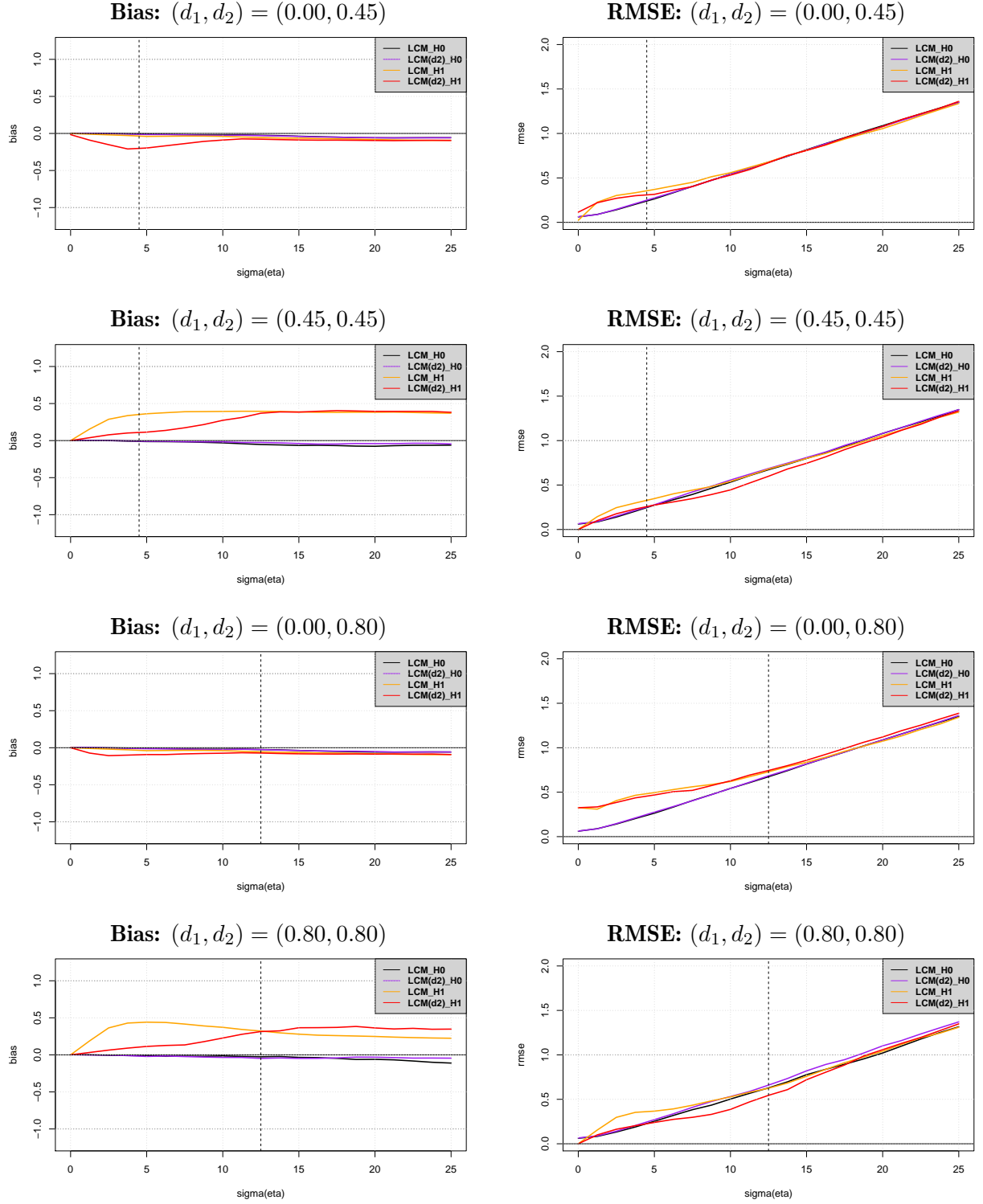


Figure 6: Bias and RMSE properties: Impact of LCMB. The left panels provide the bias of LCM using either the LCMB approach with $\hat{\gamma}_x = \hat{d}_2$ to estimate d_1 or the standard implementation in Andersen & Varneskov (2021a). The right panels provide corresponding RMSE results. See Sections 4-5 for implementation details. Moreover, the subscript with \mathcal{H}_0 indicates $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$, and \mathcal{H}_1 similarly denotes $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$. Finally, we consider a sample size $n = 650$ and use 1000 replications.

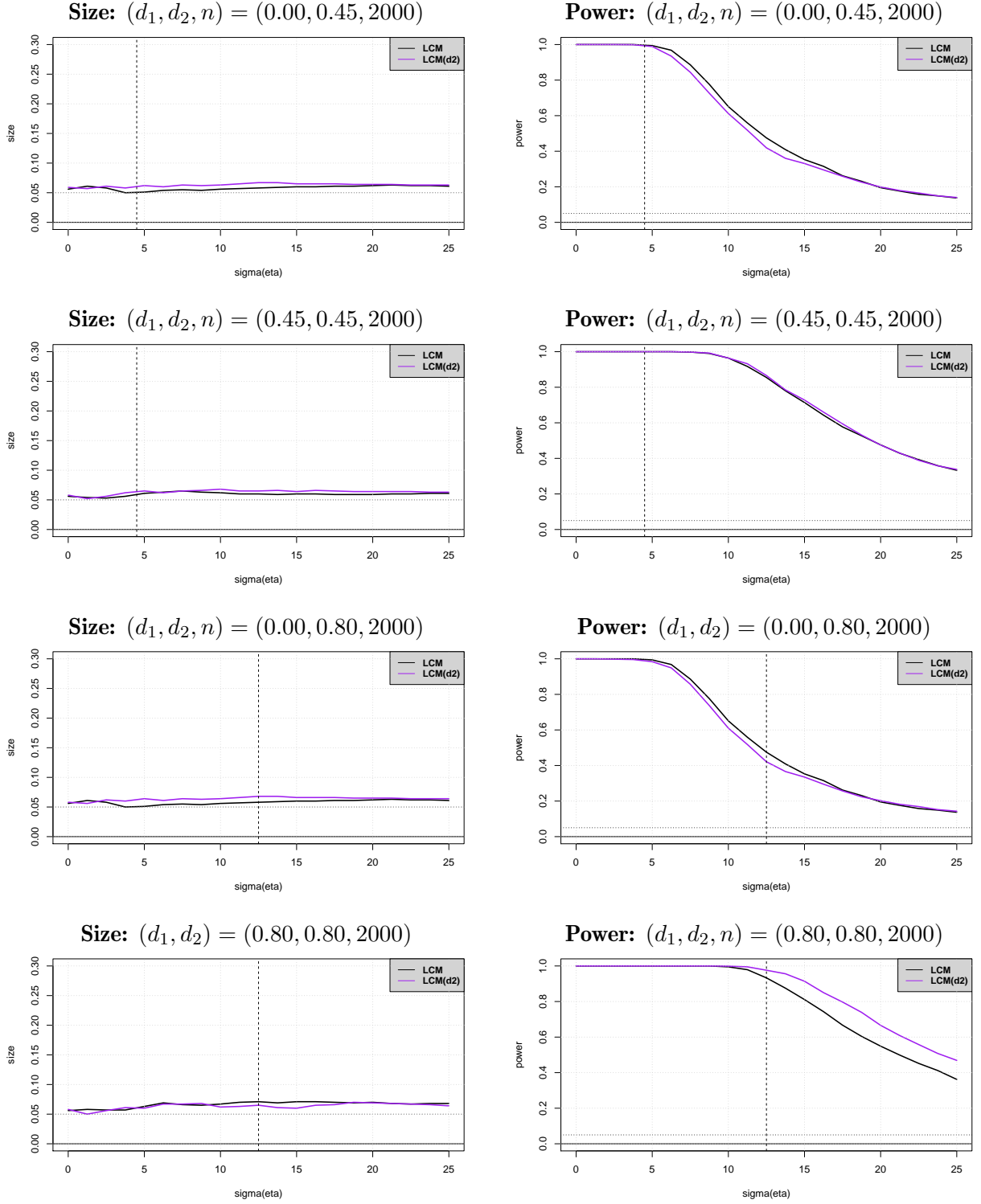


Figure 7: Size and power properties: Impact of LCMB and sample size. The left panels provide the size of LCM using either the LCMB approach with $\hat{\gamma}_x = \hat{d}_2$ to estimate d_1 or the standard implementation in Andersen & Varneskov (2021a). The right panels provide corresponding power results. See Sections 4-5 for implementation details. Size is for the parameterization $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$, and power is for $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$. Finally, we consider a sample size $n = 2000$, a 5% nominal test size and use 1000 replications.

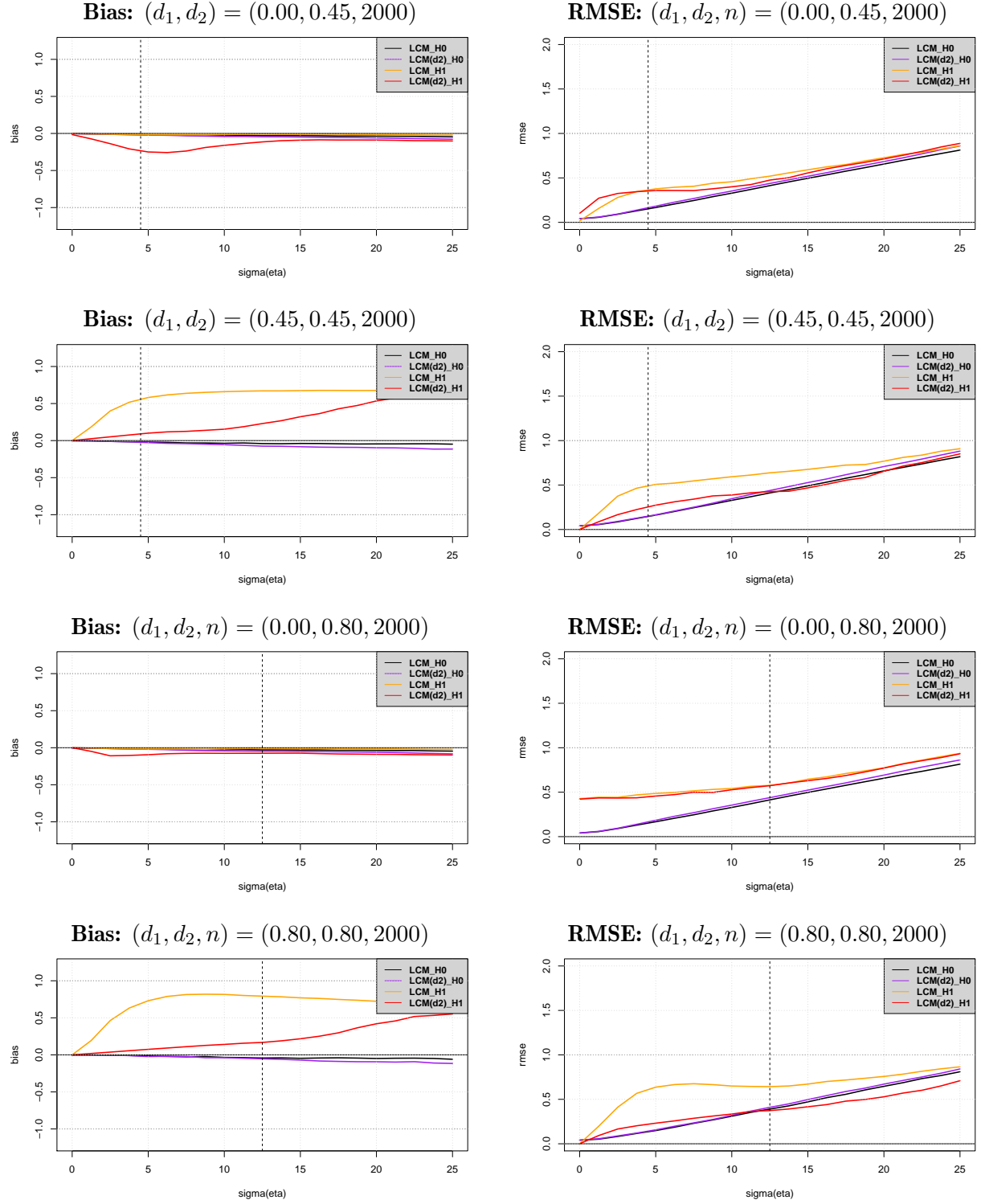


Figure 8: Bias and RMSE properties: Impact of LCMB and sample size. The left panels provide the bias of LCM using either the LCMB approach with $\hat{\gamma}_x = \hat{d}_2$ to estimate d_1 or the standard implementation in Andersen & Varneskov (2021a). The right panels provide corresponding RMSE results. See Sections 4-5 for implementation details. Moreover, the subscript with \mathcal{H}_0 indicates $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$, and \mathcal{H}_1 similarly denotes $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$. Finally, we consider a sample size $n = 2000$ and use 1000 replications.

Temporal Dependence and Rank					
	Returns _{<i>t</i>}	RV _{<i>t</i>-1}	DS _{<i>t</i>-1}	PE _{<i>t</i>-1}	TB _{<i>t</i>-1}
Mean	0.0055	0.0396	0.1346	0.0290	0.0468
Std. Dev.	0.0429	0.0239	0.0567	0.0042	0.0294
Skewness	-0.6833	3.9095	2.4834	-0.3293	0.5800
Kurtosis	5.5338	30.4239	12.5860	2.5694	3.7378
ACF(1)	0.0527	0.6468	0.9663	0.9955	0.9887
TELW	0.0789 (0.0500)	0.4227 (0.0500)	0.9982 (0.0500)	1.0356 (0.0500)	0.9388 (0.0500)
ELWM	0.0804 (0.0500)	0.4306 (0.0500)	0.9230 (0.0500)	1.1372 (0.0500)	0.9273 (0.0500)

Table 1: 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 2021*a*) 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 $\lfloor n^{0.71} \rfloor$ and, for TELW, trimming $\lfloor n^{0.1} \rfloor$ to reduce sensitivity to the mean. Finally, the sample of monthly observations spans March 1960 through March 2015 ($n = 661$).

Multivariate Return Predictions						
	RV _{<i>t</i>-1}	DS _{<i>t</i>-1}	TB _{<i>t</i>-1}	PE _{<i>t</i>-1}	Wald	\mathbb{P} -Wald
LCM(γ_x)	0.1003 (0.0980)	0.2144 (0.1016)	-0.3679 (0.3021)	-2.8461 (1.5459)	17.0281	0.0019
LCM(d_ψ)	-0.1638 (0.1167)	0.3065 (0.1158)	-0.6771 (0.3352)	-1.2283 (1.7696)	14.0414	0.0072
OLS	-0.1684 (0.1103)	0.0276 (0.0512)	-0.0885 (0.0969)	-0.8801 (0.6343)	6.4354	0.1689
IVX	-0.1646 (0.1178)	0.0268 (0.0509)	-0.0817 (0.0983)	-0.7462 (0.6493)	4.5776	0.3334

Table 2: Multivariate regressions. This table provides coefficient estimates, standard errors in parentheses and joint significance tests based on Wald statistics (and associated \mathbb{P} -values) for three different methods; OLS, IVX and LCM. Specifically, LCM is either implemented using $\hat{\gamma}_x = \max_{i=2,\dots,5} d_i = 1.0356$ from the TELW estimates in Table 1 for the return persistence, corresponding to the LCMB approach, or $\hat{d}_1(\psi) = 0.7676$. This is indicated by γ_x and d_ψ parentheses. The LCMB approach is implemented with $\nu = \nu_G = 0.2$ and $\kappa_p = 0.70$, and the subsequent LCM procedure with $\kappa = 0.69$. In both cases, κ_G is equal to the requisite bandwidth rate. Inference for OLS and IVX employs Ecker-White standard errors. IVX is implemented with two instruments as in Section 5. Finally, the sample of monthly observations spans March 1960 through March 2015 ($n = 661$).

Univariate Return Predictions						
	RV _{t-1}			DS _{t-1}		
	OLS	IVX	LCM(d_ψ)	OLS	IVX	LCM(d_ψ)
$\hat{\mathcal{B}}_c$	-0.1360	-0.1465	0.0163	0.0236	0.0200	0.2406
Wald	1.6690	1.7157	0.0241	0.3203	0.2326	5.6043
\mathbb{P} -Wald	0.1964	0.1902	0.8766	0.5714	0.6296	0.0179
	PE _{t-1}			TB _{t-1}		
	OLS	IVX	LCM(d_ψ)	OLS	IVX	LCM(d_ψ)
$\hat{\mathcal{B}}_c$	-0.4956	-0.3987	-0.7156	-0.0280	-0.0277	-0.8328
Wald	1.3162	0.8460	0.1731	0.2218	0.2098	6.3625
\mathbb{P} -Wald	0.2513	0.3577	0.6774	0.6377	0.6469	0.0117

Table 3: Univariate regressions. This table provides coefficient estimates and significance tests based on Wald statistics (and associated \mathbb{P} -values) for three different methods; OLS, IVX and LCM. Specifically, LCM is implemented using LCMB persistence to obtain $\hat{d}_1(\psi) = 0.7676$, which is indicated by d_ψ in the parenthesis. The LCMB approach is implemented with $\nu = \nu_G = 0.2$ and $\kappa_p = 0.70$, and the subsequent LCM procedure with $\kappa = 0.69$. In both cases, κ_G is equal to the requisite bandwidth rate. Inference for OLS and IVX employs Ecker-White standard errors. IVX is implemented with two instruments as in Section 5. 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. (2021a), ‘Consistent inference for predictive regressions in persistent economic systems’, *Journal of Econometrics* **224**, 215–244.
- Andersen, T. G. & Varneskov, R. T. (2021b), ‘Testing for parameter instability and structural change in persistent predictive regressions’, *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.
- Brockwell, P. J. & Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd ed. Springer Verlag, New York.
- 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.
- Hannan, E. (1970), *Multiple Time Series*, New York: Wiley.
- 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.
- Leeb, H. & Pötscher, B. (2005), ‘Model selection and inference: Fact and fiction’, *Econometric Theory* **21**, 21–59.
- 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. (1997), ‘Consistency of averaged cross-periodogram in long memory series’, *Journal of Time Series Analysis* **18**, 137–155.
- Lobato, I. (1999), ‘A semiparametric two-step estimator in a multivariate long memory model’, *Journal of Econometrics* **90**, 129–155.
- Magdalinos, T. & Phillips, P. C. B. (2009), Econometric inference in the vicinity of unity. CoFie Working Paper (7), Singapore Management University.
- 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. (2021), ‘Frequency dependent risk’, *Journal of Financial Economics* **140**, 644–675.
- 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 multivariate 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.
- Parzen, E. (1957), ‘On consistent estimates of the spectrum of a stationary time series’, *Annals of Mathematical Statistics* **28**, 329–348.
- 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. (2001), ‘Narrow-band analysis of nonstationary processes’, *The Annals of Statistics* **29**, 947–986.
- Robinson, P. M. & Marinucci, D. (2003), ‘Semiparametric frequency domain analysis of fractional cointegration’. In: Robinson, P.M. (Ed.), *Time Series with Long Memory*. Oxford University Press, Oxford, pp. 334–373.
- 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.
- Shiller, R. (2000), *Irrational Exuberance*, Princeton University Press, United States.
- Shimotsu, K. (2010), ‘Exact local whittle estimation of fractional integration with unknown 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 Assumptions and Theory

This section includes additional assumptions and provides a fractional cointegration rank testing procedure that can be used to discriminate between the model hypotheses (i)-(iv).

A.1 Additional Assumptions

Assumption D1- ζ . *The vector process ζ_{t-1} , $t = 1, \dots$, is covariance stationary with spectral density matrix satisfying $\mathbf{f}_{\zeta\zeta}(\lambda) \sim \mathbf{G}_{\zeta\zeta}$ as $\lambda \rightarrow 0^+$, where $\mathbf{G}_{\zeta\zeta}$ is finite with non-random elements. Moreover, there exists a $\varpi \in (0, 2]$ such that $|\mathbf{f}_{\zeta\zeta}(\lambda) - \mathbf{G}_{\zeta\zeta}| = O(\lambda^\varpi)$ as $\lambda \rightarrow 0^+$.*

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

Assumption D3- ζ . *For $\mathbf{A}_{\zeta}(\lambda, i)$, the i -th row of $\mathbf{A}_{\zeta}(\lambda) = \sum_{j=0}^{\infty} \mathbf{A}_{\zeta,j} e^{ij\lambda}$, its partial derivative satisfies $\|\partial \mathbf{A}_{\zeta}(\lambda, i) / \partial \lambda\| = O(\lambda^{-1} \|\mathbf{A}_{\zeta}(\lambda, i)\|)$ as $\lambda \rightarrow 0^+$, for $i = 1, \dots, k+1$.*

A.2 LCM Rank Testing and Imperfect Regressors

This section provides a new rank selection procedure under fractional cointegration, which facilitates discriminating between the model hypotheses (i)-(iv). Moreover, we establish the properties of the approach requiring that Assumption F holds, similarly to the results in Section 3. Hence, we can choose among several (semi-)parametric estimators of the integration orders for the regressors and the LCMB persistence estimator for returns, as discussed extensively in Sections 3-5.

A.2.1 LCM Covariance Estimation: Central Limit Theory

Initially, to obtain intuition, 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,

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

with, again, $\xi_{t-1} \perp\!\!\!\perp \mathbf{u}_{t-1}$ by Assumption C. Hence, we can apply this decomposition to test for the presence of ξ_{t-1} . The interpretation of our procedure, 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 3, 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 selection procedure for the presence of ξ_{t-1} using $\widehat{\mathbf{v}}_t^c = (\widehat{e}_t, (\widehat{\mathbf{u}}_{t-1}^c)')'$. This design works, since $\eta_t^{(d_1)}$ is a lower-order residual and has a degenerate spectral density for $\lambda \rightarrow 0^+$, as discussed below equations (11) and (12).

Formally, to design the model selection procedure, we leverage insights from equation (A.1) and use the fractionally filtered series, $\widehat{\mathbf{v}}_t^c$. 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 same trimmed long-run covariance estimator used to design the asymptotic variance estimator in Section 3.2,

$$\widehat{\mathbf{G}}_{\widehat{\mathbf{v}}}^c(\ell_G, m_G) = \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \Re(\mathbf{I}_{\widehat{\mathbf{v}}}^c(\lambda_j)), \quad (\text{A.2})$$

where, again, we use separate bandwidth and trimming functions, $m_G = m_G(n)$ and $\ell_G = \ell_G(n)$. Moreover, while Lemma B.1 in Appendix B and Theorem 2 establish consistency of the trimmed estimator (20) for the covariance matrix $\mathbf{G}_{\psi\psi}$ with $\psi_{t-1} \equiv (\varphi_{t-1}, \mathbf{u}_{t-1}')'$, we now require an associated central limit theory, covering models (ii)-(iv), to design a suitable rank test for (iv).²³

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 ,*

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

The first condition in Assumption T-G* is familiar from semiparametric frequency domain estima-

²³ Again, we use the definition $\psi_{t-1} \equiv (\varphi_{t-1}, \mathbf{u}_{t-1}')'$ to indicate that weakly dependent return innovations have no asymptotic impact on the limit theory, when $d_1 > 0$ in scenarios (ii)-(iv).

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

$$\frac{1 - \kappa_G}{2} \vee \frac{(1 - \kappa_G)(1 - \underline{d}) + \kappa_G/2}{2} \vee \frac{\kappa_G}{2} - (1 - \kappa_G)\underline{d}_x < \nu < \frac{\kappa_G}{2}, \quad (\text{A.3})$$

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 $\mathbf{c}_{t-1} = \mathbf{0}$, $\forall t = 1, \dots, n$. 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 6. *Suppose Assumptions D1-D3, C, M, F and T-G^{*} hold, $0 < d_1 \leq 1$, and $n^{1/2}/m_G \rightarrow 0$. Then, by letting $\mathbf{G}_{\psi\psi}^{(i)}$ be the $i = 1, \dots, k+1$ column of $\mathbf{G}_{\psi\psi}$, it follows,*

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

Theorem 6 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 $\mathbf{G}_{\psi\psi}$ 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 assess the null hypothesis $\widetilde{\mathcal{H}}_0 : \text{rank}(\mathbf{G}_{\psi\psi}) = k+1$ against the specific alternative $\widetilde{\mathcal{H}}_A : \text{rank}(\mathbf{G}_{\psi\psi}) = k$, because the spectral density matrix for the (latent) regressors \mathbf{u}_{t-1} is locally of full rank by Assumption D1 for frequencies $\lambda \rightarrow 0^+$.

Remark 9. *A result analogous to Theorem 6 still holds, if $b = d_1 = 0$, as for scenario (i). Specifically, we need to write $\psi_t \equiv \mathbf{v}_t$ and $\mathbf{G}_{\psi\psi} \equiv \mathbf{G}_{vv}$, 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), $\mathbf{G}_{\psi\psi}$ will be of full rank, and this will be indicated by our subsequent rank selection 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).*

Remark 10. *The condition $\sqrt{n}/m \rightarrow 0$ and $\sqrt{n}/m_G \rightarrow 0$, as $n \rightarrow \infty$, in Theorems 1 and 6, 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 $\sqrt{n}/(m^{1-\epsilon}\sqrt{m_d}) + \sqrt{n}/(m_G^{1-\epsilon}\sqrt{m_d}) \rightarrow 0$, 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.

A.2.2 LCM Rank Testing for Cointegration

Next, to estimate the rank, we let δ_i and $\hat{\delta}_i$, $i = 1, \dots, k+1$ denote the eigenvalues of the covariance matrices $\mathbf{G}_{\psi\psi}$ and $\hat{\mathbf{G}}_{\hat{v}\hat{v}}^c(\ell_G, m_G)$, listed in descending order, $0 < \delta_k < \dots < \delta_1$, with $0 < \delta_{k+1} < \delta_k$ and $\delta_{k+1} = 0$ under $\tilde{\mathcal{H}}_0$ and $\tilde{\mathcal{H}}_A$, 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,

$$\hat{r} = \underset{\varrho \in \{0,1\}}{\operatorname{argmin}} \mathcal{L}(\varrho), \quad \mathcal{L}(\varrho) = \vartheta(n)(k+1-\varrho) - \sum_{i=1}^{k+1-\varrho} \hat{\delta}_i, \quad (\text{A.4})$$

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

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

Theorem 7. Suppose the conditions of Theorem 6 and Assumption V hold, then,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\hat{r} = r) = 1.$$

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,

$$\hat{\mathbf{P}}_{\hat{v}\hat{v}}^c(\ell_G, m_G) \equiv \operatorname{diag}(\hat{\mathbf{G}}_{\hat{v}\hat{v}}^c(\ell_G, m_G))^{-1/2} \hat{\mathbf{G}}_{\hat{v}\hat{v}}^c(\ell_G, m_G) \operatorname{diag}(\hat{\mathbf{G}}_{\hat{v}\hat{v}}^c(\ell_G, m_G))^{-1/2}. \quad (\text{A.5})$$

In unreported simulations, using the numerical framework from Section 5, we find correlation-based procedures to perform substantially better than ones based on covariance estimates. Moreover, we have assessed alternative procedures based on feasible inference for the eigenvalues, inspired by Phillips & Ouliaris (1988), Robinson & Yajima (2002) and Nielsen & Shimotsu (2007). Consistent with the latter, we found these to perform strictly worse than model selection procedures in finite samples.

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 \mathbf{V} , let $\mathbf{V}(i)$ index the i th element, and, similarly, for

a matrix \mathbf{M} , let $\mathbf{M}(i, q)$ 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. The following two subsections provide new technical lemmas, which may be of independent interest. These are subsequently used to establish the main theoretical results.

B.1 Auxiliary Lemmas for Section 3 and Appendix A.2

This section provides two auxiliary lemmas that expand on Theorem 4 and Lemma A.4 in Andersen & Varneskov (2021a), studying endogeneity-induced errors and asymptotic variance estimation, respectively. We will henceforth refer to the latter as AV (2021) and, similarly, to their Online Appendix as AVOA (2020). Specifically, the first auxiliary lemma provides bounds for the differences,

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

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

where, as described in Section 3.2, $\widehat{\eta}_t^{(b,c)}$ constitutes an estimate of the regression residuals when implementing feasible inference. Hence, we provide asymptotic bounds to describe the errors arising when using the fractionally filtered observations $\widehat{\mathbf{v}}_t^c$ rather than the unobservable $\widehat{\mathbf{v}}_t$ when calculating key measures and statistics, thus quantifying the impact of regressor endogeneity. The auxiliary lemma differs from AV (2021, Theorem 4) by allowing for cointegration, $b > 0$, and imperfect regressors.

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

- (a) $\lambda_m^{-1}(\widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}^c(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m)) = O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon}),$
- (b) $\lambda_m^{-1}(\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^c(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}(\ell, m)) = O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon}),$
- (c) $\widehat{\mathbf{G}}_{\widehat{u}\widehat{u}}^c(\ell_G, m_G) - \widehat{\mathbf{G}}_{\widehat{u}\widehat{u}}(\ell_G, m_G) \leq O_p((m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon}),$
- (d) $\widehat{\mathbf{G}}_{\widehat{u}\widehat{e}}^c(\ell_G, m_G) - \widehat{\mathbf{G}}_{\widehat{u}\widehat{e}}(\ell_G, m_G) \leq O_p((m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon}),$
- (e) $\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b,c)}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b)}(\ell_G, m_G) \leq O_p((m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon}) + O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon}).$

Proof. First, (a) and (c) follows directly from AV (2021, Theorems 4(a) and 4(c)), since the assumptions and specification of the regressors in this paper readily follows their framework.²⁴

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

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

²⁴While AV (2021) state their results for \underline{d} rather than \underline{d}_x to maintain notational simplicity in their framework, it is clear that their results apply to \underline{d}_x as the parameter appears when applying the differencing operator to \mathbf{u}_{t-1} and \mathbf{c}_{t-1} .

for which the component $\hat{e}_t^{(1)}$ is equivalent to the case without cointegration considered by AV (2021, Theorem 4(b)) due to Assumptions D1-D3 and C. By applying the decomposition (B.3), we have

$$\hat{\mathbf{F}}_{\widehat{ue}}^c(\ell, m) - \hat{\mathbf{F}}_{\widehat{ue}}(\ell, m) = \hat{\mathbf{F}}_{\widehat{ce}}^c(\ell, m) = \hat{\mathbf{F}}_{\widehat{ce}}^{(c,1)}(\ell, m) + \hat{\mathbf{F}}_{\widehat{ce}}^{(c,2)}(\ell, m), \quad (\text{B.4})$$

where $\hat{\mathbf{F}}_{\widehat{ce}}^{(c,1)}(\ell, m)$ and $\hat{\mathbf{F}}_{\widehat{ce}}^{(c,2)}(\ell, m)$ are the TDACs between \widehat{c}_{t-1} and $\widehat{e}_t^{(1)}$, respectively, $\widehat{e}_t^{(2)}$. Now, by applying AV (2021, Theorem 4(b)) and AVOA (2020, Lemma A.12(b)), we have

$$\lambda_m^{-1} \hat{\mathbf{F}}_{\widehat{ce}}^{(c,1)}(\ell, m) \leq O_p((m/n)^{d_x}/\ell^{1+\epsilon}) \quad \text{and} \quad \mathbf{w}_{\widehat{c}}(\lambda_j, i) = O_p(\lambda_j^{d_i}), \quad (\text{B.5})$$

for $i = 2, \dots, k+1$. Moreover, we can write $\widehat{\theta}_1 = \widehat{d}_1 - d_1 = O_p(1/\sqrt{m_d})$ and

$$\widehat{e}_t^{(2)} = (1 - L)^{\widehat{d}_1 - d_1} (1 - L)^{d_1} \eta_t \equiv (1 - L)^{\widehat{\theta}_1} \widetilde{e}_t^{(2)}, \quad (\text{B.6})$$

using Assumption F, such that by AVOA (2020, Lemmas A.8 and A.9(a)), it follows that,

$$w_{\widehat{e}}^{(2)}(\lambda_j) = w_{\widehat{e}}^{(2)}(\lambda_j) \left(1 + O_p(\ln(n)/\sqrt{m_d}) + O_p(\ln(n)^2/m_d) \right), \quad \text{with} \quad (\text{B.7})$$

$$w_{\widehat{e}}^{(2)}(\lambda_j) = \lambda_j^{d_1} e^{-(\pi/2)id_1} w_{\eta}(\lambda_j) + O_p(\lambda_j^{d_1} \ln(n)/j^{-1/2}) + O_p(n^{-d_1-1}), \quad (\text{B.8})$$

and, furthermore, by AVOA (2020, Lemma A.12(b)) that $w_{\widehat{e}}^{(2)}(\lambda_j) = O_p(\lambda_j^{d_1})$ when $\ln(n)/j^{1/2} \rightarrow 0$. Hence, since $0 < d_i \leq d_1 + d_i$, $i = 2, \dots, k+1$, we may further write

$$\hat{\mathbf{F}}_{\widehat{ce}}^{(c,2)}(\ell, m) \leq \frac{2\pi}{n} \sum_{j=\ell}^m O_p(\lambda_j^{d_x}) \leq \frac{2\pi m^{1+d_x}}{n^{1+d_x}} \sum_{j=\ell}^m O_p \left(\left(\frac{j}{m} \right)^{d_x} \frac{1}{j^{1+\epsilon}} \right) \leq O_p \left(\left(\frac{m}{n} \right)^{1+d_x} \frac{1}{\ell^{1+\epsilon}} \right), \quad (\text{B.9})$$

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

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

$$\widehat{\mathbf{G}}_{\widehat{ue}}^c(\ell_G, m_G) - \widehat{\mathbf{G}}_{\widehat{ue}}(\ell_G, m_G) = \widehat{\mathbf{G}}_{\widehat{ce}}^c(\ell_G, m_G) = \widehat{\mathbf{G}}_{\widehat{ce}}^{(c,1)}(\ell_G, m_G) + \widehat{\mathbf{G}}_{\widehat{ce}}^{(c,2)}(\ell_G, m_G), \quad (\text{B.10})$$

where, again, the DFT bounds in (B.5) apply to $w_{\widehat{e}}^{(2)}(\lambda_j)$ and $\mathbf{w}_{\widehat{c}}(\lambda_j, i)$. Moreover, by AV (2021, Theorem 4(c)), we have

$$\widehat{\mathbf{G}}_{\widehat{ce}}^{(c,1)}(\ell_G, m_G) \leq O_p((m_G/n)^{d_x}/\ell_G^{1+\epsilon}). \quad (\text{B.11})$$

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

$$\widehat{\mathbf{G}}_{\widehat{ce}}^{(c,2)}(\ell_G, m_G) \leq \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{d_x})$$

$$\leq \frac{Km_G^{d_x}}{n^{d_x}} \sum_{j=\ell_G}^{m_G} O_p \left(\left(\frac{j}{m_G} \right)^{d_x} \frac{1}{j^{1+\epsilon}} \right) \leq O_p \left(\left(\frac{m_G}{n} \right)^{d_x} \frac{1}{\ell_G^{1+\epsilon}} \right), \quad (\text{B.12})$$

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

For **(e)**, recall $\hat{\eta}_t^{(b)} = \hat{e}_t - \hat{\mathbf{B}}(\ell, m)' \hat{\mathbf{u}}_{t-1}$ and let us define

$$\hat{\eta}_t^{(b,1)} = \hat{e}_t^{(1)} - \hat{\mathbf{B}}(\ell, m)' \hat{\mathbf{u}}_{t-1}, \quad \hat{\tau}_{t-1}^{(1)} = (\hat{\mathbf{B}}_c(\ell, m) - \hat{\mathbf{B}}(\ell, m))' \hat{\mathbf{u}}_{t-1}^c, \quad \hat{\tau}_{t-1}^{(2)} = \hat{\mathbf{B}}(\ell, m) \hat{e}_{t-1}, \quad (\text{B.13})$$

such that we can use $\hat{e}_t = \hat{e}_t^{(1)} + \hat{e}_t^{(2)}$ to decompose $\hat{\eta}_t^{(b)} = \hat{\eta}_t^{(b,1)} + \hat{e}_t^{(2)}$ and write

$$\hat{\eta}_t^{(b,c)} = \hat{e}_t - \hat{\mathbf{B}}_c(\ell, m)' \hat{\mathbf{u}}_{t-1}^c = \hat{\eta}_t^{(b,1)} + \hat{e}_t^{(2)} - \hat{\tau}_{t-1}^{(1)} - \hat{\tau}_{t-1}^{(2)} = \hat{\eta}_t^{(b)} - \hat{\tau}_{t-1}^{(1)} - \hat{\tau}_{t-1}^{(2)}. \quad (\text{B.14})$$

The main difference between this decomposition and the corresponding in AV (2021, Theorem 4) is the presence of $\hat{e}_t^{(2)}$ and the fact that we have $b \neq 0$ for $\eta_t^{(b)}$. Hence, we need to distinguish between cases without cointegration in scenarios (ii) and (iii), where ξ_{t-1} is driving the limit, and scenario (iv), where $\xi_{t-1} = 0, \forall t$. In both inference regimes, we have $d_1 = b > 0$. In the first case, we have fractionally differenced $\xi_{t-1}^{(-d_1)}$ such that the resulting error process is ξ_{t-1} , asymptotically.

Before treating the subtleties of the two inference regimes, we provide bounds on the error terms that are common to both. To this end, we use (B.14) and make the decomposition,

$$\begin{aligned} \hat{\mathbf{G}}_{\hat{\eta}\hat{\eta}}^{(b,c)}(\ell_G, m_G) - \hat{\mathbf{G}}_{\hat{\eta}\hat{\eta}}^{(b)}(\ell_G, m_G) &= \hat{\mathbf{G}}_{\hat{\tau}\hat{\tau}}^{(1,1)}(\ell_G, m_G) + \hat{\mathbf{G}}_{\hat{\tau}\hat{\tau}}^{(2,2)}(\ell_G, m_G) + 2\hat{\mathbf{G}}_{\hat{\tau}\hat{\tau}}^{(1,2)}(\ell_G, m_G) \\ &\quad - 2\hat{\mathbf{G}}_{\hat{\eta}\hat{\tau}}^{(b,1)}(\ell_G, m_G) - 2\hat{\mathbf{G}}_{\hat{\eta}\hat{\tau}}^{(b,2)}(\ell_G, m_G), \end{aligned} \quad (\text{B.15})$$

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

$$\hat{\mathbf{G}}_{\hat{\eta}\hat{\tau}}^{(b,i)}(\ell_G, m_G) = \hat{\mathbf{G}}_{\hat{\eta}\hat{\tau}}^{(b,i,1)}(\ell_G, m_G) + \hat{\mathbf{G}}_{\hat{\eta}\hat{\tau}}^{(b,i,2)}(\ell_G, m_G), \quad i = 1, 2, \quad (\text{B.16})$$

to indicate the decomposition of $\hat{\eta}_t^{(b)}$ into $\hat{\eta}_t^{(b,1)}$ and $\hat{e}_t^{(2)}$. Now, since **(a)** and **(b)** yield,

$$\hat{\mathbf{B}}_c(\ell, m) - \hat{\mathbf{B}}(\ell, m) \leq O_p((m/n)^{d_x}/\ell^{1+\epsilon}), \quad (\text{B.17})$$

we may use equations (A.23), (A.26) and (A.31) in AVOA (2020) to show

$$\begin{aligned} \hat{\mathbf{G}}_{\hat{\tau}\hat{\tau}}^{(1,1)}(\ell_G, m_G) &\leq O_p \left(\left(\frac{m}{n} \right)^{2d_x} \frac{1}{\ell^{2(1+\epsilon)}} \right) \times \left(1 + \left(\frac{m_G}{n} \right)^{d_x} \frac{1}{\ell_G^{1+\epsilon}} \right), \\ \hat{\mathbf{G}}_{\hat{\tau}\hat{\tau}}^{(2,2)}(\ell_G, m_G) &\leq O_p \left(\left(\frac{m}{n} \right)^{2d_x} \frac{1}{\ell^{1+\epsilon}} \right), \end{aligned}$$

$$\widehat{\mathbf{G}}_{\widehat{\tau}\widehat{\tau}}^{(1,2)}(\ell_G, m_G) \leq O_p \left(\left(\frac{m}{n} \right)^{\underline{d}_x} \frac{1}{\ell^{1+\epsilon}} \right) \times \left(\left(\frac{m_G}{n} \right)^{\underline{d}_x} \frac{1}{\ell_G^{1+\epsilon}} \right),$$

where, again, the bounds are restated with \underline{d}_x rather than \underline{d} as in AVOA (2020), since the parameter appears when applying the fractional differencing operator to \mathbf{u}_{t-1} and \mathbf{c}_{t-1} .

The case without cointegration. Here, $\widehat{e}_t^{(1)}$ will drive the asymptotic limit and $\widehat{e}_t^{(2)}$ will be a lower order error term. Moreover, the former corresponds to the case considered by AV (2021, Theorem 4), with ξ_{t-1} acting as the (regression) error process. Hence, by invoking equations (A.27) and (A.30) of AVOA (2020), it follows for the two long-run covariance terms involving $\widehat{\eta}_t^{(b,1)}$ that,

$$\begin{aligned} \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b,1,1)}(\ell_G, m_G) &\leq O_p \left(\left(\frac{m}{n} \right)^{\underline{d}_x} \frac{1}{\ell^{1+\epsilon}} \right), \\ \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,2,1)}(\ell_G, m_G) &\leq O_p \left(\left(\frac{m_G}{n} \right)^{\underline{d}_x} \frac{1}{\ell_G^{1+\epsilon}} \right) \times \left(1 + m^{-1/2} \right). \end{aligned}$$

This implies that in order to complete the proof, we need to establish corresponding asymptotic bounds for the remaining terms, $\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,1,2)}(\ell_G, m_G)$ and $\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,2,2)}(\ell_G, m_G)$, i.e., the long-run covariances involving $\widehat{e}_t^{(2)}$. To this end, let us use the discrete Fourier transform bounds in (B.5) and (B.7)-(B.8), $\widehat{\mathbf{B}}(\ell, m) = O_p(1)$, uniformly by AV (2021, Theorem 1), and $0 < d_i \leq d_1 + d_i$ to write,

$$\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,2,2)}(\ell_G, m_G) \leq \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{\underline{d}_x}) \leq O_p \left(\left(\frac{m_G}{n} \right)^{\underline{d}_x} \frac{1}{\ell_G^{1+\epsilon}} \right), \quad (\text{B.18})$$

similarly to (B.12). For the last term, make the decomposition,

$$\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,1,2)}(\ell_G, m_G) = (\widehat{\mathbf{B}}_c(\ell, m) - \widehat{\mathbf{B}}(\ell, m))' \left(\widehat{\mathbf{G}}_{\widehat{u}\widehat{e}}^{(2)}(\ell_G, m_G) + \widehat{\mathbf{G}}_{\widehat{c}\widehat{e}}^{(c,2)}(\ell_G, m_G) \right), \quad (\text{B.19})$$

where $\widehat{\mathbf{G}}_{\widehat{c}\widehat{e}}^{(c,2)}(\ell_G, m_G) \leq O_p((m_G/n)^{\underline{d}_x} 1/\ell_G^{1+\epsilon})$ by (B.12). Moreover, for $i = 2, \dots, k+1$, since

$$\mathbf{w}_{\widehat{u}}(\lambda_j, i) = \mathbf{w}_u(\lambda_j, i) + O_p \left(\frac{n^{1/2-d_i}}{j^{1-d_i}} \right) + O_p \left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j} \right), \quad \mathbf{w}_u(\lambda_j, i) = O_p(1), \quad (\text{B.20})$$

by equations (A.8), (A.60), (A.65) and Lemma A.6(a) in AVOA (2020), we may write

$$\begin{aligned} \widehat{\mathbf{G}}_{\widehat{u}\widehat{e}}^{(2)}(\ell_G, m_G) &\leq \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{d_1}) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p \left(\frac{\lambda_j^{d_1+\underline{d}_x} n^{1/2}}{j} \right) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p \left(\frac{\lambda_j^{d_1} \ln(n)n^{1/2}}{m_d^{1/2}j} \right) \\ &\leq O_p(1) + O_p \left(\left(\frac{m_G}{n} \right)^{\underline{d}_x} \frac{n^{1/2}}{m_G^{1-\epsilon} \ell_G^{1+\epsilon}} \right) + O_p \left(\frac{n^{1/2} \ln(n)}{m_G^{1-\epsilon} m_d^{1/2} \ell_G^{1+\epsilon}} \right), \end{aligned} \quad (\text{B.21})$$

for some arbitrarily small $\epsilon > 0$, using $d_1 \geq 0$ and Varneskov (2017, Lemma C.4). Hence, by combining

bounds, $n^{1/2}/m_G \rightarrow 0$, Lemmas B.1(a)-(b) in the absence of cointegration in conjunction with the continuous mapping theorem, we have $\widehat{\mathbf{G}}_{\widehat{u}\widehat{e}}^{(2)}(\ell_G, m_G) \leq O_p(1)$ and, thus,

$$\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,1,2)}(\ell_G, m_G) \leq O_p((m/n)^{d_x}/\ell^{1+\epsilon}). \quad (\text{B.22})$$

Consequently, by collecting bounds for all components in (B.16),

$$\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,1)}(\ell_G, m_G) + \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,2)}(\ell_G, m_G) \leq O_p((m_G/n)^{d_x}/\ell_G^{1+\epsilon}) + O_p((m/n)^{d_x}/\ell^{1+\epsilon}), \quad (\text{B.23})$$

which, together with bounds for the remaining terms in equation (B.15), provides the requisite result when cointegration is absent, that is, for the inference scenarios (ii)-(iii).

The case with cointegration. By **(a)** and **(b)**, whose rates are independent of cointegration, there are no difference between the treatment of the terms in (B.15) and (B.16) except for $\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,1,1)}(\ell_G, m_G)$ and $\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,2,1)}(\ell_G, m_G)$, that is, the covariance terms involving $\eta_t^{(b,1)}$. Hence, let us define

$$\widehat{e}_t^{(1)} = \widehat{e}_t^{(1)} - (1-L)^{\widehat{d}_1} \xi_{t-1}^{(-d_1)} \equiv \widehat{e}_t^{(1)} - \widehat{\xi}_{t-1}, \quad \widehat{\eta}_t^{(b,1)} = \widehat{e}_t^{(1)} - \widehat{\mathbf{B}}(\ell, m)' \widehat{\mathbf{u}}_{t-1}, \quad (\text{B.24})$$

such that the triangle inequality delivers:

$$|\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,1,1)}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,1,1)}(\ell_G, m_G)| \leq O_p\left(\widehat{\mathbf{G}}_{\widehat{\xi}\widehat{\tau}}^{(b,1,1)}(\ell_G, m_G)\right), \quad (\text{B.25})$$

$$|\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,2,1)}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\tau}}^{(b,2,1)}(\ell_G, m_G)| \leq O_p\left(\widehat{\mathbf{G}}_{\widehat{\xi}\widehat{\tau}}^{(b,2,1)}(\ell_G, m_G)\right). \quad (\text{B.26})$$

This implies that the result *with cointegration* follows by from the result *without cointegration* as well as establishing (and verifying) the bounds on the long-run covariance terms between $\widehat{\xi}_{t-1}$ and the errors $\widehat{\tau}_{t-1}^{(1)}$ and $\widehat{\tau}_{t-1}^{(2)}$. To this end, we may use AVOA (2020, Lemma A.12(b)) to write

$$w_{\widehat{\xi}}(\lambda_j) = O_p(1) + O_p\left(\frac{n^{1/2-d_1}}{j^{1-d_1}}\right) + O_p\left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right), \quad (\text{B.27})$$

$$w_{\widehat{\tau}}^{(1)}(\lambda_j) = \left(\widehat{\mathbf{B}}_c(\ell, m) - \widehat{\mathbf{B}}(\ell, m)\right)' (w_{\widehat{u}}(\lambda_j) + w_{\widehat{e}}(\lambda_j)), \quad w_{\widehat{\tau}}^{(2)}(\lambda_j) = \widehat{\mathbf{B}}(\ell, m)' w_{\widehat{e}}(\lambda_j), \quad (\text{B.28})$$

where the components in $w_{\widehat{\tau}}^{(1)}(\lambda_j)$ and $w_{\widehat{\tau}}^{(2)}(\lambda_j)$ are described by (B.5), (B.17) and (B.20). Hence, it suffices to study $\widehat{\mathbf{G}}_{\widehat{u}\widehat{\xi}}(\ell_G, m_G)$ and $\widehat{\mathbf{G}}_{\widehat{e}\widehat{\xi}}(\ell_G, m_G)$. First, for the latter,

$$\begin{aligned} \widehat{\mathbf{G}}_{\widehat{e}\widehat{\xi}}(\ell_G, m_G) &\leq \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{d_x}) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{d_1+d_x} n^{1/2}}{j}\right) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{d_x} \ln(n) n^{1/2}}{m_d^{1/2} j}\right) \\ &\leq O_p\left(\left(\frac{m_G}{n}\right)^{d_x} \frac{1}{\ell_G^{1+\epsilon}}\right) + O_p\left(\left(\frac{m_G}{n}\right)^{d_x} \frac{n^{1/2}}{m_G^{1-\epsilon} \ell_G^{1+\epsilon}}\right) + O_p\left(\frac{n^{1/2} \ln(n)}{m_G^{1-\epsilon} m_d^{1/2} \ell_G^{1+\epsilon}}\right), \end{aligned} \quad (\text{B.29})$$

by the same arguments used for $\widehat{\mathbf{G}}_{\widehat{u}\widehat{e}}^{(2)}(\ell_G, m_G)$ in (B.21). Similarly, we have

$$\begin{aligned}\widehat{\mathbf{G}}_{\widehat{u}\widehat{\xi}}(\ell_G, m_G) &\leq \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p(1) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{\lambda_j^{\frac{d}{2}} n^{1/2}}{j}\right) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right) \\ &\quad + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{n}{j^2}\left(\lambda_j^{d_1+d_x} + \lambda_j^{\frac{d}{2}} \ln(n)/\sqrt{m_d} + \ln(n)^2/m_d\right)\right) \\ &\leq O_p(1) + O_p\left(\frac{n^{1/2}}{m_G^{1-\epsilon}\ell_G^{1+\epsilon}}\right) + O_p\left(\frac{n^{1/2}\ln(n)}{m_G^{1-\epsilon}m_d^{1/2}\ell_G^{1+\epsilon}}\right) \\ &\quad + O_p\left(\frac{n}{m_G\ell_G^2}\left(\left(\frac{m_G}{n}\right)^{\frac{d_x}{2}} + \frac{\ln(n)}{\sqrt{m_d}} + \frac{\ln(n)^2}{m_d}\right)\right),\end{aligned}\tag{B.30}$$

which is $O_p(1)$ by Assumption T-G. Hence, by combining results, we have

$$\widehat{\mathbf{G}}_{\widehat{\xi}\widehat{\tau}}^{(b,2,1)}(\ell_G, m_G) \leq O_p\left(\left(\frac{m_G}{n}\right)^{\frac{d_x}{2}} \frac{1}{\ell_G^{1+\epsilon}}\right), \quad \widehat{\mathbf{G}}_{\widehat{\xi}\widehat{\tau}}^{(b,1,1)}(\ell_G, m_G) \leq O_p\left(\left(\frac{m}{n}\right)^{\frac{d_x}{2}} \frac{1}{\ell^{1+\epsilon}}\right),\tag{B.31}$$

which, together with the remaining bounds for the case *without cointegration*, provides the requisite result for the case *with cointegration*, i.e., scenario (iv), thereby concluding the proof. \square

The second auxiliary lemma establishes bounds and convergence results for $\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b)}(\ell_G, m_G)$ in the cases with and without cointegration. To this end, recall from the proof of Lemma B.1 that we have the decomposition $\widehat{\eta}_t^{(b)} = \widehat{\eta}_t^{(b,1)} + \widehat{e}_t^{(2)}$ where $\widehat{\eta}_t^{(b,1)} = \widehat{e}_t^{(1)} - \widehat{\mathbf{B}}(\ell, m)' \widehat{\mathbf{u}}_{t-1}$ and the terms $(\widehat{e}_t^{(1)}, \widehat{e}_t^{(2)})$ are defined in equation (B.3). Hence, we can equivalently write,

$$\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b)}(\ell_G, m_G) = \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b,1)}(\ell_G, m_G) + \widehat{\mathbf{G}}_{\widehat{e}\widehat{e}}^{(2)}(\ell_G, m_G) + 2\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{e}}^{(b,1,2)}(\ell_G, m_G),\tag{B.32}$$

whose components will be analyzed separately in the following.

Lemma B.2. *Suppose the conditions of Lemma B.1 hold. Then,*

$$(a) \quad \lambda_{m_G}^{-2b} \left(\widehat{\mathbf{G}}_{\widehat{e}\widehat{e}}^{(2)}(\ell_G, m_G) - G_{\eta\eta}/(1+2b) \right) = o_p(1),$$

(b) *The following convergence results hold,*

$$\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b,1)}(\ell_G, m_G) = \begin{cases} G_{\xi\xi} + o_p(1), & \text{under models (ii) and (iii),} \\ o_p(\lambda_{m_G}^{2b}), & \text{under model (iv).} \end{cases}$$

(c) *The following convergence results hold,*

$$\begin{cases} \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b)}(\ell_G, m_G) \xrightarrow{\mathbb{P}} G_{\xi\xi}, & \text{under models (ii) and (iii),} \\ \lambda_{m_G}^{-2b} \widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b)}(\ell_G, m_G) \xrightarrow{\mathbb{P}} G_{\eta\eta}/(1+2b), & \text{under model (iv).} \end{cases}$$

Proof. First, for **(a)**, we may use $d_1 = b$, (B.7) and (B.8) to write

$$w_{\hat{e}}^{(2)}(\lambda_j) = \lambda_j^b e^{-(\pi/2)ib} w_\eta(\lambda_j) + O_p(\lambda_j^b \ln(n) m_d^{-1/2}) + O_p(\lambda_j^b \ln(n) j^{-1/2}) + O_p(n^{-b-1}). \quad (\text{B.33})$$

Moreover, let us define the related long-run covariance measure,

$$\tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(\ell_G, m_G) \equiv \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \lambda_j^{2b} \Re(w_\eta(\lambda_j) \bar{w}_\eta(\lambda_j)), \quad (\text{B.34})$$

and subsequently make the error decomposition,

$$\begin{aligned} \hat{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(\ell_G, m_G) - \tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(\ell_G, m_G) &\leq \frac{K \ln(n)}{m_G} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{2b} (j^{-1/2} + m_d^{-1/2})) + \frac{K}{m_G n^{b+1}} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^b) \\ &\leq \ln(n) \left(\frac{m_G}{n}\right)^{2b} \left(\frac{m_G^\epsilon}{m_G^{1/2}} + \frac{m_G^\epsilon}{m_d^{1/2}}\right) \sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^{2b} j^{-(1+\epsilon)}\right) \\ &\quad + \left(\frac{m_G}{n}\right)^b \frac{m_G^\epsilon}{n} \sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{j}{m_G}\right)^b j^{-(1+\epsilon)}\right) \\ &\leq O_p\left(\frac{\ln(n)}{\ell_G^{1+\epsilon}} \left(\frac{m_G}{n}\right)^{2b} \left(\frac{m_G^\epsilon}{m_G^{1/2}} + \frac{m_G^\epsilon}{m_d^{1/2}}\right)\right) + O_p\left(\left(\frac{m_G}{n}\right)^b \frac{m_G^\epsilon}{n \ell_G^{1+\epsilon}}\right), \end{aligned} \quad (\text{B.35})$$

for some arbitrarily small $\epsilon > 0$, using Varneskov (2017, Lemma C.4) and that the remaining cross-product error terms arising from the product of the decomposition in (B.33) are of strictly lower asymptotic order by the tuning parameters being $\ell_G \asymp n^{\nu_G}$, $m_G \asymp n^{\kappa_G}$ and $m_d \asymp n^\varrho$, with $0 < \nu_G < \kappa_G < \varrho \leq 1$ in Assumptions F and T-G. Together with $0 \leq b \leq 1$, this implies

$$\lambda_{m_G}^{-2b} \left(\hat{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(\ell_G, m_G) - \tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(\ell_G, m_G) \right) \leq o_p(1). \quad (\text{B.36})$$

Hence, we continue by examining $\tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(\ell_G, m_G)$. By definition, we have

$$\tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(\ell_G, m_G) - \tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(1, m_G) = \frac{-1}{m_G - \ell_G + 1} \sum_{j=1}^{\ell_G-1} \lambda_j^{2b} \Re(w_\eta(\lambda_j) \bar{w}_\eta(\lambda_j)) = \frac{-\ell_G \tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(1, \ell_G - 1)}{m_G - \ell_G + 1}. \quad (\text{B.37})$$

Let $l \in \{\ell_G - 1, m_G\}$ be either of the two generic sequences of integers, then we adopt exactly the same arguments used to establish Christensen & Varneskov (2017, Equation (B.7)) (see also Robinson & Marinucci (2003, p. 361) and the steps for Lobato (1997, Theorem 1)) to show

$$\lambda_l^{-2b} \left| \tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(1, l) - G_{\eta\eta}/(1 + 2b) \right| \leq o_p(1), \quad (\text{B.38})$$

noting that $\tilde{\mathbf{G}}_{\hat{e}\hat{e}}^{(2)}(1, l)$ corresponds to $\lambda_m^{-1} \mathbf{F}_{zz}^*(1, m)$ in their notation, and b to $-d_i$. Hence, by combining

equations (B.37) and (B.38), we can establish the following bound:

$$\lambda_{m_G}^{-2b} \left(\tilde{\mathbf{G}}_{\widehat{e\widehat{e}}}^{(2)}(\ell_G, m_G) - \tilde{\mathbf{G}}_{\widehat{e\widehat{e}}}^{(2)}(1, m_G) \right) \leq O_p \left(\left(\frac{\lambda_{\ell_G}}{\lambda_{m_G}} \right)^{2b} \frac{\ell_G}{m_G} \right) = o_p(1). \quad (\text{B.39})$$

The requisite result, thus, follows by combining results (B.37), (B.38) and (B.39).

Next, for **(b)**, we will study the properties of $\widehat{\mathbf{G}}_{\widehat{\eta\eta}}^{(b,1)}(\ell_G, m_G)$ under the scenarios (ii)-(iii) as well as the cointegration setting (iv). To this end, let us write

$$\widehat{\eta}_t^{(b,1)} = \widehat{\eta}_t^{(b,1,1)} + \widehat{\eta}_t^{(b,1,2)} + \widehat{\eta}_t^{(b,1,3)} + \widehat{\eta}_t^{(b,1,4)} \quad (\text{B.40})$$

where, by addition and subtraction, the components are defined as,

$$\begin{aligned} \widehat{\eta}_t^{(b,1,1)} &= (1-L)^{\widehat{d}_1} \xi_{t-1}^{(-d_1)}, \quad \widehat{\eta}_t^{(b,1,2)} = \widehat{\mathbf{B}}(\ell, m)'(\mathbf{u}_{t-1} - \widehat{\mathbf{u}}_{t-1}), \quad \widehat{\eta}_t^{(b,1,3)} = (\mathbf{B} - \widehat{\mathbf{B}}(\ell, m))' \mathbf{u}_{t-1}, \\ \widehat{\eta}_t^{(b,1,4)} &= \widehat{e}_t^{(1)} - \mathbf{B}' \mathbf{u}_{t-1} = (1-L)^{\widehat{d}_1} a + \mathbf{B}' \left(\mathbf{Q}(L)(1-L)^{\widehat{d}_1} \mathbf{x}_{t-1} - \mathbf{u}_{t-1} \right). \end{aligned}$$

This resembles the decomposition in the proof of AVOA (2020, Lemma A.9(b)) (cf., their equation (A.77)) and we rely on similar arguments. Next, we will establish results for the discrete Fourier transforms of each term in the decomposition. First, by AVOA (2020, Lemma A.9(a)), we have

$$w_{\widehat{\eta}}^{(b,1,1)}(\lambda_j) = \begin{cases} w_{\xi}(\lambda_j) (1 + O_p(\ln(n)/\sqrt{m_d})) & \text{under models (ii) and (iii),} \\ 0 & \text{under model (iv).} \end{cases} \quad (\text{B.41})$$

Second, by combining $\widehat{\mathbf{B}}(\ell, m) = O_p(1)$ and (B.20), we have

$$w_{\widehat{\eta}}^{(b,1,2)}(\lambda_j) = O_p \left(\frac{n^{1/2-d_x}}{j^{1-d_x}} \right) + O_p \left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j} \right) \quad (\text{B.42})$$

Third, by applying Theorem 1 and (B.20), we readily have

$$w_{\widehat{\eta}}^{(b,1,3)}(\lambda_j) = \begin{cases} O_p(1/\sqrt{m}) & \text{under models (ii) and (iii),} \\ O_p(\lambda_m^b/\sqrt{m}) & \text{under model (iv).} \end{cases} \quad (\text{B.43})$$

Fourth, by applying equation (A.63) and Lemmas A.6(a)-(c) of AVOA (2020) (as on their page 32), it follows that,

$$w_{\widehat{\eta}}^{(b,1,4)}(\lambda_j) = O_p \left(\frac{n^{1/2-b}}{j^{1-b}} \right). \quad (\text{B.44})$$

Hence, using $b = d_1$ and $\underline{d} \leq \underline{d}_x$, we may combine results to show,

$$w_{\hat{\eta}}^{(b,1)}(\lambda_j) = w_{\xi}(\lambda_j) + O_p(\ln(n)/\sqrt{m_d}) + O_p(1/\sqrt{m}) + O_p\left(\frac{n^{1/2-\underline{d}}}{j^{1-\underline{d}}}\right) + O_p\left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right) \quad (\text{B.45})$$

under models (ii)-(iii), and, similarly, that

$$w_{\hat{\eta}}^{(b,1)}(\lambda_j) = O_p(\lambda_m^b/\sqrt{m}) + O_p\left(\frac{n^{1/2-\underline{d}}}{j^{1-\underline{d}}}\right) + O_p\left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right) \quad (\text{B.46})$$

under model (iv). We are now ready to study the asymptotic properties of the long-run covariance estimate $\hat{\mathbf{G}}_{\hat{\eta}\hat{\eta}}^{(b,1)}(\ell_G, m_G)$ in the two inference regimes, with and without cointegration.

The case without cointegration. The discrete Fourier transform in (B.45) allows us to write,

$$\begin{aligned} \hat{\mathbf{G}}_{\hat{\eta}\hat{\eta}}^{(b,1)}(\ell_G, m_G) - \hat{\mathbf{G}}_{\xi\xi}(\ell_G, m_G) &\leq O_p(\ln(n)/\sqrt{m_d}) + O_p(1/\sqrt{m}) \\ &+ \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{n^{1/2}}{j} \left(\lambda_j^{\underline{d}} + \frac{\ln(n)}{m_d^{1/2}}\right)\right) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{n}{j^2} \left(\lambda_j^{2\underline{d}} + \frac{\ln(n)^2}{m_d} + \frac{\ln(n)\lambda_j^{\underline{d}}}{m_d^{1/2}}\right)\right) \\ &\leq O_p(\ln(n)/\sqrt{m_d}) + O_p(1/\sqrt{m}) + O_p\left(\frac{n^{1/2}}{m_G^{1-\epsilon}\ell_G^{1+\epsilon}} \left(\left(\frac{m_G}{n}\right)^{\underline{d}} + \frac{\ln(n)}{m_d^{1/2}}\right)\right) \\ &+ O_p\left(\frac{n}{m_G\ell_G^2} \left(\left(\frac{m_G}{n}\right)^{2\underline{d}} + \frac{\ln(n)^2}{m_d} + \frac{\ln(n)}{m_d^{1/2}} \left(\frac{m_G}{n}\right)^{\underline{d}}\right)\right) \end{aligned} \quad (\text{B.47})$$

similarly to (B.30), for some arbitrarily small $\epsilon > 0$, using, again, Varneskov (2017, Lemma C.4) and that the remaining cross-product terms of the errors are of strictly lower order by the tuning parameters satisfying $\ell_G \asymp n^{\nu_G}$, $m_G \asymp n^{\kappa_G}$ and $m_d \asymp n^{\varrho}$, with $0 < \nu_G < \kappa_G < \varrho \leq 1$ in Assumptions F and T-G. Next, as for equation (B.37), we have, by definition,

$$\left| \hat{\mathbf{G}}_{\xi\xi}^{(b,1)}(\ell_G, m_G) - \hat{\mathbf{G}}_{\xi\xi}(1, m_G) \right| = \left| \frac{\ell_G \hat{\mathbf{G}}_{\xi\xi}(1, \ell_G - 1)}{m_G - \ell_G + 1} \right| \leq O_p(\ell_G/m_G). \quad (\text{B.48})$$

Finally, since Christensen & Varneskov (2017, Lemma 6) provides $|\hat{\mathbf{G}}_{\xi\xi}(1, m_G) - G_{\xi\xi}| \leq o_p(1)$, we can combine this with (B.47), (B.48) and the triangle inequality to show $|\hat{\mathbf{G}}_{\hat{\eta}\hat{\eta}}^{(b,1)}(\ell_G, m_G) - G_{\xi\xi}| \leq o_p(1)$.

The case with cointegration. The discrete Fourier transform in (B.46) allows us to write,

$$\begin{aligned} \hat{\mathbf{G}}_{\hat{\eta}\hat{\eta}}^{(b,1)}(\ell_G, m_G) &\leq O_p(\lambda_m^{2b}/m) \\ &+ \frac{K\lambda_m^b}{m_G\sqrt{m}} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{n^{1/2}}{j} \left(\lambda_j^{\underline{d}} + \frac{\ln(n)}{m_d^{1/2}}\right)\right) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{n}{j^2} \left(\lambda_j^{2\underline{d}} + \frac{\ln(n)^2}{m_d} + \frac{\ln(n)\lambda_j^{\underline{d}}}{m_d^{1/2}}\right)\right) \\ &\leq O_p\left(\left(\frac{m}{n}\right)^{2b} \frac{1}{m}\right) + O_p\left(\frac{n^{1/2}}{m_G^{1-\epsilon}\ell_G^{1+\epsilon}\sqrt{m}} \left(\frac{m}{n}\right)^b \left(\left(\frac{m_G}{n}\right)^{\underline{d}} + \frac{\ln(n)}{m_d^{1/2}}\right)\right) \end{aligned}$$

$$+ O_p \left(\frac{n}{m_G \ell_G^2} \left(\left(\frac{m_G}{n} \right)^{2\bar{d}} + \frac{\ln(n)^2}{m_d} + \frac{\ln(n)}{m_d^{1/2}} \left(\frac{m_G}{n} \right)^{\bar{d}} \right) \right), \quad (\text{B.49})$$

using the same arguments as for (B.47). Now, by $\bar{d} - b \geq 0$ in conjunction with the conditions in Assumption T-G, we have $\lambda_{m_G}^{-2b} \widehat{\mathbf{G}}_{\eta\eta}^{(b,1)}(\ell_G, m_G) \leq o_p(1)$, providing the requisite result.

For (c), since we have by the Cauchy-Schwarz inequality,

$$\left| \widehat{\mathbf{G}}_{\eta\epsilon}^{(b,1,2)}(\ell_G, m_G) \right| \leq \sqrt{\widehat{\mathbf{G}}_{\eta\eta}^{(b,1)}(\ell_G, m_G) \widehat{\mathbf{G}}_{\epsilon\epsilon}^{(2)}(\ell_G, m_G)}, \quad (\text{B.50})$$

the convergence results follow by invoking (a) and (b), concluding the proof. \square

B.2 Auxiliary Lemmas for Section 4

This section establishes a sequence of lemmas that aid in proving the theoretical results in Section 4, that is, for the LCM-bias (LMCB) approach to estimation of d_1 . To this end, let us define

$$\check{\epsilon}_t = (1 - L)^{\widehat{\gamma}_x} \alpha + \mathbf{B}'(1 - L)^{\widehat{\gamma}_x - d_1} \mathbf{u}_{t-1} + (1 - L)^{\widehat{\gamma}_x - d_1} \xi_{t-1} + (1 - L)^{\widehat{\gamma}_x} \eta_t \equiv \sum_{i=1}^4 \check{\epsilon}_t^{(i)}. \quad (\text{B.51})$$

The first lemma collects DFT result for the four components in (B.51) from AVOA (2020).

Lemma B.3. *Suppose Assumptions D1-D3, C, M-d₁, F-d₁, T-d₁, T-G-d₁ hold. Moreover, suppose that the bandwidths satisfy $n^{1/2}/m \rightarrow 0$, $n^{1/2}/m_G \rightarrow 0$. Then, for $j = 1, \dots, m$, it follows,*

- (a) $w_{\epsilon}^{(1)}(\lambda_j) = O_p((j/n)^{\gamma_x} n^{1/2} j^{-1})$.
- (b) $w_{\epsilon}^{(2)}(\lambda_j) = \lambda_j^{\psi} e^{-(\pi/2)\psi i} \mathbf{B}' \mathbf{w}_u(\lambda_j) + O_p((j/n)^{\psi} n^{1/2} j^{-1}) + O_p(\ln(n) m_d^{-1/2} n^{1/2} j^{-1})$.
- (c) $w_{\epsilon}^{(3)}(\lambda_j) = \lambda_j^{\psi} e^{-(\pi/2)\psi i} w_{\xi}(\lambda_j) + O_p((j/n)^{\psi} n^{1/2} j^{-1}) + O_p(\ln(n) m_d^{-1/2} n^{1/2} j^{-1})$.
- (d) $w_{\epsilon}^{(4)}(\lambda_j) = \lambda_j^{\gamma_x} e^{-(\pi/2)\psi i} w_{\eta}(\lambda_j) + O_p((j/n)^{\gamma_x} n^{1/2} j^{-1}) + O_p(\ln(n) m_d^{-1/2} n^{1/2} j^{-1})$.

Proof. (a) follows by the Taylor expansion in equation (A.63) of AVOA (2020) in conjunction with their Lemmas A.6(a)-(c), since $\gamma_x > 0$. (b)-(d) follow by Lemmas A.8 and A.9(a) in AVOA (2020). \square

The next lemma provide bounds on the endogeneity bias for the LCM estimator, similarly to the results in Lemma B.1, when the former is implemented with $\widehat{\gamma}_x$ imposed as the memory on y_t . Specifically, we complement Lemma B.1 by providing equivalent bounds on $\widehat{\mathbf{F}}_{u\epsilon}^c(\ell, m)$ and $\widehat{\mathbf{G}}_{\eta\eta}^{(b,c)}(\ell_G, m_G)$.

Lemma B.4. *Suppose the conditions of Lemma B.3 hold. Then, for some arbitrarily small $\epsilon > 0$,*

- (a) $\lambda_m^{-1} (\widehat{\mathbf{F}}_{u\epsilon}^c(\ell, m) - \widehat{\mathbf{F}}_{u\epsilon}(\ell, m)) \leq O_p((m/n)^{\bar{d}_x + \psi} / \ell^{1+\epsilon})$.
- (b) $\widehat{\mathbf{G}}_{\eta\eta}^{(b,c)}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\eta\eta}^{(b)}(\ell_G, m_G) \leq O_p((m/n)^{2\psi + \bar{d}_x} / \ell^{1+\epsilon}) + O_p((m/n)^{2\psi + \bar{d}_x} m^{\epsilon/2} / \ell_G^{(1+\epsilon)/2})$.

Proof. First, for **(a)**, we have by Lemma B.3 that

$$w_{\tilde{e}}(\lambda_j) = O_p((j/n)^\psi) + O_p((j/n)^\psi n^{1/2} j^{-1}) + O_p(\ln(n) m_d^{-1/2} n^{1/2} j^{-1}), \quad (\text{B.52})$$

since $\psi \geq 0$ and $\psi \leq \gamma_x$. Hence, by (B.5), we can write $\widehat{\mathbf{F}}_{\widehat{u}\tilde{e}}^c(\ell, m) = \widehat{\mathbf{F}}_{\widehat{u}\tilde{e}}(\ell, m) + \widehat{\mathbf{F}}_{\widehat{c}\tilde{e}}(\ell, m)$, where

$$\begin{aligned} \widehat{\mathbf{F}}_{\widehat{c}\tilde{e}}(\ell, m) &\leq \frac{2\pi}{n} \sum_{j=\ell}^m O_p\left(\lambda_j^{\underline{d}_x+\psi}\right) + \frac{2\pi}{n} \sum_{j=\ell}^m O_p\left(\lambda_j^{\underline{d}_x+\psi} \frac{n^{1/2}}{j}\right) + \frac{2\pi}{n} \sum_{j=\ell}^m O_p\left(\frac{\ln(n) n^{1/2}}{m_d^{1/2} j}\right) \\ &\leq O_p\left(\left(\frac{m}{n}\right)^{1+\underline{d}_x+\psi} \frac{1}{\ell^{1+\epsilon}}\right) + O_p\left(\left(\frac{m}{n}\right)^{\underline{d}_x+\psi} \frac{m^\epsilon}{n^{1/2} \ell^{1+\epsilon}}\right) + O_p\left(\frac{m^\epsilon}{m_d^{1/2} n^{1/2} \ell^{1+\epsilon}}\right), \end{aligned} \quad (\text{B.53})$$

which provides the requisite result, since the second and third asymptotic bound are dominated by the first bound when $\sqrt{n}/m \rightarrow 0$, $m/m_d \rightarrow 0$, $\underline{d}_x \leq 1$ and $\psi \leq 1$.

For **(b)**. The proof proceeds similarly to that of Lemma B.1(e). First, let us write $\check{\eta}_t^{(b)} = \check{e}_t - \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x)' \widehat{\mathbf{u}}_{t-1}$, where $\widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x)$ is defined in equation (B.68) and corresponds to $\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x)$, but free from endogeneity-related bias. Moreover, using the notation from (B.51), define

$$\check{e}_t^{(1)} \equiv \check{e}_t^{(1)} + \check{e}_t^{(2)} + \check{e}_t^{(3)}, \quad \check{e}_t^{(2)} \equiv \check{e}_t^{(4)}, \quad \check{\eta}_t^{(b,1)} = \check{e}_t^{(1)} - \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x)' \widehat{\mathbf{u}}_{t-1}, \quad \check{\eta}_t^{(b)} = \check{\eta}_t^{(b,1)} + \check{e}_t^{(2)}, \quad (\text{B.54})$$

such that, with $\check{\tau}_{t-1}^{(1)} = (\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x) - \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x))' \widehat{\mathbf{u}}_{t-1}^c$ and $\check{\tau}_{t-1}^{(2)} = \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x)' \widehat{\mathbf{c}}_{t-1}$, we have

$$\check{\eta}_t^{(b,c)} = \check{e}_t - \widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x)' \widehat{\mathbf{u}}_{t-1}^c = \check{\eta}_t^{(b,1)} + \check{e}_t^{(2)} - \check{\tau}_{t-1}^{(1)} - \check{\tau}_{t-1}^{(2)} = \check{\eta}_t^{(b)} - \check{\tau}_{t-1}^{(1)} - \check{\tau}_{t-1}^{(2)}. \quad (\text{B.55})$$

Next, using these, we make the decomposition,

$$\begin{aligned} \widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b,c)}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b)}(\ell_G, m_G) &= \widehat{\mathbf{G}}_{\check{\tau}\check{\tau}}^{(1,1)}(\ell_G, m_G) + \widehat{\mathbf{G}}_{\check{\tau}\check{\tau}}^{(2,2)}(\ell_G, m_G) + 2\widehat{\mathbf{G}}_{\check{\tau}\check{\tau}}^{(1,2)}(\ell_G, m_G) \\ &\quad - 2\widehat{\mathbf{G}}_{\check{\eta}\check{\tau}}^{(b,1)}(\ell_G, m_G) - 2\widehat{\mathbf{G}}_{\check{\eta}\check{\tau}}^{(b,2)}(\ell_G, m_G), \end{aligned} \quad (\text{B.56})$$

where the first three terms are long-run (co)variance estimates for $\check{\tau}_{t-1}^{(1)}$ and $\check{\tau}_{t-1}^{(2)}$, and the final two terms are their respective long-run covariances with $\check{\eta}_t^{(b)}$. Next, let us consider,

$$w_{\check{\tau}}^{(1)}(\lambda_j) = \left(\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x) - \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x)\right) \left(\mathbf{w}_{\widehat{u}}(\lambda_j) + \mathbf{w}_{\widehat{c}}(\lambda_j)\right) \leq O_p\left((m/n)^{\underline{d}_x+\psi} / \ell^{1+\epsilon}\right), \quad (\text{B.57})$$

using equations (B.5), (B.20) and (B.84). Hence, we readily have

$$\widehat{\mathbf{G}}_{\check{\tau}\check{\tau}}^{(1,1)}(\ell_G, m_G) \leq O_p\left((m/n)^{2(\underline{d}_x+\psi)} / \ell^{2(1+\epsilon)}\right). \quad (\text{B.58})$$

Similarly, by Theorem 3 and (B.20), $w_{\check{\tau}}^{(2)}(\lambda_j) \leq O_p(\lambda_m^\psi \lambda_j^{\underline{d}_x})$ such that, by $m = m_G$ from Assumption

T-G- d_1 ,

$$\widehat{\mathbf{G}}_{\check{\tau}\check{\tau}}^{(2,2)}(\ell_G, m_G) \leq \frac{\lambda_m^{2\psi} K}{m_G} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{2\underline{d}_x}) \leq O_p\left((m/n)^{2(\underline{d}_x+\psi)} m^\epsilon / \ell_G^{1+\epsilon}\right). \quad (\text{B.59})$$

Next, since $\widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b)}(\ell_G, m_G) = O_p((m_G/n)^{2\psi})$ by Lemma B.9(c), the Cauchy-Schwarz inequality provides

$$\begin{aligned} \widehat{\mathbf{G}}_{\check{\tau}\check{\tau}}^{(1,2)}(\ell_G, m_G) &\leq O_p\left((m/n)^{2(\underline{d}_x+\psi)} m^{\epsilon/2} / (\ell^{1+\epsilon} \ell_G^{(1+\epsilon)/2})\right), \\ \widehat{\mathbf{G}}_{\check{\eta}\check{\tau}}^{(b,1)}(\ell_G, m_G) &\leq O_p\left((m/n)^{2\psi+\underline{d}_x} / \ell^{1+\epsilon}\right), \\ \widehat{\mathbf{G}}_{\check{\eta}\check{\tau}}^{(b,2)}(\ell_G, m_G) &\leq O_p\left((m/n)^{2\psi+\underline{d}_x} m^{\epsilon/2} / \ell_G^{(1+\epsilon)/2}\right), \end{aligned}$$

Hence, since $0 < \underline{d}_x \leq 1$ and $\psi \geq 0$, the dominant asymptotic bounds are provided by the covariance terms $\widehat{\mathbf{G}}_{\check{\eta}\check{\tau}}^{(b,1)}(\ell_G, m_G)$ and $\widehat{\mathbf{G}}_{\check{\eta}\check{\tau}}^{(b,2)}(\ell_G, m_G)$, thereby establishing the requisite result. \square

The next lemma turns attention to the TDAC between $\widehat{\mathbf{u}}_{t-1}$ and \check{e}_t without an endogenous component in the regressors. Specifically, using the decomposition in (B.51), we can write,

$$\widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}(\ell, m) \equiv \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(1)}(\ell, m) + \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(2)}(\ell, m) + \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(3)}(\ell, m) + \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(4)}(\ell, m),$$

and establish convergence results for each term separately. To this end, we further define

$$\begin{aligned} \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(2,1)}(\ell, m) &= \frac{2\pi}{n} \sum_{j=\ell}^m \Re\left(\mathbf{w}_{\widehat{\mathbf{u}}}(\lambda_j) \overline{\mathbf{w}}_{\check{e}}^{(2,1)}(\lambda_j)\right), \quad \mathbf{w}_{\check{e}}^{(2,1)}(\lambda_j) = \lambda_j^\psi e^{-(\pi/2)\psi i} \mathbf{B}' \mathbf{w}_u(\lambda_j), \\ \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(3,1)}(\ell, m) &= \frac{2\pi}{n} \sum_{j=\ell}^m \Re\left(\mathbf{w}_{\widehat{\mathbf{u}}}(\lambda_j) \overline{\mathbf{w}}_{\check{e}}^{(3,1)}(\lambda_j)\right), \quad \mathbf{w}_{\check{e}}^{(3,1)}(\lambda_j) = \lambda_j^\psi e^{-(\pi/2)\psi i} \mathbf{w}_\xi(\lambda_j), \\ \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(4,1)}(\ell, m) &= \frac{2\pi}{n} \sum_{j=\ell}^m \Re\left(\mathbf{w}_{\widehat{\mathbf{u}}}(\lambda_j) \overline{\mathbf{w}}_{\check{e}}^{(4,1)}(\lambda_j)\right), \quad \mathbf{w}_{\check{e}}^{(4,1)}(\lambda_j) = \lambda_j^{\gamma_x} e^{-(\pi/2)\gamma_x i} \mathbf{w}_\eta(\lambda_j), \end{aligned}$$

corresponding, again, the TDACs for the different components in (B.51), albeit without the (higher-order) terms that inducing approximation errors in the representations in Lemma B.3.

Lemma B.5. *Suppose the conditions of Lemma B.3 hold. Then, for some arbitrarily small $\epsilon > 0$,*

- (a) $\lambda_m^{-1} \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(1)}(\ell, m) \leq O_p((m/n)^{\gamma_x} n^{1/2} / (m^{1-\epsilon} \ell^{1+\epsilon})) + O_p((m/n)^{\gamma_x+\underline{d}_x} n / (m \ell^2)).$
- (b) $\lambda_m^{-1} (\widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(2)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(2,1)}(\ell, m)) \leq O_p((m/n)^\psi n^{1/2} / (m^{1-\epsilon} \ell^{1+\epsilon})) + O_p((m/n)^{\psi+\underline{d}_x} n / (m \ell^2)).$
- (c) $\lambda_m^{-1} (\widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(3)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(3,1)}(\ell, m)) \leq O_p((m/n)^\psi n^{1/2} / (m^{1-\epsilon} \ell^{1+\epsilon})) + O_p((m/n)^{\psi+\underline{d}_x} n / (m \ell^2)).$
- (d) $\lambda_m^{-1} (\widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(4)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{\mathbf{u}}\check{e}}^{(4,1)}(\ell, m)) \leq O_p((m/n)^{\gamma_x} n^{1/2} / (m^{1-\epsilon} \ell^{1+\epsilon})) + O_p((m/n)^{\gamma_x+\underline{d}_x} n / (m \ell^2)).$

Proof. First, for **(a)**, we may combine (B.20) and Lemma B.3(a) to write,

$$\begin{aligned}\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(1)}(\ell, m) &\leq \frac{2\pi}{n} \sum_{j=\ell}^m O_p \left(\lambda_j^{\gamma_x} \frac{n^{1/2}}{j} \right) + \frac{2\pi}{n} \sum_{j=\ell}^m O_p \left(\lambda_j^{\gamma_x + \underline{d}_x} \frac{n}{j^2} \right) + \frac{2\pi}{n} \sum_{j=\ell}^m O_p \left(\lambda_j^{\gamma_x} \frac{\ln(n)n}{m_d^{1/2} j^2} \right) \\ &\leq O_p \left(\left(\frac{m}{n} \right)^{\gamma_x} \frac{m^\epsilon}{n^{1/2} \ell^{1+\epsilon}} \right) + O_p \left(\left(\frac{m}{n} \right)^{\gamma_x + \underline{d}_x} \frac{1}{\ell^2} \right) + O_p \left(\left(\frac{m}{n} \right)^{\gamma_x} \frac{\ln(n)}{m_d^{1/2} \ell^2} \right).\end{aligned}\quad (\text{B.60})$$

Since the third term is dominated by the second as $m/m_d \rightarrow 0$, $\sqrt{n}/m \rightarrow 0$ and $\underline{d}_x \leq 1$, this gives the requisite bound. Second, for **(b)**, we may combine (B.20) and Lemma B.3(b) to write,

$$\begin{aligned}\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(1)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m) &\leq \frac{2\pi}{n} \sum_{j=\ell}^m O_p \left(\frac{n^{1/2}}{j} \left(\lambda_j^\psi + \frac{\ln(n)}{\sqrt{m_d}} + \lambda_j^{\psi + \underline{d}_x} + \lambda_j^\psi \frac{\ln(n)}{\sqrt{m_d}} \right) \right) \\ &\quad + \frac{2\pi}{n} \sum_{j=\ell}^m O_p \left(\frac{n}{j^2} \left(\lambda_j^{\psi + \underline{d}_x} + \frac{\ln(n)^2}{m_d} + \lambda_j^{\underline{d}_x} \frac{\ln(n)}{\sqrt{m_d}} + \lambda_j^\psi \frac{\ln(n)}{\sqrt{m_d}} \right) \right) \\ &\leq O_p \left(\frac{m^\epsilon}{n^{1/2} \ell^{1+\epsilon}} \left(\left(\frac{m}{n} \right)^\psi + \frac{\ln(n)}{\sqrt{m_d}} \right) \right) + \\ &\quad + O_p \left(\frac{1}{\ell^2} \left(\left(\frac{m}{n} \right)^{\psi + \underline{d}_x} + \frac{\ln(n)^2}{m_d} + \left(\frac{m}{n} \right)^{\underline{d}_x} \frac{\ln(n)}{\sqrt{m_d}} + \left(\frac{m}{n} \right)^\psi \frac{\ln(n)}{\sqrt{m_d}} \right) \right) \\ &= O_p \left(\frac{m^\epsilon}{n^{1/2} \ell^{1+\epsilon}} \left(\frac{m}{n} \right)^\psi \right) + O_p \left(\frac{1}{\ell^2} \left(\frac{m}{n} \right)^{\psi + \underline{d}_x} \right)\end{aligned}\quad (\text{B.61})$$

using, similarly, $m/m_d \rightarrow 0$, $\sqrt{n}/m \rightarrow 0$ and $0 \leq \psi \leq 1$ for the final equality. This gives the requisite bound. Finally, **(c)** and **(d)** follow by the same arguments as **(b)**, concluding the proof. \square

Next, let us define

$$\widetilde{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m) \equiv \lambda_m^\psi \frac{\cos(\psi\pi/2)}{1 + \psi} \widehat{\mathbf{F}}_{uu}(\ell, m) \mathbf{B}, \quad (\text{B.62})$$

concentrate on the differences,

$$\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m) - \widetilde{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m), \quad \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(1, m), \quad \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(1, m), \quad (\text{B.63})$$

and study the properties of $\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(1, m)$ and $\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(1, m)$.

Lemma B.6. *Suppose the conditions of Lemma B.3 hold. Then, for some arbitrarily small $\epsilon > 0$,*

$$(a) \quad \lambda_m^{-1-\psi} \left(\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m) - \widetilde{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m) \right) \leq O_p(m^{-1}) + O_p(m^{-1/2} \lambda_m^{\varpi/2}) + O_p(\lambda_m^{\varpi}).$$

$$(b) \quad \sqrt{m} \lambda_m^{-1-\psi} \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(1, m) \xrightarrow{\mathbb{D}} N(\mathbf{0}, \mathbf{G}_{uu} G_{\xi\xi} / (2(1 + 2\psi))), \text{ under models (ii) and (iii).}$$

$$(c) \quad \sqrt{m} \lambda_m^{-1-\gamma_x} \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(1, m) \xrightarrow{\mathbb{D}} N(\mathbf{0}, \mathbf{G}_{uu} G_{\eta\eta} / (2(1 + 2\gamma_x))).$$

Proof. First, for **(a)**, recall that,

$$\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m) = \frac{2\pi}{n} \sum_{j=\ell}^m \lambda_j^\psi \Re(e^{(\pi/2)\psi i} \mathbf{w}_u(\lambda_j) \overline{\mathbf{w}}_u(\lambda_j)) \mathbf{B}.$$

Hence, using $\Re(e^{(\pi/2)\psi i}) = \cos(\pi\psi/2)$, we seek to establish an error bound for the decomposition,

$$\begin{aligned} \lambda_m^{-1-\psi} (\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m) - \widetilde{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m)) &= \frac{\cos(\psi\pi/2)}{m} \sum_{j=\ell}^m \Re(\mathbf{I}_{uu}(\lambda_j)) \mathbf{B} \left(\left(\frac{j}{m} \right)^\psi - \frac{1}{1+\psi} \right) \\ &= \frac{\cos(\psi\pi/2)}{m} \sum_{j=\ell}^m \mathbf{G}_{uu} \mathbf{B} \left(\left(\frac{j}{m} \right)^\psi - \frac{1}{1+\psi} \right) \\ &\quad + \frac{\cos(\psi\pi/2)}{m} \sum_{j=\ell}^m (\Re(\mathbf{I}_{uu}(\lambda_j)) - \mathbf{G}_{uu}) \mathbf{B} \left(\left(\frac{j}{m} \right)^\psi - \frac{1}{1+\psi} \right) \equiv \mathbf{\mathcal{E}}_1 + \mathbf{\mathcal{E}}_2. \end{aligned} \quad (\text{B.64})$$

First, for $\mathbf{\mathcal{E}}_1$, we have

$$\|\mathbf{\mathcal{E}}_1\| \leq K \left| \frac{1}{m} \sum_{j=\ell}^m \left(\left(\frac{j}{m} \right)^\psi - \frac{1}{1+\psi} \right) \right| \leq O(m^{-1}), \quad (\text{B.65})$$

using Shimotsu & Phillips (2005, Lemma 5.4). Next, for $\mathbf{\mathcal{E}}_2$, we may use Assumptions D1-D3 to invoke Theorem 2 and Corollary 1 on Hannan (1970, pp. 248-249)²⁵, providing,

$$\begin{aligned} \mathbb{E}[\Re(\mathbf{I}_{uu}(\lambda_j))] &= \Re(\mathbf{f}_{uu}(\lambda_j)) + O(n^{-1}), \quad \|\mathbf{f}_{uu}(\lambda_j) - \mathbf{G}_{uu}\| = O(\lambda_j^\varpi), \\ \mathbb{V}[\Re(\mathbf{I}_{uu}(\lambda_j))] &= \Re(\mathbf{f}_{uu}(\lambda_j) \otimes \mathbf{f}_{uu}(\lambda_j)) + O(n^{-1}), \quad \text{Cov}[\Re(\mathbf{I}_{uu}(\lambda_j), \Re(\mathbf{I}_{uu}(\lambda_k)))] = O(n^{-1}), \end{aligned}$$

for $j, k = 1, \dots, m$, $m/n \rightarrow 0$ and $j \neq k$. Hence, we obtain the following bound,

$$\mathbb{E}[\mathbf{\mathcal{E}}_2] \leq \frac{K}{m} \sum_{j=\ell}^m \|\mathbf{I}_{uu}(\lambda_j) - \mathbf{G}_{uu}\| \times \left| \left(\frac{j}{m} \right)^\psi - \frac{1}{1+\psi} \right| \leq O(n^{-1}) + O(\lambda_m^\varpi). \quad (\text{B.66})$$

Moreover, using the (co-)periodogram second moment results, we have

$$\begin{aligned} \mathbb{V}[\mathbf{\mathcal{E}}_2] &= \cos(\pi\psi/2)^2 \frac{1}{m^2} \sum_{j=\ell}^m \left(\left(\frac{j}{m} \right)^\psi - \frac{1}{1+\psi} \right) \sum_{k=\ell}^m \left(\left(\frac{k}{m} \right)^\psi - \frac{1}{1+\psi} \right) \\ &\quad \times \mathbb{E}[(\mathbf{I}_{uu}(\lambda_j) - \mathbf{G}_{uu}) \otimes (\mathbf{I}_{uu}(\lambda_k) - \mathbf{G}_{uu})] \\ &\leq K(1 + O(n^{-1})) \left| \frac{1}{m} \sum_{j=\ell}^m \left(\left(\frac{j}{m} \right)^\psi - \frac{1}{1+\psi} \right) \right|^2 + \frac{1}{m^2} \sum_{j=\ell}^m \left| \left(\frac{j}{m} \right)^\psi - \frac{1}{1+\psi} \right|^2 O(\lambda_m^\varpi(1 + \lambda_m^\varpi)) \end{aligned}$$

²⁵See also Parzen (1957, Theorem 4) and Brockwell & Davis (1991, Theorem 10.3.2) for similar results.

$$\leq O(m^{-2}) + O(m^{-1}\lambda_m^\varpi). \quad (\text{B.67})$$

The moment results for \mathcal{E}_2 readily implies $\mathcal{E}_2 \leq O_p(m^{-1}) + O_p(m^{-1/2}\lambda_m^{\varpi/2}) + O_p(\lambda_m^\varpi)$, which, together with the asymptotic bound in equation (B.65), provides the requisite result.

Finally, (b) and (c) follow directly from AVOA (2020, Lemma A.3), concluding the proof. \square

Lemma B.7. *Suppose the conditions of Lemma B.3 hold. Then, for some arbitrarily small $\epsilon > 0$,*

$$(a) \quad \sqrt{m}\lambda_m^{-1-\psi} \left(\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(1, m) \right) \leq O_p(\ell^{1+\varpi}/(m^{1/2}n^\varpi)(\ell/m)^\psi).$$

$$(b) \quad \sqrt{m}\lambda_m^{-1-\gamma_x} \left(\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(1, m) \right) \leq O_p(\ell^{1+\varpi}/(m^{1/2}n^\varpi)(\ell/m)^{\gamma_x}).$$

Proof. We have $\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(i,1)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(i,1)}(1, m) = -\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(i,1)}(1, \ell - 1)$, $i = 3, 4$. Hence, the results follow by applying AVOA (2020, Lemma A.1(d)) for both (a) and (b). \square

Lemma B.8. *Suppose the conditions of Lemma B.3 hold. Moreover, assume that the conditions $0 < b = d_1 < 1$, $b \leq \underline{d}$, $n^{1/2}/m \rightarrow 0$, and $\max(0, (1 - 3\kappa/2)/(1 + \kappa/2), 2b) < \varpi \leq 2$ are satisfied, then,*

$$\begin{cases} \sqrt{m}\lambda_m^{-\psi} \left(\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x) - \lambda_m^\psi c(\psi)\mathbf{B} \right) \xrightarrow{\mathbb{D}} N(\mathbf{0}, \mathbf{V}(\xi, \psi)), & \text{under models (ii) and (iii),} \\ \sqrt{m}\lambda_m^{-\gamma_x} \left(\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x) - \lambda_m^\psi c(\psi)\mathbf{B} \right) \xrightarrow{\mathbb{D}} N(\mathbf{0}, \mathbf{V}(\eta, \gamma_x)), & \text{under model (iv).} \end{cases}$$

Proof. First, recall that $\check{\mathbf{v}}_t = (\check{e}_t, \widehat{\mathbf{u}}'_{t-1})'$, then, by invoking Lemmas B.1(a) and B.4(a) as well as the continuous mapping theorem, while recalling that $\psi = \gamma_x - b$,

$$\sqrt{m}\lambda_m^{-\gamma_x} \left(\widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x) - \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x) \right) \leq O_p \left((m/n)^{\underline{d}_x - b} \sqrt{m}/\ell^{1+\epsilon} \right), \quad (\text{B.68})$$

for some arbitrarily small $\epsilon > 0$, where $\widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x) = \widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1} \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}(\ell, m)$. Hence, we continue working with the estimate without endogeneity, $\check{\mathbf{v}}_t$. Next, using the definitions in Lemma B.6, we have

$$\sqrt{m}\lambda_m^{-1-\gamma_x} \left(\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}(\ell, m) - \widetilde{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(2,1)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(\ell, m) - \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(\ell, m) \right) = \sum_{i=1}^6 \mathcal{A}_i, \quad (\text{B.69})$$

where the asymptotic bounds on the right-hand-side error terms are,

$$\begin{aligned} \mathcal{A}_1 &\leq O_p \left(\left(\frac{n}{m} \right)^{1/2+b} \frac{m^\epsilon}{\ell^{1+\epsilon}} \right), \quad \mathcal{A}_2 \leq O_p \left(\left(\frac{m}{n} \right)^{\underline{d}_x - b} \frac{n}{m^{1/2}\ell^2} \right), \quad \mathcal{A}_3 \leq O_p \left(\left(\frac{n}{m} \right)^b \frac{1}{m^{1/2}} \right), \\ \mathcal{A}_4 &\leq O_p \left(\left(\frac{m}{n} \right)^{\varpi/2-b} \right), \quad \mathcal{A}_5 \leq O_p \left(\left(\frac{m}{n} \right)^{\varpi-b} m^{1/2} \right), \quad \mathcal{A}_6 \leq O_p \left(\left(\frac{n}{m} \right)^b \left(\frac{\ell}{m} \right)^\psi \frac{\ell^{1+\varpi}}{m^{1/2}n^\varpi} \right). \end{aligned}$$

using Lemma B.5 for \mathcal{A}_1 and \mathcal{A}_2 ; Lemma B.6 for \mathcal{A}_3 , \mathcal{A}_4 and \mathcal{A}_5 ; and Lemma B.7 for \mathcal{A}_6 . These are

all $o_p(1)$ by Assumption T- d_1 , $0 < b < 1$ and the mutual consistency condition. Moreover, we have

$$\begin{aligned}\widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m) - \widehat{\mathbf{F}}_{uu}(\ell, m) &\leq \frac{2\pi}{n} \sum_{j=\ell}^m O_p \left(\frac{n^{1/2}}{j} \left(\lambda_j^{d_x} + \frac{\ln(n)}{\sqrt{m_d}} \right) \right) \\ &\leq \frac{2\pi}{n} \sum_{j=\ell}^m O_p \left(\frac{n}{j^2} \left(\lambda_j^{2d_x} + \lambda_j^{d_x} \frac{\ln(n)}{\sqrt{m_d}} + \frac{\ln(n)^2}{m_d} \right) \right) \\ &\leq O_p \left(\left(\frac{m}{n} \right)^{d_x} \frac{m^\epsilon}{n^{1/2} \ell^{1+\epsilon}} \right) + O_p \left(\left(\frac{m}{n} \right)^{2d_x} \frac{1}{\ell^2} \right),\end{aligned}\tag{B.70}$$

using (B.20), $\underline{d}_x \leq 1$, $m/m_d \rightarrow 0$ and $\sqrt{m}/n \rightarrow 0$. Hence

$$\sqrt{m} \lambda_m^{-1-\gamma_x} \left(\widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m) - \widehat{\mathbf{F}}_{uu}(\ell, m) \right) \leq O_p \left(\left(\frac{n}{m} \right)^{1/2+\gamma_x-\underline{d}_x} \frac{m^\epsilon}{\ell^{1+\epsilon}} \right) + O_p \left(\left(\frac{m}{n} \right)^{2\underline{d}_x-\gamma_x} \frac{n}{m^{1/2} \ell^2} \right),$$

which is $o_p(1)$ using, again, the conditions in Assumption T- d_1 . Furthermore, we have $\lambda_m^{-1} \widehat{\mathbf{F}}_{uu}(\ell, m) \xrightarrow{\mathbb{P}} \mathbf{G}_{uu}$ by AVOA (2020, Lemmas A.1 and A.2). Hence, we can combine results to write,

$$\sqrt{m} \lambda_m^{-\gamma_x} \left(\widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x) - \lambda_m^\psi c(\psi) \mathbf{B} \right) = \sqrt{m} \lambda_m^{-\gamma_x} \widehat{\mathbf{F}}_{uu}(\ell, m)^{-1} \left(\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(\ell, m) + \widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(\ell, m) \right) + o_p(1).$$

The final results in (a) and (b) follow by applying Lemmas B.6(b) and (c) with scale factors $\sqrt{m} \lambda_m^{-\psi}$ and $\sqrt{m} \lambda_m^{-\gamma_x}$, respectively, to the right-hand-side terms. In models (ii) and (iii), the limit for $\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(\ell, m)$ dominates the corresponding for $\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(\ell, m)$, which becomes a higher-order error. In contrast, in model (iv), we have $\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(3,1)}(\ell, m) = \mathbf{0}$ and the limit is driven by $\widehat{\mathbf{F}}_{\widehat{u}\widehat{e}}^{(4,1)}(\ell, m)$. The requisite central limit theorems, thus, follow by the continuous mapping theorem and Slutsky's theorem.

Besides requiring additional tuning parameter restrictions in Assumption T- d_1 as well as the conditions $0 < b < 1$ and $2b < \varpi$, there are no differences between this and the corresponding treatment of the mutual consistency condition in Theorem 1, concluding this proof. \square

The next lemma establishes bounds and convergence results for $\widehat{\mathbf{G}}_{\eta\eta}^{(b)}(\ell_G, m_G)$ in the cases with and without cointegration, similarly to Lemma B.2. To this end, recall from equation (B.51) and the proof of Lemma B.4 that we have $\check{\eta}_t^{(b)} = \check{\eta}_t^{(b,1)} + \check{e}_t^{(2)}$, where $\check{\eta}_t^{(b,1)} = \check{e}_t^{(1)} - \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x)' \widehat{\mathbf{u}}_{t-1}$ and the two terms $(\check{e}_t^{(1)}, \check{e}_t^{(2)})$ are defined in equation (B.54). Hence, we can equivalently write,

$$\widehat{\mathbf{G}}_{\eta\eta}^{(b)}(\ell_G, m_G) = \widehat{\mathbf{G}}_{\eta\eta}^{(b,1)}(\ell_G, m_G) + \widehat{\mathbf{G}}_{\check{e}\check{e}}^{(2)}(\ell_G, m_G) + 2\widehat{\mathbf{G}}_{\eta\check{e}}^{(b,1,2)}(\ell_G, m_G),\tag{B.71}$$

whose components will be analyzed separately in the following.

Lemma B.9. *Suppose the conditions of Lemma B.4 hold. Then,*

$$(a) \quad \lambda_{m_G}^{-2\gamma_x} \left(\widehat{\mathbf{G}}_{\check{e}\check{e}}^{(2)}(\ell_G, m_G) - G_{\eta\eta}/(1 + 2\gamma_x) \right) = o_p(1),$$

(b) The following convergence results hold,

$$\begin{cases} \lambda_{m_G}^{-2\psi} \left(\widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b,1)}(\ell_G, m_G) - G_{\xi\xi}/(1+2\psi) \right) = o_p(1), & \text{in (ii),} \\ \lambda_{m_G}^{-2\psi} \left(\widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b,1)}(\ell_G, m_G) - G_{\xi\xi}/(1+2\psi) - (1/(1+2\psi) - c(\psi)^2) \mathbf{B}' \mathbf{G}_{uu} \mathbf{B} \right) = o_p(1), & \text{in (iii),} \\ \lambda_{m_G}^{-2\psi} \left(\widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b,1)}(\ell_G, m_G) - (1/(1+2\psi) - c(\psi)^2) \mathbf{B}' \mathbf{G}_{uu} \mathbf{B} \right) = o_p(1), & \text{in (iv).} \end{cases}$$

(c) The following convergence results hold,

$$\lambda_{m_G}^{-2\psi} \widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b)}(\ell_G, m_G) \xrightarrow{\mathbb{P}} \begin{cases} G_{\xi\xi}/(1+2\psi), & \text{in model (ii),} \\ G_{\xi\xi}/(1+2\psi) + (1/(1+2\psi) - c(\psi)^2) \mathbf{B}' \mathbf{G}_{uu} \mathbf{B}, & \text{in model (iii),} \\ (1/(1+2\psi) - c(\psi)^2) \mathbf{B}' \mathbf{G}_{uu} \mathbf{B}, & \text{in model (iv).} \end{cases}$$

Proof. First, **(a)** follows by the same arguments as Lemma B.2(a). Second, for **(b)**, we follow the same steps as in the proof of Lemma B.2(b), implying we will study the properties of $\widehat{\mathbf{G}}_{\check{\eta}\check{\eta}}^{(b,1)}(\ell_G, m_G)$ under the model scenarios (ii)-(iii) as well as the cointegration setting (iv). To this end, let us write

$$\check{\eta}_t^{(b,1)} = \check{\eta}_t^{(b,1,1)} + \check{\eta}_t^{(b,1,2)} + \check{\eta}_t^{(b,1,3)} + \check{\eta}_t^{(b,1,4)} \quad (\text{B.72})$$

where, by addition and subtraction, the components are defined as,

$$\begin{aligned} \check{\eta}_t^{(b,1,1)} &= (1-L)^{\widehat{\gamma}_x - \gamma_x} \xi_{t-1}^{(\psi)}, \quad \check{\eta}_t^{(b,1,2)} = \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x)' (\mathbf{u}_{t-1} - \widehat{\mathbf{u}}_{t-1}), \\ \check{\eta}_t^{(b,1,3)} &= (\mathbf{B} - \widehat{\mathbf{B}}(\ell, m, \widehat{\gamma}_x))' \mathbf{u}_{t-1}, \quad \check{\eta}_t^{(b,1,4)} = (1-L)^{\widehat{\gamma}_x - d_1} \mathbf{u}_{t-1} - \mathbf{u}_{t-1}. \end{aligned}$$

The proof for the case $\psi = 0$ follows by the same arguments as Lemma B.2(b). Hence, we focus on the case $\psi > 0$ and proceed by initially establishing results for the respective discrete Fourier transforms of each term in the decomposition. First, by Lemma B.3(c), we have

$$w_{\check{\eta}}^{(b,1,1)}(\lambda_j) = \begin{cases} \lambda_j^\psi e^{-(\pi/2)\psi i} w_\xi(\lambda_j) + O_p\left(\frac{n^{1/2-\psi}}{j^{1-\psi}}\right) + O_p\left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right), & \text{in models (ii) and (iii),} \\ 0, & \text{in model (iv).} \end{cases} \quad (\text{B.73})$$

Second, by combining $\widehat{\mathbf{B}}(\ell, m) = O_p(\lambda_m^\psi)$ in Theorem 3 and (B.20), we have

$$w_{\check{\eta}}^{(b,1,2)}(\lambda_j) = O_p\left(\lambda_m^\psi \frac{n^{1/2-d_x}}{j^{1-d_x}}\right) + O_p\left(\lambda_m^\psi \frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right) \quad (\text{B.74})$$

Third, by applying Theorem 3 and (B.20), we have

$$w_{\tilde{\eta}}^{(b,1,3)}(\lambda_j) = \begin{cases} O_p(\lambda_m^\psi m^{-1/2}), & \text{in model (ii),} \\ (1 - c(\psi)\lambda_m^\psi) \mathbf{B}' \mathbf{w}_u(\lambda_j) + O_p(\lambda_m^\psi m^{-1/2}), & \text{in model (iii),} \\ (1 - c(\psi)\lambda_m^\psi) \mathbf{B}' \mathbf{w}_u(\lambda_j) + O_p(\lambda_m^{\gamma_x} m^{-1/2}), & \text{in model (iv).} \end{cases} \quad (\text{B.75})$$

Fourth, by the same arguments as given when establishing Lemma B.2(a)-(b), we have

$$w_{\tilde{\eta}}^{(b,1,4)}(\lambda_j) = \begin{cases} O_p\left(\frac{n^{1/2-\gamma_x}}{j^{1-\gamma_x}}\right), & \text{in model (ii),} \\ -(1 - \lambda_j^\psi e^{-(\pi/2)\psi i}) \mathbf{B}' \mathbf{w}_u(\lambda_j) + O_p\left(\frac{n^{1/2-\psi}}{j^{1-\psi}}\right) + O_p\left(\frac{\ln(n)n^{1/2}}{m_d^{1/2}j}\right), & \text{in (iii)-(iv).} \end{cases} \quad (\text{B.76})$$

Hence, in model (ii), it follows by $0 < \underline{d}_x < 1$, $\psi \geq 0$ and $\psi \leq \gamma_x$,

$$w_{\tilde{\eta}}^{(b,1)}(\lambda_j) = \lambda_j^\psi e^{-(\pi/2)\psi i} w_\xi(\lambda_j) + O_p\left(\frac{n^{1/2}}{j} \left(\lambda_j^\psi + \lambda_j^{\underline{d}_x} \lambda_m^\psi + \frac{\ln(n)}{m_d^{1/2}}\right)\right) + O_p\left(\frac{\lambda_m^\psi}{m^{1/2}}\right). \quad (\text{B.77})$$

Similarly, in model (iii), we have

$$\begin{aligned} w_{\tilde{\eta}}^{(b,1)}(\lambda_j) &= \lambda_j^\psi e^{-(\pi/2)\psi i} w_\xi(\lambda_j) + O_p\left(\frac{n^{1/2}}{j} \left(\lambda_j^\psi + \lambda_j^{\underline{d}_x} \lambda_m^\psi + \frac{\ln(n)}{m_d^{1/2}}\right)\right) + O_p\left(\frac{\lambda_m^\psi}{m^{1/2}}\right) \\ &\quad + \left(\lambda_j^\psi e^{-(\pi/2)\psi i} - c(\psi)\lambda_m^\psi\right) \mathbf{B}' \mathbf{w}_u(\lambda_j), \end{aligned} \quad (\text{B.78})$$

and model (iv) exhibits a DFT on the same form, but having $w_\xi(\lambda_j) = 0$ as well as the error term $O_p(\lambda_m^\psi m^{-1/2})$ being replaced by one of order $O_p(\lambda_m^{\gamma_x} m^{-1/2})$, with $\psi \leq \gamma_x$. Hence, we will explicitly treat model (iii) in the following since the same arguments may readily be applied to establish the corresponding results for models (ii) and (iv). To this end, define

$$\begin{aligned} \tilde{G}_{\xi\xi}(\ell_G, m_G) &= \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \lambda_j^{2\psi} \Re(I_{\xi\xi}(\lambda_j)), \\ \tilde{B}_{uu}(\ell_G, m_G) &= \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} \left(\lambda_j^{2\psi} + c(\psi)^2 \lambda_m^{2\psi} - 2\lambda_j^\psi \lambda_m^\psi \cos(\pi\psi/2) c(\psi)\right) \Re(\mathbf{B}' \mathbf{I}_{uu}(\lambda_j) \mathbf{B}), \end{aligned}$$

Moreover, since Assumptions D1-D3 and M together with Theorem 2 and Corollary 1 on Hannan (1970, pp. 248-249) imply $\mathbb{E}[\mathbf{I}_{u\xi}(\lambda_j)] = O(n^{-1})$ and $\mathbb{V}[\mathbf{I}_{u\xi}(\lambda_j)] = O(n^{-1})$, for $j = 1, \dots, m$, we have,

$$\begin{aligned} \widehat{G}_{\tilde{\eta}\tilde{\eta}}^{(b,1)}(\ell_G, m_G) - \tilde{G}_{\xi\xi}(\ell_G, m_G) - \tilde{B}_{uu}(\ell_G, m_G) &\leq O_p\left(\lambda_m^{2\psi} m^{-1/2}\right) + O_p\left(\lambda_m^{2\psi} n^{-1/2}\right) \\ &\quad + \frac{K\lambda_m^\psi}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{n^{1/2}}{j} \left(\lambda_m^\psi + \frac{\ln(n)}{m_d^{1/2}}\right)\right) + \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\frac{n}{j^2} \left(\lambda_m^{2\psi} + \frac{\ln(n)^2}{m_d} + \frac{\ln(n)\lambda_m^\psi}{m_d^{1/2}}\right)\right) \end{aligned}$$

$$\begin{aligned}
&\leq O_p\left(\lambda_m^{2\psi} m^{-1/2}\right) + O_p\left(\frac{n^{1/2}}{m_G^{1-\epsilon} \ell_G^{1+\epsilon}} \left(\frac{m_G}{n}\right)^\psi \left(\left(\frac{m_G}{n}\right)^\psi + \frac{\ln(n)}{m_d^{1/2}}\right)\right) \\
&+ O_p\left(\frac{n}{m_G \ell_G^2} \left(\left(\frac{m_G}{n}\right)^{2\psi} + \frac{\ln(n)^2}{m_d} + \frac{\ln(n)}{m_d^{1/2}} \left(\frac{m_G}{n}\right)^\psi\right)\right)
\end{aligned} \tag{B.79}$$

similarly to (B.30), for some arbitrarily small $\epsilon > 0$, using, again, $0 < \underline{d}_x \leq 1$, $\psi \geq 0$, Varneskov (2017, Lemma C.4) and that the remaining cross-product terms of the errors are of strictly lower order by the tuning parameters satisfying $\ell_G \asymp n^{\nu_G}$, $m_G \asymp n^{\kappa_G}$ and $m_d \asymp n^\varrho$, with $0 < \nu_G < \kappa_G < \varrho \leq 1$ in Assumptions F- d_1 and T- d_1 . Hence, by invoking the latter, we have

$$\lambda_m^{-2\psi} \left(\widehat{G}_{\tilde{\eta}\tilde{\eta}}^{(b,1)}(\ell_G, m_G) - \tilde{G}_{\xi\xi}(\ell_G, m_G) - \tilde{B}_{uu}(\ell_G, m_G) \right) \leq o_p(1). \tag{B.80}$$

Finally, by the definition $c(\psi) = \cos(\pi\psi/2)/(1+\psi)$ and the same arguments for Lemma B.2(a),

$$\lambda_m^{-2\psi} \left(\tilde{G}_{\xi\xi}(\ell_G, m_G) - G_{\psi\psi}/(1+2\psi) \right) = o_p(1), \tag{B.81}$$

$$\lambda_m^{-2\psi} \left(\tilde{B}_{uu}(\ell_G, m_G) - (1/(1+2\psi) - c(\psi)^2) \mathbf{B}' \mathbf{G}_{uu} \mathbf{B} \right) = o_p(1). \tag{B.82}$$

This delivers the requisite convergence result for model (iii), and equivalent arguments establish the corresponding results for models (ii) and (iv), which appear as special cases of the limit.

For (c), since we have by the Cauchy-Schwarz inequality,

$$\left| \widehat{G}_{\tilde{\eta}\tilde{\epsilon}}^{(b,1,2)}(\ell_G, m_G) \right| \leq \sqrt{\widehat{G}_{\tilde{\eta}\tilde{\eta}}^{(b,1)}(\ell_G, m_G) \widehat{G}_{\tilde{\epsilon}\tilde{\epsilon}}^{(2)}(\ell_G, m_G)}, \tag{B.83}$$

the convergence results follow by invoking (a) and (b), concluding the proof. \square

B.3 Proof of Theorem 1

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

$$\sqrt{m} \lambda_m^{-b} \left(\widehat{\mathbf{B}}_c(\ell, m) - \widehat{\mathbf{B}}(\ell, m) \right) \leq O_p \left((m/n)^{\underline{d}_x - b} \sqrt{m}/\ell^{1+\epsilon} \right), \tag{B.84}$$

for some arbitrarily small $\epsilon > 0$. Hence, we may continue by working with the corresponding estimate without regressor endogeneity, $\widehat{\mathbf{v}}_t$. As for the proof of Theorem 6, define $\widehat{\mathbf{A}}(L) \equiv \widehat{\mathbf{D}}(L) \mathbf{D}(L)^{-1}$ and $\mathbf{a}_t \equiv \mathbf{D}(L) \mathbf{z}_t$ such that $\widehat{\mathbf{v}}_t = \widehat{\mathbf{A}}(L) \mathbf{a}_t$, and further write $\mathbf{a}_t = \boldsymbol{\mu}_t + \mathbf{v}_t$, where $\boldsymbol{\mu}_t \equiv \mathbf{D}(L) \boldsymbol{\mu} \mathbf{1}_{\{t \geq 1\}}$ and, again, $\mathbf{v}_t = (e_t, \mathbf{u}'_{t-1})'$ with $e_t = \varphi_{t-1} + \eta_t^{(d_1)}$, $d_1 = b$, and $\varphi_{t-1} = \mathbf{B}' \mathbf{u}_{t-1} + \xi_{t-1}$. 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,

$$\widehat{\mathbf{B}}(\ell, m) - \mathbf{B} = \widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1} \widehat{\mathbf{F}}_{u\eta}^{(b)}(1, m) + \widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1} \widehat{\mathbf{F}}_{u\xi}(1, m) - \mathbf{C}_1 + \mathbf{C}_2 + \mathbf{C}_3^{(b)} + \mathbf{C}_4, \tag{B.85}$$

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

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

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

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

$$\sqrt{m} \lambda_m^{-b} (\mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3^{(b)}) = o_p(1), \quad \lambda_m^{-1} \widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m) \xrightarrow{\mathbb{P}} \mathbf{G}_{uu}. \quad (\text{B.86})$$

The result, then, follows by applying AVOA (2020, Lemma A.3) to $\sqrt{m} \lambda_m^{-1-b} \widehat{\mathbf{F}}_{u\eta}^{(b)}(1, m)$ in conjunction with (B.86), the continuous mapping theorem and Slutsky's theorem.

Inference for models (ii) and (iii): Since $\widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1} \widehat{\mathbf{F}}_{u\eta}^{(b)}(1, m) = O_p(\lambda_m^b m^{-1/2})$, with $0 < b = d_1 \leq 1$, and we may use AVOA (2020, Lemma A.2) to show $\sqrt{m} \mathcal{C}_4 = o_p(1)$, despite ξ_{t-1} being non-trivial, the central limit theory follows by applying AVOA (2020, Lemma A.3) to $\sqrt{m} \lambda_m^{-1} \widehat{\mathbf{F}}_{u\xi}(1, m)$ in conjunction with (B.86), 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. \square

B.4 Proof of Theorem 2

The result follows using Lemmas B.1(c) and (e) to eliminate the impact of endogeneity, Lemma B.2(c) to establish the requisite convergence results for $\widehat{\mathbf{G}}_{\widehat{\eta}\widehat{\eta}}^{(b)}(\ell_G, m_G)$, AVOA (2020, Lemma A.4(a)) to establish an equivalent convergence result for $\widehat{\mathbf{G}}_{\widehat{u}\widehat{u}}(\ell_G, m_G)$ and the continuous mapping theorem. \square

B.5 Proof of Theorem 3

The result follows by Lemma B.8 in conjunction with the Cramér-Wold theorem. \square .

B.6 Proof of Theorem 4

First, define $\mathbf{B}_m = \lambda_m^\psi c(\psi) \mathbf{B}$ and the function $g(x_1, x_2) = x_1/x_2$, for $x_2 \neq 0$. Second, note that we can write $\mathcal{R}(\mathcal{L}) = g(\mathcal{L}' \widehat{\mathbf{B}}_c(\ell, m, \widehat{\gamma}_x), \mathcal{L}' \widehat{\mathbf{B}}_c(\ell, \widehat{m}, \widehat{\gamma}_x))$. Hence, using Theorem 3 and $\mathcal{L}' \mathbf{B} \neq 0$, we may use the delta method to establish the asymptotic distribution of $\mathcal{R}(\mathcal{L})$ and subsequently the distribution of the estimator $\widehat{\psi}(\mathcal{L})$. We provide the explicit steps of the proof for models (ii)-(iii), since identical arguments deliver the corresponding result for model (iv). As an initial step, we compute the gradient

vector as,

$$g'(\mathcal{L}'\mathcal{B}_m, \mathcal{L}'\mathcal{B}_{\tilde{m}}) = \frac{\lambda_{\tilde{m}}^{-\psi}}{c(\psi)\mathcal{L}'\mathcal{B}} \begin{pmatrix} 1 \\ -\varkappa^\psi \end{pmatrix} \quad (\text{B.87})$$

implying that one component of the asymptotic variance becomes,

$$g'(\mathcal{L}'\mathcal{B}_m, \mathcal{L}'\mathcal{B}_{\tilde{m}})' \Phi(\varkappa, \psi) g'(\mathcal{L}'\mathcal{B}_m, \mathcal{L}'\mathcal{B}_{\tilde{m}}) = \lambda_{\tilde{m}}^{-2\psi} \Theta(\varkappa, \psi, \psi).$$

Hence, by invoking Theorem 3(a), we have

$$\sqrt{m}(\mathcal{R}(\mathcal{L}) - \varkappa^\psi) \xrightarrow{\mathbb{D}} N(0, \varkappa^{2\psi} \mathcal{S}(\mathcal{L}, \xi, \psi) \Theta(\varkappa, \psi, \psi)), \quad (\text{B.88})$$

when $\psi \neq 1/4$. Next, we define another function $h(x) = \ln(x)/\ln(\varkappa)$, with $x > 0$, whose gradient is given by $h'(\varkappa^\psi) = \varkappa^{-\psi}/\ln(\varkappa)$. Hence, by another application of the delta method, it follows,

$$\sqrt{m}(\widehat{\psi}(\mathcal{L}) - \psi) \xrightarrow{\mathbb{D}} N(0, \mathcal{S}(\mathcal{L}, \xi, \psi) \Theta(\varkappa, \psi, \psi) / \ln(\varkappa)^2), \quad (\text{B.89})$$

thus providing the distribution result for models (ii)-(iii), concluding the proof. \square

B.7 Proof of Theorem 5

The result follows using Lemmas B.1(c) and B.4(b) to eliminate the impact of endogeneity, Lemma B.9(c) to establish the requisite convergence results for $\widehat{\mathbf{G}}_{\tilde{\eta}\tilde{\eta}}^{(b)}(\ell_G, m_G)$, AVOA (2020, Lemma A.4(a)) to establish an equivalent convergence result for $\widehat{\mathbf{G}}_{\tilde{u}\tilde{u}}(\ell_G, m_G)$ and the continuous mapping theorem. \square

B.8 Proof of Theorem 6

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

$$\widehat{\mathbf{G}}_{\widehat{\mathbf{v}}\widehat{\mathbf{v}}}^c(\ell_G, m_G) - \widehat{\mathbf{G}}_{\widehat{\mathbf{v}}\widehat{\mathbf{v}}}(\ell_G, m_G) \leq O_p\left((m_G/n)^{d_x}/\ell_G^{1+\epsilon}\right), \quad (\text{B.90})$$

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

$$\begin{aligned} \widehat{\mathbf{G}}_{\widehat{\mathbf{v}}\widehat{\mathbf{v}}}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(1, m_G) &= \left(\widehat{\mathbf{G}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(1, m_G) \right) + \left(\widehat{\mathbf{G}}_{\mathbf{v}\mathbf{v}}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\ell_G, m_G) \right) \\ &\quad + \left(\widehat{\mathbf{G}}_{\widehat{\mathbf{v}}\widehat{\mathbf{v}}}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\mathbf{v}\mathbf{v}}(\ell_G, m_G) \right) \equiv \mathbf{u}_1^{(G)} + \mathbf{u}_2^{(G)} + \mathbf{u}_3^{(G)}, \end{aligned} \quad (\text{B.91})$$

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

Lemma B.10. *Under the conditions of Theorem 6, the following uniform bounds hold:*

(a) $m_G^{1/2} \mathbf{u}_1^{(G)} = O_p(\ell_G / \sqrt{m_G})$.

(b) $m_G^{1/2} \mathbf{u}_2^{(G)} \leq O_p((\ell_G / \sqrt{m_G})(\ell_G / n)^{d_1}) + O_p((m_G / n)^{d_1} / \sqrt{m_G})$.

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

$$m_G^{1/2} \mathbf{u}_3^{(G)} \leq O_p\left(\frac{n}{\sqrt{m_G} \ell_G^2} \left(\frac{m_G}{n}\right)^{\frac{2d_x}{2}}\right) + O_p\left(\frac{n^{1/2}}{\sqrt{m_G}} \left(\frac{m_G}{n}\right)^{\frac{d}{2}} \frac{m_G^\epsilon}{\ell_G^{1+\epsilon}}\right) + O_p\left(\ln(n)(m_G / m_d)^{1/2}\right).$$

(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

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

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

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

$$\mathbf{u}_2^{(G)} = \widehat{\mathbf{G}}_{\check{\mathbf{c}}\check{\mathbf{c}}}(\ell_G, m_G) + \widehat{\mathbf{G}}_{\psi\check{\mathbf{c}}}(\ell_G, m_G) + \widehat{\mathbf{G}}_{\check{\mathbf{c}}\psi}(\ell_G, m_G) \quad (\text{B.92})$$

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 $j \asymp n^j$, $\forall j > 0$. Hence, uniformly,

$$\widehat{\mathbf{G}}_{\check{\mathbf{c}}\check{\mathbf{c}}}(\ell_G, m_G) = \frac{1}{m_G - \ell_G + 1} \sum_{j=\ell_G}^{m_G} O_p(\lambda_j^{2d_1}) \leq \left(\frac{m_G}{n}\right)^{2d_1} \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p\left(\left(\frac{\ell_G}{m_G}\right)^{2d_1}\right). \quad (\text{B.93})$$

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

$$\frac{1}{m_G} \sum_{j=\ell_G}^{m_G} \left(\frac{\ell_G}{m_G}\right)^{2d_1} = \int_{\frac{\ell_G}{m_G}}^1 x^{2d_1} dx + O(m_G^{-1}) = O\left((\ell_G / m_G)^{1+2d_1}\right) + O(m_G^{-1}), \quad (\text{B.94})$$

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

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

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

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

$$\mathbf{u}_{31}^{(G)} = \hat{\mathbf{G}}_{\mu\mu}(\ell_G, m_G) + \left(\hat{\mathbf{G}}_{\psi\mu}(\ell_G, m_G) + \hat{\mathbf{G}}_{\mu\psi}(\ell_G, m_G) \right) + \left(\hat{\mathbf{G}}_{\check{\mathbf{c}}\mu}(\ell_G, m_G) + \hat{\mathbf{G}}_{\mu\check{\mathbf{c}}}(\ell_G, m_G) \right),$$

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

$$\begin{aligned} \mathbf{u}_{311}^{(G)} &\equiv \hat{\mathbf{G}}_{\mu\mu}(\ell_G, m_G) + \left(\hat{\mathbf{G}}_{\psi\mu}(\ell_G, m_G) + \hat{\mathbf{G}}_{\mu\psi}(\ell_G, m_G) \right) \\ &\leq O_p \left(\frac{n}{m_G \ell_G^2} \left(\frac{m_G}{n} \right)^{2d_x} \right) + O_p \left(\frac{n^{1/2}}{m_G} \left(\frac{m_G}{n} \right)^{\frac{d_x}{2}} \frac{m_G^\epsilon}{\ell_G^{1+\epsilon}} \right), \end{aligned} \quad (\text{B.95})$$

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),

$$\hat{\mathbf{G}}_{\mu\check{\mathbf{c}}}(\ell_G, m_G) \leq \frac{K}{m_G} \sum_{j=\ell_G}^{m_G} O_p \left(n^{1/2} j^{-1} \right) \times O_p \left(\lambda_j^{d_1} \right) \leq \frac{K m_G^\epsilon n^{1/2}}{m_G} \left(\frac{m_G}{n} \right)^{d_1} \sum_{j=\ell_G}^{m_G} O_p(j^{-1-\epsilon}), \quad (\text{B.96})$$

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

$$\mathbf{u}_{31}^{(G)} \leq O_p \left(\frac{n}{m_G \ell_G^2} \left(\frac{m_G}{n} \right)^{2d_x} \right) + O_p \left(\frac{n^{1/2}}{m_G} \left(\frac{m_G}{n} \right)^{\frac{d}{2}} \frac{m_G^\epsilon}{\ell_G^{1+\epsilon}} \right). \quad (\text{B.97})$$

Next, for the second term $\mathbf{u}_{32}^{(G)}$, we may write,

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

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

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

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

$$\bar{f}_G(\ell_G, m_G, n) = 1 \vee \frac{n^{1/2} m_G^\epsilon}{m_G \ell_G^{1+\epsilon}} \vee \frac{n}{m_G \ell_G^2}, \quad \text{with} \quad \bar{f}_G(\ell_G, m_G, n) \rightarrow 1, \quad (\text{B.98})$$

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

$$\widehat{\mathbf{G}}_{\tilde{v}\tilde{v}}^{(1)}(\ell_G, m_G) - \widehat{\mathbf{G}}_{bb}(\ell_G, m_G) \leq O_p\left(\frac{\ln(n)^2}{m_d} \bar{f}_G(\ell_G, m_G, n)\right) + O_p\left(\frac{\ln(n)}{\sqrt{m_d}} \sqrt{\bar{f}_G(\ell_G, m_G, n)}\right), \quad (\text{B.99})$$

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

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

for some arbitrarily small $\epsilon > 0$, using (B.94) and the same arguments as for (B.95) and (B.96). Hence, by Assumptions T and T-G * , $\widehat{\mathbf{G}}_{\tilde{v}\tilde{v}}^{(2)}(\ell_G, m_G) - \widehat{\mathbf{G}}_{\tilde{c}\tilde{c}}(\ell_G, m_G) \leq o_p(\mathbf{U}_{311}^{(G)})$. Similarly,

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

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

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

B.9 Proof of Theorem 7

The result follows by combining Theorem 6 and Robinson & Yajima (2002, Theorem 4). \square