Consistent Local Spectrum Inference for Predictive Return Regressions

Andersen, Torben G. ; Varneskov, Rasmus T.

Document Version
Accepted author manuscript

Published in:
Econometric Theory

DOI:
10.1017/S0266466622000354

Publication date:
2022

License
Unspecified

Citation for published version (APA):
Andersen, T. G., & Varneskov, R. T. (2022). Consistent Local Spectrum Inference for Predictive Return Regressions. *Econometric Theory*, *38*(6), 1253-1307. https://doi.org/10.1017/S0266466622000354

Consistent Local Spectrum Inference for Predictive Return Regressions*

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

June 20, 2022

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, an empirical application to monthly S&P 500 return predictions provides evidence for a fractionally integrated conditional mean component. Our new LCM procedure and tools indicate significant predictive power for future returns stemming from key state variables such as the default spread and treasury interest rates.

Keywords: Cointegration, Fractional Integration, Frequency Domain Inference, Local Spectrum Procedure, Return Predictability, Semiparametric Estimation, VAR Models.

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

*We wish to thank the Co-Editor Guido Kuersteiner, two anonymous referees as well as seminar 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).

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

‡Department of Finance, Copenhagen Business School, Frederiksberg, Denmark; CREATES, Aarhus, Denmark; the Danish Finance Institute; Multi Assets at Nordea Asset Management, Copenhagen, Denmark; e-mail: 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 and Ludvigson (2010) and Campbell (2018, Chapter 5). Similarly, equilibrium asset pricing models such as the long-run risk model (Bansal and Yaron, 2004), dynamic disaster model (Gabaix, 2012) or regime-switching CCAPM (Lettau, Ludvigson and 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 and 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 and 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,

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

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

[1] These assumptions 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 $\widehat{\mathcal{B}}_{\text{OLS}} = \sum_{t=1}^{n} y_t x_{t-1} / \sum_{t=1}^{n} x_{t-1}^2$. The null hypothesis of no predictability implies $\mathcal{B} = 0$, 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,

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

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

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

Whether this endogeneity correction ensures satisfactory inference hinges on the quality of the asymptotic approximation to the distribution for the regression coefficient, $\widehat{\mathcal{B}}_{\text{OLS}}$. In this regard, the strong persistence of many candidate regressors points towards a potential "spurious regression" problem, although the absence of strong return correlation may alleviate this concern. Still, under the alternative hypothesis, $\mathcal{B} \neq 0$, the mean return inherits the persistence of the (true) regressor, even if it likely will be disguised by the large return innovations. The theoretical justification for predictability implies we should pay close attention to this scenario. Indeed, through extensive simulations under carefully calibrated, strictly stationary, alternatives, Ferson, Sarkissian and 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.

[2] The 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.

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

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

The presence of a highly persistent mean return has implications beyond the need to adapt the finite-sample inference accordingly. On the one hand, it improves our ability to identify the true predictive relationship, as the signal-to-noise is enhanced, when we examine the "correct" regressor. On the other hand, the concern about misleading inference is exacerbated by the high correlation among many candidate regressors. If one is found significant, a number of others are also likely to display predictive ability. This implies that a significant regressor 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 and 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 and Whitelaw (2008), and more recently Kostakis, Magdalinos and Stamatogiannis (2015), find that no significant gains are obtained through this approach.

1.2 The Local-to-Unit Root Approach

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

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

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

The LUR approximation to the asymptotic distribution in the near unit root scenario for the predictor has two important implications. First, the rate of convergence of $\widehat{\mathcal{B}}_{\text{OLS}}$ increases to n, reflecting the enhanced signal-to-noise ratio associated with unit root-style regressions. Second, inference generally becomes non-standard. Specifically, if $\sigma_{vw} \neq 0$, the interaction between the persistent regressor and the lagged return residual generates a random endogeneity bias that depends on C_ϕ. Under the LUR specification, the deviation of the autoregressive root ϕ_n from unity shrinks at the same speed as the rate of convergence, rendering consistent inference for this coefficient infeasible. This implies that C_ϕ is an unidentified nuisance parameter, and the asymptotic distribution for $\mathcal{B}$ has a discontinuity around unity, relative to the stationary case ($\phi_n = \phi < 1$), complicating inference in the absence of prior knowledge about the underlying strength of the regressor persistence.

Various techniques have been developed to handle the 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 and 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 Phillips and Magdalinos (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

[5]For another procedure to obtain near optimal tests in the univariate setting, see Elliott, Müller and 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, an LUR representation, moderate integration ($C_\phi > 0$, $0 < \delta_\phi < 1$), and stationarity ($C_\phi > 0$, $\delta_\phi = 0$). Specifically, in this setting, the IVX procedure obtains valid inference by generating an instrument for $x_t = \sum_{s=1}^{t} \Delta x_s$ directly from the series itself through a filter that ensures a mild reduction in the degree of persistence,

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

When β_z is chosen below, but near, unity, $\widetilde{z}_t$ is at most mildly integrated, and its dynamics are governed exclusively through deliberate choices of C_z and β_z, which may, thus, be designed to generate a desirable limit distribution.[6] The IVX estimator is, then, simply the standard IV estimator, with $\widetilde{z}_t$ serving as instrument, $\widehat{\mathcal{B}}_{\text{IVX}} = \sum_{t=1}^{n} y_t \widetilde{z}_{t-1} / \sum_{t=1}^{n} x_{t-1} \widetilde{z}_{t-1}$. In the unit root and LUR scenarios, the estimation error for OLS, $\sum_{t=1}^{n} \upsilon_t\, x_{t-1} / \sum_{t=1}^{n} x_{t-1}^2$, has an 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 and Magdalinos (2007). Specifically, letting the errors obey an mds, then, under suitable regularity conditions, $n^{(1+\beta_z)/2}(\widehat{\mathcal{B}}_{\text{IVX}} - \mathcal{B}) \xrightarrow{\mathbb{D}} MN(0, \sigma^2_{\text{IVX}})$. The asymptotic variance, σ^2_{IVX}, is generally stochastic if the IVX instrument is moderately integrated, but a feasible, consistent estimator may be obtained 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 nonstationary 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 valu-

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

ation ratios, macroeconomic variables, lagged returns, and realized volatility measures, it is a very strong requirement. Phillips and 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 and 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 is 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 and 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 and Phillips (1989) and Sims, Stock and 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 and Demetrescu (2015) compare the size and power properties of IVX and related variable addition techniques in an LUR setting; Ren, Tu and 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

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

[8] The point is illustrated in Hamilton (1994, Chapter 18) for scenarios subject to potential spurious regression issues in a unit root setting, while Choi (1993) explores inference in AR systems with $I(1)$ processes. These procedures are studied more broadly for inference in possibly (co-)integrated VAR systems by, e.g., Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996). Moreover, Bauer and 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.

χ^2-distributed Wald tests. Likewise, Liu, Yang, Cai and 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 and 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 and Stambaugh (2009)) cointegration. Finally, Georgiev, Harvey, Leybourne and Taylor (2019) develop 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 and Varneskov (2021*a*), we develop a different asymptotic framework for analyzing predictive regressions within persistent systems. Specifically, we assume that all variables are fractionally integrated of potentially different orders, and that the regression may, or may not, feature cointegration. Let L and $(1-L)^d$ be the usual lag and fractional differencing operators, then, drawing parallels to the predictive systems (1)-(4), we stipulate a predictive relation of the form,

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

where $u_{t-1} \in I(0)$ is weakly dependent, and $\upsilon_t \in I(d_y - b)$ with $0 \leq b \leq d_y$ captures the possibility of cointegration (when $b > 0$).[9] The differencing operator $(1-L)^{d_x - d_y}$ ensures that the regression is balanced, while $b > 0$ indicates that the order of integration in the error term is reduced relative to d_y, indicating (fractional) cointegration. In this formulation, we allow $y_t \in I(d_y)$ and $x_{t-1} \in I(d_x)$ to exhibit either weak or strong dependence by having their fractional integration orders fall within

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

a wide range $0 \leq d < 2$, for $d = \{d_y, d_x\}$. Importantly, the framework is not confined to univariate predictive regressions (with trivial means or initial values), but accommodates diverse persistence (i.e., d's) among the predictors, thus providing a flexible setting to analyze systems with various financial and macroeconomic variables. This feature corresponds to having different localization coefficients in the LUR setting (3).[10] Andersen and 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".[11]

In this paper, we extend the framework in Andersen and Varneskov (2021*a*) to allow for imperfect regressors (in the spirit of Pastor and Stambaugh (2009)) that may exhibit general forms of endogeneity, which is similar to treating an omitted regressor problem, with the latter allowed to be persistent. That is, we expand the LCM approach to handle empirically relevant scenarios, where the regressors may be imperfect, persistent and endogenous, for which there is currently no uniform solution available in the literature. However, the LCM procedure still relies critically on consistent estimation of the fractional integration orders of the variables. This is particularly difficult for return regressions, because the signal-to-noise ratio of the conditional mean return to its innovations typically is too "low" for standard univariate time series techniques to detect (strong) serial dependence in finite samples. We begin to address this issue by developing: (i) a new feasible inference procedure, which holds irrespective of whether the regressors are "imperfect" or "perfect"; (ii) 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 significance test for (a subset of) the regressors, which avoids estimation of d_y altogether. These contributions provide substantial extensions to the LCM approach in Andersen and Varneskov (2021*a*) and allow us to examine return regressions under various levels of generality. Specifically, if we can safely assume that a subset of the variables is significant (e.g., due to compelling asset pricing theory), we may use LCMB to estimate d_y and test for return predictability of the remaining variables. If this assumption is deemed too strong, we can adopt the testing procedure in (iii) which, while less powerful, remains asymptotically valid.

We establish the theoretical properties of LCM and each of the new components (i)-(iii) in an endogenous, imperfect, and persistent regressor setting, demonstrating that the asymptotic distribution theory is Gaussian, regardless of the inference scenario, stationary versus nonstationary persistence, and perfect versus imperfect predictors. Moreover, we examine the finite sample properties of pre-

[10] In fact, Duffy and Kasparis (2021) show there is a close link between certain nonstationary fractionally integrated processes with d close to $1/2$ and autoregressive processes of the LUR type in a regression context.

[11] These issues are not treated in the companion paper Andersen and Varneskov (2021*b*) either, which considers testing for parameter instability and structural breaks in persistent predictive relations (that is, testing on $\mathcal{B}$), adopting the setting of Andersen and Varneskov (2021*a*). The latter, in particular, rules out the simultaneous presence of cointegration and endogeneity as well as imperfect regressors which, as will be detailed below, have important implications for the asymptotic properties of LCM. Hence, all asymptotic results in this paper are new.

dictability tests using OLS, IVX and LCM procedures. Specifically, OLS and IVX may suffer from considerable size distortions in our long memory setting, thus providing "spurious" inference and non-trivial biases in coefficient estimates. In comparison, our LCM procedures can be designed to display desirable size properties, non-trivial power, and bias robustness in general settings.

Finally, in an empirical application to monthly S&P 500 return prediction, we find evidence corroborating the presence of a fractionally integrated conditional mean return component. Furthermore, by applying an LCMB-augmented procedure, we find that key state variables, such as the default spread and treasury rates, do possess significant predictive power for the future returns.

The paper proceeds as follows. Section 2 introduces the setting, draws parallels to the imperfect regressor model of Pastor and Stambaugh (2009) and describes the LCM procedure. Section 3 provides limit theory and feasible inference, conditional on being able to estimate the return persistence. Section 4 introduces our LCMB approach to persistence estimation and 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 assumptions, technical lemmas and proofs of the main theoretical results. Proofs of the technical lemmas are deferred to an Online Appendix.

2 Predictive Returns Regressions with Persistent Variables

This section introduces a predictive regression framework for asset returns, where all variables may exhibit fractional integration of different orders. It is inspired by the persistent economic systems studied in Andersen and Varneskov (2021*a*) and the predictive system for expected returns with imperfect predictors developed by Pastor and Stambaugh (2009). Finally, we motivate and review the Local speCtruM (LCM) approach, introduced by the former.

2.1 Predictive System and Assumptions

The predictive regression model consists of several components, which we introduce and describe sequentially. Before proceeding, we formally define the fractional differencing operator,

$$(1-L)^d \;=\; \sum_{i=0}^{\infty} \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}\, L^i, \tag{6}$$

for some integration order d, where $\Gamma(\,\cdot\,)$ is the gamma function. We assume to observe a $(k+1)\times 1$ vector $\boldsymbol{Z}_t = (y_t, \boldsymbol{\mathcal{X}}_{t-1}')'$ at times $t = 1,\ldots,n$, where y_t denotes the asset return and $\boldsymbol{\mathcal{X}}_{t-1}$ is a vector of candidate predictors, stipulated to have a multi-component structure,

$$\boldsymbol{\mathcal{X}}_{t-1} = \boldsymbol{x}_{t-1} + \boldsymbol{c}_{t-1}, \qquad \boldsymbol{x}_t \perp\!\!\!\perp \boldsymbol{c}_s, \quad \text{for all } t,s, \tag{7}$$

with $\boldsymbol{x}_{t-1}$ capturing the most persistent signal, and the mean zero process $\boldsymbol{c}_{t-1} \in I(0)$ consisting of either measurement errors, additional weakly dependent components in the variables, or both.

Moreover, we define $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_{t-1}')'$, which is assumed to obey a Type II fractional model,

$$\boldsymbol{D}(L)(\boldsymbol{z}_t - \boldsymbol{\mu}) \;=\; \boldsymbol{v}_t \, \mathbf{1}_{\{t \geq 1\}}, \quad \text{with} \quad \boldsymbol{v}_t \;=\; (e_t, \boldsymbol{u}_{t-1}')' \tag{8}$$

being a weakly dependent process, and $\boldsymbol{\mu} = (\mu_y, \boldsymbol{\mu}_x')'$ is a $(k+1) \times 1$ vector of nonrandom, unknown finite numbers, capturing either the means or initial values of $\boldsymbol{z}_t$.[12] That is, we allow y_t and the persistent signal of the regressors, $\boldsymbol{x}_{t-1}$, to display fractional integration with persistence described by,

$$\boldsymbol{D}(L) = \text{diag}\left[(1-L)^{d_1}, \boldsymbol{D}_x(L)\right], \quad \boldsymbol{D}_x(L) = \text{diag}\left[(1-L)^{d_2}, \ldots, (1-L)^{d_{k+1}}\right], \tag{9}$$

while the observed regressors may exhibit even more flexible dynamics due to the inclusion of $\boldsymbol{c}_{t-1}$.

In this setting, where all variables may possess a high degree of persistence, we need a regression model that is compatible with the observed dynamics in equations (7)–(8). This is achieved by defining the predictive relation between y_t and the observable regressors $\boldsymbol{\mathcal{X}}_{t-1}$ through the weakly dependent components of the unobservable persistent signals $\boldsymbol{x}_{t-1}$. Specifically, we assume,

$$e_t = \varphi_{t-1} + \eta_t^{(b)}, \quad \varphi_{t-1} = \boldsymbol{\mathcal{B}}'\boldsymbol{u}_{t-1} + \xi_{t-1}, \quad \boldsymbol{u}_t \perp\!\!\!\perp \xi_s, \quad \text{for all } t, s, \tag{10}$$

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 (8) and (10), this is tantamount to a balanced prediction model for asset returns,

$$y_t = a + \boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{Q}}(L)\boldsymbol{x}_{t-1} + \xi_{t-1}^{(-d_1)} + \upsilon_t, \quad t = 1, \ldots, n\,, \tag{11}$$

where $\boldsymbol{\mathcal{Q}}(L) = \boldsymbol{D}_x(L)(1-L)^{-d_1}$, $a = \mu_y - \boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{Q}}(L)\boldsymbol{\mu}_x$ as well as the innovations,

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

This implies that we may summarize the observable system, $\boldsymbol{Z}_t$, in vector form,

$$\begin{pmatrix} y_t \\ \boldsymbol{\mathcal{X}}_{t-1} \end{pmatrix} = \begin{pmatrix} a + \boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{Q}}(L)\boldsymbol{x}_{t-1} + \xi_{t-1}^{(-d_1)} + \upsilon_t \\ \boldsymbol{x}_{t-1} + \boldsymbol{c}_{t-1} \end{pmatrix} = \begin{pmatrix} \mu_y + \boldsymbol{\mathcal{B}}'(1-L)^{-d_1}\boldsymbol{u}_{t-1} + \xi_{t-1}^{(-d_1)} + \upsilon_t \\ \boldsymbol{\mu}_x + \boldsymbol{D}_x(L)^{-1}\boldsymbol{u}_{t-1} + \boldsymbol{c}_{t-1} \end{pmatrix}, \tag{12}$$

where the main objective is to estimate and draw inference on the parameter vector $\boldsymbol{\mathcal{B}}$. Unfortunately, as discussed in detail below, the regression model in equation (11) is latent, with three nonstandard layers disguising the predictive relation. First, we observe $\boldsymbol{\mathcal{X}}_{t-1}$, not the persistent signals $\boldsymbol{x}_{t-1}$ nor the measurement errors $\boldsymbol{c}_{t-1}$. Second, the error ξ_{t-1} cloaks the conditional mean, similar to an omitted regressor. Third, the persistence of the variables, measured by $\boldsymbol{D}(L)$, is unknown a priori.

Despite these challenges, and assuming a latent predictive relation, it is important to realize that the system (7)-(12) encompasses most multivariate fractionally integrated systems in the literature. To see this, suppose $\boldsymbol{c}_{t-1} = \mathbf{0}$ and $\xi_{t-1} = 0$, $\forall t$, as well as $0 \leq b \leq d_1$, then the most persistent components

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

of the explanatory variables, captured by $\boldsymbol{x}_{t-1}$, 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 describes the inferential issues surrounding return regressions. In particular, $\boldsymbol{c}_{t-1}$ is included to accommodate endogeneity, multiple components and measurement errors in the regressors, rendering their signals latent, and φ_{t-1} captures the possibility that the predictors may imperfectly describe the conditional mean. Finally, letting,

$$b = d_1 \qquad \text{such that} \qquad \upsilon_t = \eta_t, \tag{13}$$

which we maintain throughout, the regression system (12) has only weakly dependent innovations. Yet, as noted previously, the latter may well dominate the persistent signal in finite samples, effectively disguising the serial correlation and predictability of the persistent mean component.[13]

Moreover, our regression system is balanced, regardless of the forecasting prowess of the regressors, as $y_t \in I(d_1)$ under both $\mathcal{H}_0 : \boldsymbol{\mathcal{B}} = \mathbf{0}$ and $\mathcal{H}_A : \boldsymbol{\mathcal{B}} \neq \mathbf{0}$. $\mathcal{H}_0$ allows for the regressors to 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, $\boldsymbol{x}_{t-1}$, to ensure regression balance. This allows d_1 to differ from the persistence of the candidate predictors, $\{d_2, \ldots, d_{k+1}\}$. Note, however, that under $\mathcal{H}_A$, implicitly, the predictive relation is between y_t and a "persistence transformed" signal, $\widetilde{\boldsymbol{x}}_{t-1} \equiv \boldsymbol{\mathcal{Q}}(L)\boldsymbol{x}_{t-1}$ rather than $\boldsymbol{x}_{t-1}$. Of course, if $\boldsymbol{\mathcal{Q}}(L) = \boldsymbol{I}_k$, a k-dimensional identity matrix, this adjustment is negligible.

The next section discusses these points, provides examples and draws parallels to the extant literature, particularly Pastor and Stambaugh (2009) and Andersen and Varneskov (2021*a*). Before proceeding, however, we impose some formal structure on the system. The regularity conditions mirror those imposed by Andersen and Varneskov (2021*a*) and the assumptions for the semiparametric fractional cointegration analyses in, e.g., Robinson and Marinucci (2003), Christensen and Nielsen (2006) and Christensen and 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 on the system in terms of $\boldsymbol{q}_t = (\boldsymbol{u}'_{t-1}, \eta_t)'$, the weakly dependent components of the persistent regressor signals and regression errors in equations (8) and (10), as well as $\boldsymbol{\zeta}_{t-1} = (\boldsymbol{c}'_{t-1}, \xi_{t-1})'$, the measurement errors in the observable regressors and the conditional mean; see equations (7), (10) and (12). Finally, we define the filtration $\mathcal{F}_t = \sigma(y_s, \boldsymbol{\mathcal{X}}_s, s \leq t)$.

Assumption D1. *The vector process $\boldsymbol{q}_t$, $t = 1, \ldots,$ is covariance stationary with spectral density matrix satisfying $\boldsymbol{f}_{qq}(\lambda) \sim \boldsymbol{G}_{qq}$ as $\lambda \to 0^+$, where $\boldsymbol{G}_{qq}$ is finite with non-random elements, the upper left $k \times k$ submatrix, $\boldsymbol{G}_{uu}$, 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 $|\boldsymbol{f}_{qq}(\lambda) - \boldsymbol{G}_{qq}| = O(\lambda^{\varpi})$ as $\lambda \to 0^+$. Finally, let $\boldsymbol{G}_{qq}(i, k+1)$ be the $(i, k+1)$th element of $\boldsymbol{G}_{qq}$, which has $\boldsymbol{G}_{qq}(i, k+1) = \boldsymbol{G}_{qq}(k+1, i) = 0$ for all $i = 1, \ldots, k$.*

[13] The presence of both d_1 and b in the regression errors of the system (7)-(12) signals our ability to accommodate fractional cointegration models that do not fully purge the residuals of persistence. Since this feature typically is not contemplated for predictive *return* regressions, we simplify notation by ignoring this possibility henceforth. Nonetheless, we note that equivalent results apply for $0 \leq b \leq d_1$, subject only to slight changes in the tuning parameter assumptions.

Assumption D2. *$\boldsymbol{q}_t$ is a linear vector process, $\boldsymbol{q}_t = \sum_{j=0}^{\infty} \boldsymbol{A}_j \boldsymbol{\epsilon}_{t-j}$, whose lower $k \times (k+1)$ submatrix of $\boldsymbol{A}_0$ contains zeros and coefficient matrices satisfy the condition $\sum_{j=0}^{\infty} j^{1/2} \|\boldsymbol{A}_j\| < \infty$. Moreover, the innovation sequence has, almost surely, $\mathbb{E}[\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\mathcal{F}_{t-1}] = \boldsymbol{I}_{k+1}$, and the higher-order 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. There exists a random variable ζ such that $\mathbb{E}[\zeta^2] < \infty$ and, for all c and some C, $\mathbb{P}[\|\boldsymbol{q}_t\| > c] \leq C\mathbb{P}[|\zeta| > c]$. Finally, the periodogram of $\boldsymbol{\epsilon}_t$ is denoted by $\boldsymbol{J}(\lambda)$.*

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

Assumption C. *Suppose $\boldsymbol{\zeta}_{t-1} = \boldsymbol{\zeta}_{t-1}\mathbf{1}_{\{t\geq1\}}$ is a mean-zero $(k+1)\times1$ vector satisfying conditions similar to those imposed by Assumption D1-D3 on $\boldsymbol{q}_t$. These assumptions, however, are formally stated in Appendix A for brevity. Moreover, the following, additional, conditions are required:*

(a) The co-spectrum between $\boldsymbol{\zeta}_{t-1}$ and η_t satisfies $\boldsymbol{f}_{\zeta\eta}(\lambda) \sim \boldsymbol{G}_{\zeta\eta}$, as $\lambda \to 0^+$, where the finite and non-random vector $\boldsymbol{G}_{\zeta\eta}$ may have non-zero entries.

(b) $\boldsymbol{u}_t \perp\!\!\!\perp \boldsymbol{\zeta}_s$ for all $t, s \geq 1$.

(c) When the i-th element of $\boldsymbol{\zeta}_{t-1}$ is trivial, i.e., when $\boldsymbol{\zeta}_{t-1}(i) = 0$ for all $t \geq 1$, then the co-spectrum condition $\boldsymbol{G}_{\zeta\eta}(i) = \boldsymbol{G}_{\zeta\zeta}(i, g) = \boldsymbol{G}_{\zeta\zeta}(g, i) = 0$ for $g = 1, \ldots, k+1$, naturally also holds.

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

Assumptions D1-D3 are standard in the literature on fractional (co-)integration. Specifically, D1 and D3 impose a rate of convergence for the spectral density $\boldsymbol{f}_{qq}(\lambda)$ as $\lambda \to 0^+$, which depends on the smoothness parameter $\varpi \in (0, 2]$. In addition, D1 requires locally full rank of $\boldsymbol{u}_{t-1}$ and it being locally exogenous to η_t as $\lambda \to 0^+$, but not global exogeneity. Finally, condition D2 specifies linearity, martingale and moment conditions for $\boldsymbol{q}_t$, allowing for general multivariate dependence among the variables, thus accommodating flexible lead-lag and predictive structures.[14]

Whereas D1 allows the latent predictive signals, $\boldsymbol{x}_{t-1}$, to exhibit mild endogeneity (as $\lambda \to c > 0$) through $\boldsymbol{u}_{t-1}$, Assumption C lets the observable explanatory variables exhibit stronger forms of endogeneity. This is captured via the co-spectrum between the less persistent component (and/or measurement errors) $\boldsymbol{c}_{t-1}$ and the innovations η_t, and these may, furthermore, both be 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 and Stambaugh (2009).

Assumption M imposes a mild structure on the memory of the system. Specifically, we restrict the persistent component of returns to exhibit at most unit root persistence, whereas the observable

[14] Note that Assumption D2 restricts elements of the lower $k \times (k+1)$ submatrix of $\boldsymbol{A}_0$ to be equal to zero since the weakly dependent predictor components, $\boldsymbol{u}_{t-1}$, 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 and Varneskov (2021*a*), which is necessary to derive the asymptotic properties for our new LCM-bias estimator in Section 4.

variables may be explosive, $d_i > 1$. However, as explained above, if the persistence of the regressors deviate from d_1, a transformation is required to ensure that the unobserved persistent signals $\widetilde{\boldsymbol{x}}_{t-1} \equiv \boldsymbol{Q}(L)\boldsymbol{x}_{t-1}$ comply with balance of the predictive relation (11). In general, our setting accommodates a very flexible persistence structure; if $0 < d_i < 1/2$, the variable is (asymptotically) stationary with long memory; if $d_i \geq 1/2$, the variable is nonstationary, 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 also display substantially different degrees of persistence. This is often the case for applications involving multiple financial and macroeconomic variables.

Finally, we impose $b = d_1 \leq \underline{d}_x$ as in (13), which, as noted, implies $\upsilon_t = \eta_t$ 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. Note that this is slightly more restrictive than the corresponding balanced cointegration requirement $0 \leq b \leq \min(d_1, \underline{d}_x)$ in Andersen and Varneskov (2021*a*, Eq. (8)), implying that the degree of fractional cointegration cannot be stronger than the persistence of the regressors. Both conditions can be relaxed at the expense of more complicated assumptions on the tuning parameters below.

Remark 1. *Assumption M stipulates that $\underline{d}_x > 0$, i.e., that all predictors have long memory. This condition is necessary, when the requisite elements of $\boldsymbol{c}_{t-1}$ are non-zero with positive probability. That is, we obtain identification of the persistent predictive signals through differences in memory relative to their weakly dependent components (and by using the LCM approach). We can accommodate cases, where $d_i = 0$, when $\boldsymbol{c}_{t-1}(i) = 0$, $\forall t \geq 1$, which is analogous to assuming exogeneity in OLS settings. Our assumption is reminiscent of the approach in Pastor and 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 and Yaron (2014) and Campbell, Giglio, Polk and Turley (2018), exhibit strong persistence and may be characterized as either stationary or nonstationary 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 $\boldsymbol{c}_{t-1} \in I(0)$ are of strictly lower order than $\boldsymbol{x}_{t-1}$, 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. Specifically, 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 also small and, since $\boldsymbol{f}_{\zeta\eta}(\lambda) \sim \boldsymbol{G}_{\zeta\eta}$ 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 (7)-(12) 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, $\boldsymbol{x}_{t-1}$, and the corresponding modified signal, $\widetilde{\boldsymbol{x}}_{t-1}$ ensures that the predictor signal is transformed to match the persistence of the (latent) conditional mean. To appreciate the mechanics of the filter $\boldsymbol{\mathcal{Q}}(L)$, 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 $\widetilde{x}_{t-1} = \mathcal{Q}(L)x_{t-1} = (1 - L)x_{t-1}$ with memory $d_{\widetilde{x}} = d_x - 1 = 0.8$. 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 $\widetilde{\mathcal{X}}_{t-1} = (1-L)\mathcal{X}_{t-1}$ via $\widetilde{x}_{t-1}$.[15] 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.[16] 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 $\boldsymbol{\mathcal{X}}_{t-1}$. In contrast, if the predictors are "perfect", we have $\varphi_{t-1} = \boldsymbol{\mathcal{B}}'\boldsymbol{u}_{t-1}$.

Third, the system accommodates endogenous regressors through $\boldsymbol{c}_{t-1}$, which is independent of the persistent signal, $\boldsymbol{x}_{t-1}$. To motivate this feature, we draw an analogy to the long-run risk model of Bansal and 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 and 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 about the conditional equity premium, but they facilitate richer system dynamics.[17]

[15] More generally, our setting accommodates a different degree of fractional persistence among the signals $\boldsymbol{x}_{t-1}$, necessitating different adjustments through $\boldsymbol{\mathcal{Q}}(L)$. The common feature is that they all must match the persistence of the conditional mean, i.e., have a post-transformation fractional integration order of d_1. Despite Assumption M excluding the possibility, our framework can also accommodate cases with $d_x < d_1$, whose persistence must be "enhanced" for compatibility. For example, if $d_x = 0.2$ and $d_1 = 0.6$, the signal must be aggregated by weights reflecting the fractional memory differential $d_1 - d_x$. However, such cases demand more complex assumptions on the tuning parameters below and are ruled out for ease of exposition.

[16] Specifically, ξ_{t-1} is said to be non-trivial if the variable is non-zero for some t with strictly positive probability. Similarly, we say that the variable ξ_{t-1} is trivial if $\xi_{t-1} = 0$ for all $t \geq 1$. This nomenclature also applies to other variables.

[17] A multi-component structure of the conditional mean of consumption growth is consistent with the dynamic decomposition in, e.g., Ortu, Tamoni and 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 and Benzoni (2009) and many references therein.

Fourth, the model facilitates non-negligible correlation between the unspanned component of the conditional mean, ξ_{t-1}, and the observable explanatory variables (again, through $\boldsymbol{c}_{t-1}$) as well as with the innovations to asset returns, η_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 and Goyal (2008), Lettau and 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 and Stambaugh (2009), despite arising in our fractionally integrated setting rather than their first-order AR economy. The following remark elaborates on the similarities in these approaches.

Remark 3. *Pastor and 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,*

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

where $0 < \phi < 1$*, the eigenvalues of* $\boldsymbol{\mathcal{A}}$ *are inside the unit circle, and the innovation vector* $(\eta_t, w_t, \boldsymbol{u}_t')'$ *is i.i.d. Gaussian. The system features return predictability via the conditional mean (since* $\phi > 0$*), endogenous regressors, and it accommodates imperfect predictors, when* $\varphi_{t-1} \neq a_\varphi + \boldsymbol{\mathcal{B}}'\boldsymbol{\mathcal{X}}_{t-1}$*. Moreover, if the predictors are imperfect, this generates unspanned return persistence, as captured in our setup via the inclusion of* $\xi_{t-1}^{(-d_1)}$ *in equation* (11)*. Finally, their key identifying assumption for* $\boldsymbol{\mathcal{B}}$ *is* $0 < \phi < 1$*, allowing the persistent conditional mean to be disentangled from the noise. If this assumption fails, they require exogenous regressors. It is analogous to assuming* $\underline{d}_x > 0$ *in Assumption M.*

The model (7)-(12) features four competing hypotheses for the return dynamics:

(i) $\boldsymbol{\mathcal{B}} = \mathbf{0}$ and $\xi_{t-1} = 0$, $\forall t = 1, \ldots, n$; returns are not predictable by other variables.

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

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

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

Hypotheses (i) and (ii) imply that $\boldsymbol{\mathcal{X}}_{t-1}$ 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.[18] 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 particularly 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 four hypotheses imply different inference regimes for persistent variables, for which standard OLS delivers spurious inference; see, e.g., Granger and Newbold (1974), Phillips (1987), and Tsay and Chung (2000). Of course, if we knew that y_t and $\boldsymbol{\mathcal{X}}_{t-1} = \boldsymbol{x}_{t-1}$ 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 and Marinucci (2003), Robinson and Hualde (2003), Christensen and Nielsen (2006) and Johansen and Nielsen (2012). Generally, however, we do not know, a priori, which of the hypotheses captures the given scenario, i.e., whether the regressors are endogenous and/or the predictors are "perfect", and we must estimate the persistence of $\boldsymbol{z}_t$, which is impeded 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 and Varneskov (2021*a*), we allow for "imperfect" predictors and the joint presence of endogeneity and cointegration.[19] Hence, the theoretical analysis and 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 the decomposition of the spectral density for the observable regressors, $\boldsymbol{\mathcal{X}}_{t-1}$, and their co-spectrum with the asset returns, y_t. For ease of exposition, let us suppose here that all series are asymptotically stationary with $d_i < 1/2$, $i = 1, \ldots, k+1$, and utilize $\boldsymbol{f}_{xc}(\lambda) \sim \boldsymbol{0}$, as $\lambda \to 0^+$, then we may write[20],

$$\boldsymbol{f}_{\mathcal{X}\mathcal{X}}(\lambda) \simeq \boldsymbol{\Lambda}_{xx}(\lambda)^{-1}\, \boldsymbol{G}_{uu}\, \overline{\boldsymbol{\Lambda}}_{xx}(\lambda)^{-1} + \boldsymbol{G}_{cc}\,, \tag{14}$$

$$\boldsymbol{f}_{\mathcal{X}y}(\lambda) \simeq \boldsymbol{\Lambda}_{xx}(\lambda)^{-1}\, \boldsymbol{G}_{uu}\, \boldsymbol{\mathcal{B}}\, \overline{\Lambda}_{yy}(\lambda)^{-1} + \boldsymbol{f}_{x\xi}^{(-d_1)}(\lambda) + \boldsymbol{f}_{x\eta}(\lambda) + \boldsymbol{G}_{c\xi}\, \overline{\Lambda}_{yy}(\lambda)^{-1} + \boldsymbol{G}_{c\eta}\,, \tag{15}$$

[18] We classify predictability with respect to specific model components, $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{x}_{t-1}$, not the information filtration $\mathcal{F}_t = \sigma(y_s, \boldsymbol{\mathcal{X}}_s, s \leq t)$. In principle, lagged y_t values may be informative about an omitted regressor, ξ_{t-1}. However, empirically, the return autocorrelation is minimal and, thus, the signal-to-noise ratio for mean returns is very low, rendering lagged returns ineffective for forecasting purposes, and of little practical use in financial applications. Hence, we focus on the predictive relation between returns and persistent regressors.

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

[20] We invoke asymptotic stationarity here to avoid additional approximation errors in the decompositions (14)-(15) under nonstationary integration, see Phillips and Shimotsu (2004), Shimotsu and Phillips (2005) and Robinson (2005). Of course, we allow for nonstationarity and deal with such errors in developing the asymptotic theory for LCM below.

for λ near zero, where $\overline{\Lambda}_{yy}(\lambda)$ and $\overline{\mathbf{\Lambda}}_{xx}(\lambda)$ are the complex conjugates of $\Lambda_{yy}(\lambda)$ and $\mathbf{\Lambda}_{xx}(\lambda)$,

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

These decompositions are intuitive. First, $\boldsymbol{f}_{\mathcal{X}\mathcal{X}}(\lambda)$ shares the multi-component structure of the observable regressors $\boldsymbol{\mathcal{X}}_{t-1}$, with the spectral density of the persistent signal dominating the frequencies in the vicinity of the origin. However, the speed of divergence may differ across elements, depending on the fractional integration orders of the regressors. Second, $\boldsymbol{f}_{\mathcal{X}y}(\lambda)$ contains information about the forecasting prowess of the regressors through $\boldsymbol{\mathcal{B}}$, and this effect, captured in the first term, asymptotically dominates the remaining ones at lower frequency ordinates, i.e., as $\lambda \to 0^+$. Hence, if the cross-spectrum is appropriately scaled to account for the divergence rates induced by $\mathbf{\Lambda}_{xx}(\lambda)^{-1}$ and $\overline{\Lambda}_{yy}(\lambda)^{-1}$, it converges to a limit given by $\boldsymbol{G}_{uu}\boldsymbol{\mathcal{B}}$. Third, $\boldsymbol{G}_{c\xi}\,\overline{\Lambda}_{yy}(\lambda)^{-1}$ and $\boldsymbol{G}_{c\eta}$ capture endogeneity-induced bias terms. They stem from the potentially non-zero co-spectra between the regressor measurement errors, $\boldsymbol{c}_{t-1}$, and the processes ξ_{t-1} and η_t, respectively, capturing 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 $\lambda \to 0^+$, but at a slower rate than the first term, since $\underline{d}_x > 0$. Finally, the co-spectra $\boldsymbol{f}^{(-d_1)}_{x\xi}(\lambda)$ and $\boldsymbol{f}_{x\eta}(\lambda)$ introduce sampling errors for estimators of $\boldsymbol{\mathcal{B}}$, with their respective asymptotic orders differing due to $\xi^{(-d_1)}_{t-1} \in I(d_1)$ and $\eta_t \in I(0)$.

In general, the (co-)spectral densities in equations (14) and (15) 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, $\boldsymbol{u}_{t-1}$ and e_t, are,

$$\boldsymbol{f}_{uu}(\lambda) \simeq \boldsymbol{G}_{uu} \qquad \text{and} \qquad \boldsymbol{f}_{ue}(\lambda) \simeq \boldsymbol{G}_{uu}\boldsymbol{\mathcal{B}} + \boldsymbol{f}_{u\xi}(\lambda) + \boldsymbol{f}^{(d_1)}_{u\eta}(\lambda), \tag{16}$$

which both are asymptotically bounded, yet convey similar information about $\boldsymbol{\mathcal{B}}$. This suggests that inference based on the stationary components, $\boldsymbol{u}_{t-1}$ and e_t may circumvent issues regarding balance, degeneracy of point estimates and spurious inference, motivating Andersen and Varneskov (2021*a*) to introduce the LCM procedure, consisting of two main steps. First, the procedure *fractionally filters* the observed variables $\boldsymbol{Z}_t = (y_t, \boldsymbol{\mathcal{X}}'_{t-1})'$ to obtain an estimate of $\boldsymbol{v}_t = (e_t, \boldsymbol{u}'_{t-1})'$. Second, it uses medium band least squares (MBLS) estimation for robust inference. These steps are detailed next, along with certain 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, $\widehat{d}_i$ for $i = 1, \ldots, k+1$, that satisfies mild consistency requirements.

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

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

Assumption F is mild, requiring only the existence of an estimator which, under suitable assumptions on equation (8), is consistent with an appropriate convergence rate. However, since we accommodate both (asymptotically) stationary and nonstationary variables in Assumption M, the estimator must apply for a wide range of d_i. The conditions on m_d are slightly strengthened below to ensure that fractional filtering has no asymptotic impact on the MBLS inference. Examples of estimators that satisfy our assumptions include the semi-parametric exact local Whittle (ELW), see Shimotsu and Phillips (2005) and Shimotsu (2010), the trimmed ELW (TELW) by Andersen and Varneskov (2021*a*), and parametric (long) fractional ARIMA(p, d, q) models using information criteria to determine the short-memory dynamics; see, e.g., Hualde and Robinson (2011) and Nielsen (2015).

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

$$\widehat{\boldsymbol{v}}_t^c \equiv (\,\widehat{e}_t, (\widehat{\boldsymbol{u}}_{t-1}^c)'\,)' = \widehat{\boldsymbol{D}}(L)\,\boldsymbol{Z}_t, \tag{17}$$

where $\widehat{\boldsymbol{u}}_{t-1}^c = \widehat{\boldsymbol{u}}_{t-1} + \widehat{\boldsymbol{c}}_{t-1}$, with $\widehat{\boldsymbol{u}}_{t-1} = \widehat{\boldsymbol{D}}_x(L)\,\boldsymbol{x}_{t-1}$ and $\widehat{\boldsymbol{c}}_{t-1} = \widehat{\boldsymbol{D}}_x(L)\,\boldsymbol{c}_{t-1}$.[21] Similarly, we define $\widehat{\boldsymbol{v}}_t \equiv (\widehat{e}_t, \widehat{\boldsymbol{u}}_{t-1}')'$, which is the equivalent, albeit unobservable, estimate of $\boldsymbol{v}_t$, without an endogenous component in the regressors. Using frequency domain techniques, we may then extract asymptotically similar information from $\widehat{\boldsymbol{v}}_t$ and $\widehat{\boldsymbol{v}}_t^c$. 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, $\widehat{\boldsymbol{D}}(L)\,\boldsymbol{\mu}\,\mathbf{1}_{\{t\geq 1\}}$, in a unified manner during second stage estimation.

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

$$\boldsymbol{w}_h(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \boldsymbol{h}_t\, e^{\mathrm{i}t\lambda_j}, \quad \overline{\boldsymbol{w}}_k(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \boldsymbol{k}_t\, e^{-\mathrm{i}t\lambda_j}, \quad \boldsymbol{I}_{hk}(\lambda_j) = \boldsymbol{w}_h(\lambda_j)\,\overline{\boldsymbol{w}}_k(\lambda_j), \tag{18}$$

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 $\Re(\boldsymbol{I}_{hk}(\lambda_j))$, as,

$$\widehat{\boldsymbol{F}}_{hk}(\ell, m) = \frac{2\pi}{n} \sum_{j=\ell}^{m} \Re(\boldsymbol{I}_{hk}(\lambda_j)), \qquad 1 \leq \ell \leq m \leq n, \tag{19}$$

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

$$\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) = \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}^c(\ell, m)^{-1}\,\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}^c(\ell, m), \tag{20}$$

[21] Consistent with the Type II nature of the fractional model in (8), the operator $\widehat{\boldsymbol{D}}(L)$ uses all available observations.

for which $\ell, m \to \infty$ and $\ell/m + m/n \to 0$, as $n \to \infty$. The MBLS estimator has some distinct advantages for predictive inference and testing with persistent variables. Specifically, combining sample-size-dependent trimming with a bandwidth $m/n \to 0$, equation (20) is first-order equivalent to,

$$\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) \;=\; \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1}\, \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{e}}(\ell, m), \tag{21}$$

that is, the corresponding estimator based on $\widehat{\boldsymbol{v}}_t$. 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 (14) and (15), are dominated by information about $\boldsymbol{\mathcal{B}}$, whereas 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 the 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 suggests 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 under the model hypotheses (ii)-(iv), conditional on Assumption F being satisfied, thus tackling the issues surrounding imperfect regressors.[22] Subsequently, in Section 4, we provide a new LCM-bias (LCMB) approach to estimating d_1, that relies on regression slopes, and develop a significance test for $\boldsymbol{\mathcal{B}}$, which circumvents the need to estimate d_1.

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 and Varneskov (2021*a*) by allowing for "imperfect" predictors and the simultaneous presence of endogeneity and cointegration. Moreover, we introduce a new feasible inference procedure that works across the competing hypothesis (ii)-(iv), without prior knowledge of which scenario applies to the inference problem.

[22] For ease of presentation, we focus on the three model hypotheses (ii)-(iv) in our exposition of the asymptotic theory. Moreover, as noted previously, substantial empirical and theoretical literatures suggest that the null hypothesis in (ii) is more appropriate than (i). However, as discussed in Remark 4 below, similar arguments can be applied to establish asymptotic theory for model (i). Hence, when appropriate, we discuss corresponding results for this case.

3.1 Central Limit Theory

Our development of the LCM inference procedure focuses on the scenarios (ii)-(iv), where the main distinction is between the imperfect regressor case, models (ii)-(iii), versus the perfect regressor case, model (iv). We let "$\asymp$" signify that the ratio of two terms converges to a positive constant and may 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 $n \to \infty$,*

$$\frac{m^{1+2\varpi}}{n^{2\varpi}} + \frac{\ell^{1+\varpi+d_1}}{n^{\varpi}\, m^{1/2+d_1}} + \frac{n^{1/2+d_1}}{m_d^{1/2}\, m^{d_1}\, \ell} + \frac{n^{1-d_1}}{m^{1/2-d_1}\, \ell^2} + \frac{n^{d_1}}{m^{1/2+d_1}} + \frac{m^{1/2+\underline{d}_x-d_1}}{n^{\underline{d}_x-d_1}\, \ell} + \frac{n^{1/2}}{m} \to 0.$$

The restrictions in Assumption T are mild. The first term is standard for semiparametric estimation in the frequency domain, see, e.g., Robinson (1995) and Lobato (1999), while the remaining conditions are specific to the second-stage MBLS estimator, adopted in the LCM procedure. Specifically, condition two, implying $\nu < (\varpi + \kappa(1/2 + d_1))/(1 + \varpi + d_1)$, restricts the information loss from frequency trimming; three, $(1 - \varrho)/2 + d_1(1 - \kappa) < \nu$ in conjunction with $0 < \nu < \kappa < \varrho \leq 1$, eliminates errors from estimating the integration orders; four, $(1 - \kappa/2 - d_1(1 - \kappa))/2 < \nu$, alleviates the low-frequency bias from mean-slippage following fractional filtering; five, $d_1/(1/2 + d_1) < \kappa$ imposes a mild bound on the bandwidth; six, $\kappa/2 - (1 - \kappa)(\underline{d}_x - d_1) < \nu$ eliminates the endogeneity bias. Finally, condition seven requires $1/2 < \kappa$ to avoid further cross-restrictions on the tuning parameters.[23]

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 $(3/5 - d_1/5)/2 < \nu < 4/5$. The lower bound is strictly decreasing in $0 \leq d_1 \leq 1$, implying that its most restrictive scenario is for $d_1 = 0$, 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 $1/10 + d_1/5 \leq 3/10 < \nu$. Finally, if the regressors are endogenous and we select κ close to 4/5 to enhance the efficiency of the MBLS estimator, we require $2/5 - (\underline{d}_x - d_1)/5 < \nu$, with the most conservative bound obtained when $\underline{d}_x - d_1 = 0$. 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.

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

Theorem 1. *Suppose Assumptions D1-D3, C, M, F and T hold as well as* $0 < d_1 \leq 1$*, which rules out model hypothesis (i), and that the condition* $\max(0, (1 - 3\kappa/2)/(1 + \kappa/2)) < \varpi \leq 2$ *applies, then,*

$$\begin{cases} \sqrt{m} \left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \boldsymbol{\mathcal{B}} \right) \xrightarrow{\mathbb{D}} N\big(\mathbf{0}, \boldsymbol{G}_{uu}^{-1} G_{\xi\xi}/2\big), & \text{under models (ii) and (iii)}, \\ \sqrt{m}\, \lambda_m^{-d_1} \left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \boldsymbol{\mathcal{B}} \right) \xrightarrow{\mathbb{D}} N\big(\mathbf{0}, \boldsymbol{G}_{uu}^{-1} G_{\eta\eta}/(2(1 + 2d_1))\big), & \text{under model (iv)}. \end{cases}$$

When consistent estimation of d_1 is feasible, Theorem 1 demonstrates that the LCM procedure is asymptotically Gaussian for the predictive models with imperfect regressors, (ii) and (iii), and for 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 and 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 $\sqrt{m}\lambda_m^{-d_1} \asymp \sqrt{m}(n/m)^{d_1}$ and the asymptotic variance is scaled by $1/(2(1+2d_1))$. 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), the limiting distribution remains Gaussian, regardless of whether the variables are (asymptotically) stationary or nonstationary, whether there is cointegration, and irrespective of the cointegration being weak ($d_1 < 1/2$) or strong ($d_1 \geq 1/2$). This is unique within a fractional cointegration context. Similar properties do not hold for OLS, narrow band least squares (NBLS), or maximum likelihood in fractionally cointegrated VAR models, where inference exhibits varying forms of non-Gaussianity in nonstationary cases; see Robinson and Marinucci (2003), Christensen and Nielsen (2006) and Johansen and Nielsen (2012).[24] Likewise, the Gaussian limit theory for the MBLS estimator without fractional filtering in Christensen and 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 and Phillips (2005).

Moreover, the limiting theory of the LCM procedure is correctly centered, and thus free from biases induced by persistent and endogenous regressors, as detailed by Stambaugh (1999), Pastor and Stambaugh (2009) and Phillips and 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 describing the

[24]Such methods generally do not accommodate non-trivial means, or initial values, as well as strong endogeneity among the regressors, that may or may not be "perfect". Moreover, the limit theories for these alternatives rely on the presence of cointegration. Finally, as demonstrated by Andersen and Varneskov (2021*a*, 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.

system. We introduce such a feasible procedure in the following section.

Remark 4. *We impose $0 < d_1 \leq 1$ in Theorem 1 to rule out model (i), but our results are readily extendable to the case $d_1 = 0$. Careful inspection of the proofs shows that this only requires 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 $G_{\xi\xi} + G_{\eta\eta}$ since ξ_{t-1} and η_t are of the same asymptotic order. Moreover, since cointegration no longer features in model (iv), given $d_1 = 0$, the result is obtained by setting $d_1 = 0$ in the limit theory, implying a lower convergence rate. The results for model (i) will mirror model (iv) in this case, with, obviously, $\boldsymbol{\mathcal{B}} = \mathbf{0}$. Similar comments apply to some subsequent results.*

3.2 Feasible Inference

This section provides a feasible inference procedure for the scenarios (ii)-(iv), without knowledge of which of these scenarios actually describes the fractional system. As shown in Theorem 1, we obtain a different asymptotic limit theory depending on whether or not the system features cointegration in model (iv). Hence, feasible inference requires consistent estimators of the long-run covariance matrix $\boldsymbol{G}_{uu}$ and either $G_{\xi\xi}$ for models (ii)-(iii) or $G_{\eta\eta}/(1+2d_1)$ for model (iv). Similarly, we must account for the different rates of convergence for the MBLS estimator in the distinct inference regimes. The main challenges are, first, that we observe $\widehat{\boldsymbol{v}}_t^c$ instead of either $\widehat{\boldsymbol{v}}_t$ or $\boldsymbol{v}_t$. Second, the residual components ξ_{t-1} and $\eta_t^{(d_1)}$ are latent, and we estimate them, following the estimation of $\boldsymbol{\mathcal{B}}$, as,

$$\widehat{\eta}_t^{(d_1,c)} \;=\; \widehat{e}_t \;-\; \widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m)'\,\widehat{\boldsymbol{u}}_{t-1}^c. \tag{22}$$

Importantly, and unlike the corresponding procedure in Andersen and Varneskov (2021*a*), we do not rely on separate fractional filtering of the estimate $\widehat{\eta}_t^{(d_1,c)}$ to ensure asymptotic $I(0)$ behavior, but rather utilize the different asymptotic orders in the two sources of error.[25] This is important, as it allows us to remain agnostic about what inference regime describes the system. Specifically, if the regressors are "imperfect," as in models (ii)-(iii), the contribution from ξ_{t-1} will asymptotically dominate that from $\eta_t^{(d_1)}$, since $d_1 > 0$, and we may, thus, use $\widehat{\eta}_t^{(d_1,c)}$ to estimate $G_{\xi\xi}$. In contrast, for model (iv), we recover information about the long-run variance of $\eta_t^{(d_1)}$, since $\xi_{t-1} = 0$, $\forall t \geq 1$.

Once the error process $\widehat{\eta}_t^{(d_1,c)}$ is estimated, we use a class of trimmed long-run covariance estimators to obtain the asymptotic variance components. In particular, for a generic vector $\boldsymbol{h}_t$, we adopt,

$$\widehat{\boldsymbol{G}}_{hh}(\ell_G, m_G) \;=\; \frac{1}{m_G - \ell_G + 1}\sum_{j=\ell_G}^{m_G} \Re\left(\boldsymbol{I}_{hh}(\lambda_j)\right), \tag{23}$$

with bandwidth and trimming functions, $m_G = m_G(n)$ and $\ell_G = \ell_G(n)$. This class of estimators is

[25] Moreover, Andersen and Varneskov (2021*a*) implement a separate estimator for the cointegration strength parameter b, which may differ from d_1 in their setting. Despite this, the approach in this paper will also apply to their setting. Note that the superscript (d_1, c), describing the residual estimate in (22), indicates that the error process may asymptotically depend on the integration order d_1, despite the fractional filtering having been carried out prior to MBLS estimation, and that the process is impacted by endogeneity via $\widehat{\boldsymbol{u}}_{t-1}^c$. Similar comments apply to related definitions below.

used for inference and testing in Andersen and Varneskov (2021a), but with the regression errors $\widehat{\eta}_t^{(d_1,c)}$ defined differently. Instead, it is similar to the procedure in Christensen and Varneskov (2017) in this respect. If we restrict $\ell_G = 1$, the estimator also resembles those employed by Robinson and Yajima (2002) and Nielsen and Shimotsu (2007) to design semiparametric tests for fractional cointegration rank in LW and ELW settings, respectively. However, we face additional challenges due to the, possibly, endogenous regressors, fractional filtering induced mean-slippage, and estimation errors as well as the lower-order filtering error $\eta_t^{(d_1)}$, 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 (23) with either the fractionally filtered regressors, $\widehat{\boldsymbol{u}}_{t-1}^c$, or the estimated residuals, $\widehat{\eta}_t^{(d_1,c)}$. These estimates, denoted by $\widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}^c(\ell_G, m_G)$ and $\widehat{G}_{\widehat{\eta}\widehat{\eta}}^{(d_1,c)}(\ell_G, m_G)$, are, then, combined to form,

$$\widehat{\mathbf{AVAR}}(\ell_G, m_G) \;=\; \widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}^c(\ell_G, m_G)^{-1}\,\widehat{G}_{\widehat{\eta}\widehat{\eta}}^{(d_1,c)}(\ell_G, m_G)/(2m_G)\,, \tag{24}$$

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 $0 < \nu_G < \kappa_G < \varrho \le 1$, and define $\underline{m}_n = m_G \wedge m$. Then, the following cross-restrictions are imposed,*

$$\frac{n}{m_G\,\ell_G^2} + \frac{n^2}{m_G\,\ell_G^2\,\underline{m}_n} + \left(\frac{n}{m_G}\right)^{d_1}\frac{1}{\sqrt{m_d}} + \frac{n^{1/2}}{m_G} \;\to\; 0, \quad \text{as } n \to \infty.$$

Whereas the first two regularity conditions mirror those imposed by Andersen and Varneskov (2021a, Assumption T-G), the last two conditions are unique to our asymptotic variance estimator. While the last condition allows us to avoid further cross-restrictions on the tuning parameters, similarly to Assumption T, the third is necessary for our feasible inference procedure in model (iv), as we must estimate the long-run variance $G_{\eta\eta}/(1+2d_1)$ and simultaneously recover information about the scale factor $\lambda_m^{2d_1}$, arising from the faster rate of convergence for LCM in this scenario.

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

$$\begin{cases} m_G\,\widehat{\mathbf{AVAR}}(\ell_G, m_G) \;\overset{\mathbb{P}}{\to}\; \boldsymbol{G}_{uu}^{-1}\,G_{\xi\xi}/2\,, & \textit{under models (ii) and (iii)}\,, \\ m_G\,\lambda_{m_G}^{-2d_1}\,\widehat{\mathbf{AVAR}}(\ell_G, m_G) \;\overset{\mathbb{P}}{\to}\; \boldsymbol{G}_{uu}^{-1}\,G_{\eta\eta}/(2(1+2d_1)), & \textit{under model (iv).} \end{cases}$$

Importantly, the estimator converges to the correct asymptotic variance in both inference regimes, with the exception of being scaled by m_G instead of the bandwidth m. Hence, by invoking $m_G = m$, 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. *Imposing $m_G = m$ and assuming the conditions of Theorems 1 and 2 hold, then,*

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

In summary, our feasible inference procedure is valid across scenarios (ii)-(iv), without prior knowledge of which describes the system, as long as Assumption F holds and we can estimate d_1 consistently.

4 An LCM-Bias Approach to Estimating d_1

This section introduces a new LCMB approach to estimation of the fractional integration order 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 $\widehat{d_1} \simeq 0$. 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 first "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, yet still 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 $0 < d_i \leq 1$, for $i \in \mathcal{S}_x$. Moreover, define $\gamma_x = \max_{i \in \mathcal{S}_x}(d_i)$ and suppose that $b = d_1$, $0 < d_1 < 1$ and $d_1 \leq \gamma_x$.*

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

The additional restrictions in Assumption M-d_1 are rather innocuous; they are satisfied when a "sensible" set of regressors has 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 $\widehat{\gamma}_x$ as the initial "guess" for the return persistence but, importantly, do not require that $\gamma_x = d_1$. Instead, we define,

$$\psi = \gamma_x - d_1, \quad \text{such that} \quad 0 \leq \psi \leq \gamma_x \,. \tag{25}$$

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 $0 < d_1 < 1$, is critical for obtaining the Gaussian central limit theory. We conjecture the condition $d_1 \leq \gamma_x$ may be relaxed, but this may impact the limit theory, as discussed in the following remark.

Remark 5. *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 and Nielsen (2006) and Christensen and Varneskov (2017), we conjecture that a Gaussian limit theory can be obtained, if our initial persistence guess $\widehat{\gamma}_x$ implies $\psi > -1/2$. Similarly, the study of NBLS in nonstationary environments by, e.g., Robinson and 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.*

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 $\widehat{\gamma}_x$ as an initial benchmark value for the return persistence. To this end, we define a constant $\varkappa > 1$, such that $\widetilde{m} = m/\varkappa > \ell$. The bandwidth $\widetilde{m}$ diverges at the same rate as m, but letting $\varkappa > 1$ enables us to asymptotically identify ψ. We further define,

$$\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x) \;=\; \widehat{\boldsymbol{F}}^c_{\widehat{u}\widehat{u}}(\ell, m)^{-1}\, \widehat{\boldsymbol{F}}^c_{\widehat{u}\check{e}}(\ell, m), \quad \text{where} \quad \check{e}_t \;=\; (1-L)^{\widehat{\gamma}_x}\, y_t \,, \tag{26}$$

as the LCM estimator implemented using $\widehat{\gamma}_x$ as the value for d_1, and equivalently write $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, \widetilde{m}, \widehat{\gamma}_x)$. 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*

$$\frac{m^{1+2(\varpi-2)}}{n^{2(\varpi-d_1)}} + \frac{\ell^{1+\varpi+\psi}}{m^{1/2+d_1+\psi} n^{\varpi-d_1}} + \frac{1}{\ell}\left(\left(\frac{n}{m}\right)^{1/2+d_1} + \left(\frac{n}{m}\right)^{1/2+\gamma_x-\underline{d}_x}\right)$$
$$+ \frac{m^{1/2}}{\ell^2}\left(\frac{n}{m}\right)\left(\left(\frac{m}{n}\right)^{\underline{d}_x-d_1} + \left(\frac{m}{n}\right)^{2\underline{d}_x-\gamma_x}\right) \;\to\; 0, \qquad \text{as} \quad n \to \infty\,.$$

The regularity conditions in Assumption T-d_1 tighten the corresponding restrictions one through four in Assumption T, especially if d_1 is large. To see this, note that condition one in Assumption T-d_1 and five in Assumption T imply $d_1/(1/2 + d_1) < \kappa < 2(\varpi - d_1)(1 + 2(\varpi - d_1))$, thus inducing a tight bound around the bandwidth selection $\kappa = 2/3$, if $\varpi = 2$ and d_1 is close to one. This explains why we require $d_1 < 1$, and why we impose $2d_1 < \varpi$ for Assumptions D1, T and T-d_1 to be consistent with Theorem 3 below. Hence, if d_1 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 restrictive. For example, if we set $\kappa = 0.70$, the condition implies $0.30(1/2 + d_1) < \nu$, which may become binding for d_1 close to one.

Finally, to characterize the LCM bias and variance in this setting, we write,

$$c(\psi) = \frac{\cos(\pi\psi/2)}{(1+\psi)}, \quad \boldsymbol{\mathcal{V}}(\eta,\gamma_x) = \boldsymbol{G}_{uu}^{-1}\frac{G_{\eta\eta}}{2(1+2\gamma_x)}, \quad \boldsymbol{\Phi}(\varkappa,\gamma_x) = \begin{pmatrix} 1 & \varkappa^{\gamma_x-1/2} \\ \varkappa^{\gamma_x-1/2} & \varkappa^{2\gamma_x-1} \end{pmatrix}, \tag{27}$$

with the matrices $\boldsymbol{\mathcal{V}}(\xi,\psi)$ and $\boldsymbol{\Phi}(\varkappa,\psi)$ defined analogously.

Theorem 3. *Suppose Assumptions D1-D3, C, M-d_1, F-d_1 and T-d_1 hold along with the implied condition* $\max(0,(1-3\kappa/2)/(1+\kappa/2),2d_1) < \varpi \leq 2$ *for Assumptions D1, T and T-d_1, then,*

(a) In the imperfect regressor models (ii) and (iii), it holds that,

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

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

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

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 $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m,\widehat{\gamma}_x)$ and $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,\widetilde{m},\widehat{\gamma}_x)$ depends on whether the system is stationary in the respective modeling regimes. For example, under model (iv), the asymptotic variance of $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m,\widehat{\gamma}_x)$ is smaller than the corresponding one for $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,\widetilde{m},\widehat{\gamma}_x)$, 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 nonstationary cointegration models, which is due to the signal being stronger at lower frequencies; see, e.g., Robinson and 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 $\boldsymbol{\mathcal{L}}$ be a $k\times 1$ vector, which is used to linearly transform $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x)$ and, thus, asymptotically $\lambda_m^{\psi}\, c(\psi)\, \boldsymbol{\mathcal{B}}$ to a scalar. Importantly, for this strategy to work, we require that $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$ to construct our ratio estimator for ψ.

Remark 6. *The assumption $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$ is not innocuous, imposing that not only do significant return predictors exist, we have also included them in $\boldsymbol{\mathcal{X}}_{t-1}$. It can be motivated, however, by the numerous asset pricing theories, which imply that the conditional mean return is driven by, e.g., valuation ratios, discount rates, volatility, default and tail risk, etc. Hence, if such variables are included among the regressors, it seems reasonable to assume that $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$ is, indeed, satisfied, with high probability.*

Now, if we assume that $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$ holds, our ratio estimator of ψ is defined as,

$$\widehat{\psi}(\mathcal{L}) \;=\; \frac{\ln \mathcal{R}(\mathcal{L})}{\ln(\varkappa)}, \quad \text{where} \quad \mathcal{R}(\mathcal{L}) \;=\; \frac{\boldsymbol{\mathcal{L}}'\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x)}{\boldsymbol{\mathcal{L}}'\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, \widetilde{m}, \widehat{\gamma}_x)}\,, \tag{28}$$

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

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

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

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

$$\begin{cases} \sqrt{m}\,\big(\widehat{\psi}(\mathcal{L}) - \psi\big) \;\overset{\mathbb{D}}{\to}\; N\Big(0,\, \frac{\mathcal{S}(\mathcal{L},\xi,\psi)\,\Theta(\varkappa,\psi,\psi)}{\ln(\varkappa)^2}\Big), & \textit{in models (ii)-(iii), when } \psi \neq 1/4\,, \\ \sqrt{m}\,\lambda_m^{-d_1}\,\Big(\widehat{\psi}(\mathcal{L}) - \psi\Big) \;\overset{\mathbb{D}}{\to}\; N\Big(0,\, \frac{\mathcal{S}(\mathcal{L},\eta,\gamma_x)\,\Theta(\varkappa,\psi,\gamma_x)}{\ln(\varkappa)^2}\Big), & \textit{in model (iv), when } \psi + \gamma_x \neq 1/2\,. \end{cases}$$

Theorem 4 shows that our LCMB estimator for $\psi = \gamma_x - d_1$ inherits desirable asymptotic properties from the LCM procedure, including the rate of convergence. Specifically, for the imperfect regressor models (ii)-(iii), we obtain the semiparametric rate $\widehat{\psi}(\mathcal{L}) - \psi = O_p\big(1/\sqrt{m}\big)$, when $\psi \neq 1/4$, and $\widehat{\psi}(\mathcal{L}) - \psi = o_p\big(1/\sqrt{m}\big)$, when $\psi = 1/4$. Moreover, for the cointegration model (iv), the rate is improved to $\widehat{\psi}(\mathcal{L}) - \psi = O_p\big(1/\sqrt{m}\lambda_m^{d_1}\big)$, when $\psi + \gamma_x \neq 1/2$, and $\widehat{\psi}(\mathcal{L}) - \psi = o_p\big(1/\sqrt{m}\lambda_m^{d_1}\big)$, if $\psi + \gamma_x = 1/2$, in analogy with super consistency. Since we have $\widehat{\gamma}_x - \gamma_x = O_p\big(1/\sqrt{m_d}\big)$ by Assumption F-d_1, Theorem 4 shows that $\widehat{d}_1(\psi) \equiv \widehat{\gamma}_x - \widehat{\psi}(\mathcal{L})$ obtains a conservative convergence rate of $\widehat{d}_1(\psi) - d_1 = O_p\big(1/\sqrt{m}\big)$. Critically, if our LCMB persistence estimator is implemented with a preliminary bandwidth $m_p \asymp n^{\kappa_p}$, where $\kappa < \kappa_p < \varrho \leq 1$, our original Assumption F is satisfied, the central limit theory in Theorem 1 holds, and we may rely on the feasible inference procedure from Corollary 1 for hypothesis testing on $\boldsymbol{\mathcal{B}}$. The validity of this procedure hinges on the assumption $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$. Hence, if we impose, backed by robust arguments, that this should apply for a subset of the regressors, we may utilize this to estimate d_1 using Theorem 4 prior to testing for significance of the remaining predictors via the LCM inference procedure. On the other hand, if the assumption $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$ is deemed too strong, we require a significance test for the null hypothesis in model (ii) that avoids estimation of d_1. We develop such a test in the next section by leveraging the limiting distribution in Theorem 3.

Remark 7. *The limiting distributions in Theorem 4 are pointwise and scenario dependent. As such, the small sample properties of the LCMB estimate* $\widehat{\psi}(\mathcal{L})$ *may be problematic, especially if* $\boldsymbol{\mathcal{L}}'\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x)$ *or* $\boldsymbol{\mathcal{L}}'\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, \widetilde{m}, \widehat{\gamma}_x)$ *are close to zero, when constructing the ratio* $\mathcal{R}(\mathcal{L})$*. Hence, we recommend imposing theory-inspired bounds* $0 \leq \widehat{\psi}(\mathcal{L}) \leq \widehat{\gamma}_x$ *to regularize the estimator in small samples.*

4.3 Testing Return Predictability without Estimating d_1

As an alternative to adopting Assumption F when implementing the LCM procedure, with or without the LCMB persistence estimator and its associated assumption $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$, we design a significance test for model hypothesis (ii) that avoids estimation of d_1. This is achieved by combining the LCM estimator based on our initial guess, $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x)$, its limiting distribution in Theorem 3 and our feasible inference procedure in Section 3.2. As before, we define,

$$\breve{\eta}_t^{(d_1,c)} \;=\; \breve{e}_t - \widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x)'\,\widehat{\boldsymbol{u}}_{t-1}^c, \quad \widehat{G}_{\breve{\eta}\breve{\eta}}^{(d_1,c)}(\ell_G, m_G) \;=\; \frac{1}{m_G - \ell_G + 1}\sum_{j=\ell_G}^{m_G} \Re\left(\boldsymbol{I}_{\breve{\eta}\breve{\eta}}^{(d_1,c)}(\lambda_j)\right), \tag{29}$$

as well as the asymptotic variance estimator,

$$\widehat{\mathbf{AVAR}}(\ell_G, m_G, \widehat{\gamma}_x) \;=\; \widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}^{c}(\ell_G, m_G)^{-1}\,\widehat{G}_{\breve{\eta}\breve{\eta}}^{(d_1,c)}(\ell_G, m_G)/(2m_G). \tag{30}$$

While the intuition and motivation behind the estimator (30) is similar to the corresponding one in equation (24), their asymptotic properties differ, as evidenced by the following theorem.

Assumption T-G-d_1. *Suppose Assumption T-G,* $m \asymp m_G$ *and* $(n/m_G)^{\psi}/\sqrt{m_G} \to 0$ *hold.*

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

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

The result for model (ii) in Theorem 5 is the same as in Theorem 2, but the remaining parts differ. Specifically, when $\boldsymbol{\mathcal{B}} \neq \mathbf{0}$ under the alternative hypothesis, $\mathcal{H}_A$, the inconsistency of the LCMB coefficient estimator $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x)$ spills over to the residual estimate, $\breve{\eta}_t^{(d_1,c)}$, with a rate inherited from the degenerate 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 the significance of $\boldsymbol{\mathcal{B}}$ and hypotheses of the form $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}}$. To see this, let $m_G = m$ and define,

$$\widehat{\mathfrak{T}}(\mathcal{L}) \;=\; \frac{\boldsymbol{\mathcal{L}}'\,\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x)}{\sqrt{\boldsymbol{\mathcal{L}}'\,\widehat{\mathbf{AVAR}}(\ell_G, m, \widehat{\gamma}_x)\,\boldsymbol{\mathcal{L}}}}\,, \qquad \text{where} \quad \widehat{\mathfrak{T}}(\mathcal{L}) \overset{\mathbb{D}}{\to} N(0,1), \tag{31}$$

under model (ii), by Theorems 3(a) and 5 as well as the continuous mapping theorem, Cramér-Wold and Slutsky theorems. Moreover, $\widehat{\mathfrak{T}}(\mathcal{L}) = O_p\big(\sqrt{m}\big)$ 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. Not surprisingly, this robustness is likely to come with a loss of finite sample power. Theorem 5 shows $\widehat{\mathbf{AVAR}}(\ell_G, m, \widehat{\gamma}_x)$ is inflated for model (iii), and the rate of divergence is strictly slower than the rate $\sqrt{m}\lambda_m^{-d_1}$, achieved by LCM in Theorem 1, for model (iv). Moreover, as noted above, the estimator $\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m, \widehat{\gamma}_x)$ is inconsistent, unless $\psi = 0$. Nonetheless, the test should prove useful in some scenarios, where we cannot reasonably argue that Assumption F holds for d_1 and, similarly, in settings, where we cannot invoke $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$ for the LCMB approach.

Remark 8. *An important direction for further research is to combine the test $\widehat{\mathfrak{T}}(\mathcal{L})$, the LCMB persistence estimator, and the LCM inference procedure to design a fully adaptive approach for analyzing return regressions that simultaneously delivers asymptotically correct inference and avoids the assumption $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$. However, this likely necessitates sequential testing, multi-step inference and hypothesis tests with nuisance parameters under the null. Hence, the requisite procedure must handle the complications arising from pre-testing, see, e.g., Leeb and Pötscher (2005). We are currently pursuing this task, but it requires additional theoretical developments beyond the scope of the present paper.*

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. Moreover, we assess the test size within an imperfect predictor specification. Finally, we study the finite-sample properties of 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 nonstationary fractional integration. To render the analysis manageable, we assume $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{\mathcal{X}}_{t-1}$ are univariate (written as $\mathcal{B}$ and $\mathcal{X}_{t-1}$, respectively) and stipulate that $\mathcal{X}_{t-1} = x_{t-1}$, which ensures the signal of the persistent regressor is directly observable and excludes endogeneity. Then, we generate fractional ARMA$(0,0)$ processes as,

$$(1-L)^{d_1}(y_t-\mu_y) \;=\; \varphi_{t-1}+\eta_t^{(d_1)}, \quad \varphi_{t-1} \;=\; \mathcal{B}u_{t-1}+\mathcal{B}_\xi\xi_{t-1}\,, \quad (1-L)^{d_2}(x_{t-1}-\mu_x) \;=\; u_{t-1}\,, \quad (32)$$

and $\eta_t^{(d_1)} = (1-L)^{d_1}\eta_t$, where $\zeta_t \sim$ i.i.d. $N(0,\sigma_\zeta^2)$ for $\zeta_t \in \{\eta_t, \xi_{t-1}, u_{t-1}\}$. Moreover, to highlight the impact of the noise-to-signal ratio for drawing inference about return persistence and predictability, we 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 three distinct long-memory scenarios,

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

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 nonstationary, yet mean-reverting, conditional mean process. In addition, we analyze an alternative DGP, where the conditional mean is considerable less persistent ($d_1 = 0$) than the observed regressor. This is included to assess the impact of "over-differencing," when wrongly imposing an initial integration order, $\widehat{\gamma}_x$, 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 $\widehat{d}_1(\psi) \simeq 0.7676$). As $\sigma_\eta \in [0, 25]$ varies, the noise in the predictive relation changes, impacting the quality of finite-sample inference regarding the persistence of the conditional mean return.[26]

Initially, we entertain univariate predictions using x_{t-1}, but fix $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$ in equation (32), implying that asset returns have a persistent mean, but the empiricist employs an irrelevant "imperfect" predictor, so that the persistence "spills over" into the residuals. In this scenario, we assess if and when σ_η is sufficiently large to induce "incorrect" inference regarding the fractional integration order of the returns, $\widehat{d}_1 \simeq 0$, as is generally found empirically. Moreover, we examine the size properties of predictability tests with, seemingly, $I(0)$ returns using either OLS, IVX or LCM inference. Furthermore, we consider four different implementations of the LCM procedure: One based on "standard" univariate time series estimates of d_1 as in Andersen and Varneskov (2021*a*); one "oracle" version where we treat (d_1, d_2) as known; one that adopts an initial guess $\widehat{\gamma}_x = \widehat{d}_2$ for d_1; and finally one where we construct the LCMB estimate $\widehat{d}_1(\psi)$ from an initial LCM regression using the "correct" predictor.[27] As discussed

[26]We have run similar experiments with ARMA$(1,0)$ short-run dynamics and/or $(d_1, d_2) = \{(0.00, 0.45), (0.45, 0.80)\}$. The corresponding results, when allowing for mild autoregressive dynamics as well as alternative long memory configurations, are qualitatively identical to those in Figures 1-5 and, thus, omitted for brevity.

[27]Specifically, we pretend to observe the conditional mean error process, $(1-L)^{d_2}\xi_{t-1}$, and initially use this "regressor" to generate an LCMB estimate for d_1, before implementing the LCM procedure for x_{t-1} with $\widehat{d}_1(\psi)$ and $\widehat{d}_2$ adopted in the fractional filtering step. As explained in Section 4.2, this corresponds to a setting, where we assume that $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$ holds for a subset of the predictors, and then use LCMB to test significance of the remaining predictors.

in Sections 4.2 and 4.3, the third implementation of LCM offers an asymptotically valid test of the null hypothesis, but delivers inconsistent point estimates of $\mathcal{B}$ under the alternative, when $\gamma_x \neq d_1$, while the last implementation of LCM relies on the auxiliary assumption $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$. In addition, we study the case $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$, corresponding to a "perfect" predictor scenario, to assess the bias of the LCM estimators and the power of the LCM-based significance tests.[28] 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 do not apply to fractionally (co)integrated systems, 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 instrument, an additional deterministic instrument $\sin((t-1)\pi/n)$, $t = 1, \ldots, n$, and Eicker-White standard errors, in line with the recommendations in Breitung and Demetrescu (2015) and Kostakis et al. (2015). Similarly, we employ Eicker-White inference for OLS.[29] Moreover, we implement the standard LCM, oracle LCM, LCM with $\widehat{\gamma}_x = \widehat{d}_2$, 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 in Andersen and Varneskov (2021*a*) and reflect the dynamic properties of returns and, especially, the persistent predictor variables in Section 6. Specifically, the bandwidth is chosen locally $(m/n \to 0)$ 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 and Varneskov (2021*a*), 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 $\widetilde{m}$ be 75% of m, as,

$$\max\left(\widehat{d}_1(\psi), \widehat{d}_1(\text{TELW})\right),$$

to ameliorate downward bias in the estimate of d_1, as σ_η is increased. However, we almost always have $\widehat{d}_1(\psi) \geq \widehat{d}_1(\text{TELW})$, whenever $\sigma_\eta > 1$. Hence, when using the LCMB-augmented approach to extract information about d_1 under the assumption $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$, 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 $\nu_G = \nu$

[28] We also considered DGPs with $\mathcal{B} = 1.2$ and $\mathcal{B}_\xi = 1.2$. They deliver similar results.

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

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

We first consider the estimation of d_1. Figure 1 displays the estimated integration order of returns as a function of σ_η for the baseline sample size $n = 650$ (top panels) using either the TELW estimator or the LCMB approach based on $\widehat{\gamma}_x = \widehat{d}_2$. Moreover, we provide corresponding estimates in the bottom panel for the larger sample size, $n = 2000$. Several features are noteworthy. One, the estimated persistence from the TELW procedure decreases as a function of σ_η, eventually implying failure to reject $d_1 = 0$. This occurs, not surprisingly, more rapidly for the weaker signal, $d_1 = 0.45$, 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. 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 for the persistence configurations $(d_1, d_2) \in \{(0.45, 0.45), (0.80, 0.80)\}$. Figure 2 provides results for OLS, IVX and standard LCM in its left panels and for oracle LCM, LCM with $\widehat{\gamma}_x = \widehat{d}_2$ and the LCMB-augmented procedure in the right panels. Notably, OLS and IVX are oversized for a wide range of σ_η, when 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 and Chung (2000), and the size distortions for return regressions, when applying persistent AR(1) predictors, e.g., Ferson et al. (2003). Our findings indicate that similar problems arise for return regressions with "imperfect" predictors in fractionally integrated settings for empirically relevant scenarios. Moreover, these size distortions grow as the sample size is increased from $n = 650$ to $n = 2000$. It is also important to note that the regressor persistence in Figure 2 is lower than obtained, in Section 6, for common return predictors like the default spread, price-earnings ratio, and 3-month T-bill rate, suggesting that this issue is relevant across a variety of empirical studies. Finally, 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.

Figure 3 provides power results for LCM based on oracle memory, $\widehat{\gamma}_x$ and $\widehat{d}_1(\psi)$ using the identical persistence configurations. The corresponding results for OLS and IVX are omitted due to their serious size distortions in Figure 2. Interestingly, not only do we find that all three implementations exhibit non-trivial power, which improves with the sample size, but we also observe that LCM-based significance tests using $\widehat{\gamma}_x$ and $\widehat{d}_1(\psi)$ match the power of the oracle memory version.

While all significance tests based on LCM exhibit good size properties, Figure 4 reveals that biases plague the original LCM implementation from Andersen and Varneskov (2021*a*) under the alternative hypothesis $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0)$. This is not surprising, since the procedure is implemented with the "wrong" fractional integration order for the returns for most values of σ_η, as visualized by Figure 1. Specifically, the downward bias in the TELW estimates (eventually, obtaining $\widehat{d}_1 \simeq 0$) causes the fractional filtering of y_t to be insufficient, leaving the numerator of higher asymptotic order than the denominator, where $\widehat{\boldsymbol{u}}^c_{t-1}$ has been filtered "correctly", and, thus, inflates the estimate, as $\mathcal{B} > 0$. We label this a *fractional filtering bias*. In contrast, since $d_1 = d_2$ in these scenarios, LCM based on $\widehat{\gamma}_x = \widehat{d}_2$ uses a precise first-step estimate of the conditional mean persistence and, thus, appears to be unbiased, similarly to the oracle implementation. Importantly, we also find that the LCM results based on $\widehat{d}_1(\psi)$ are closer to the oracle ones than to those obtained the standard LCM, reflecting the better quality of the persistence estimates from the LCMB approach in Figure 1. None of the estimators exhibit bias under the null hypothesis $(\mathcal{B}, \mathcal{B}_\xi) = (0, 1.2)$.

Finally, we examine size and power of OLS, IVX and the four LCM implementations in Figure 5 for the persistence configuration $(d_1, d_2) = (0.00, 0.80)$, inplying that the true persistence of returns is substantially smaller than the persistence of the observed regressor. All procedures have good size properties in this setting. However, the tests based on LCM, especially those exploiting oracle memory and the LCMB estimate $\widehat{d}_1(\psi)$, are substantially more powerful, i.e., more reliable in terms of detecting the significance of the regressor. This is intuitive. For OLS and IVX, the asymptotic order of the denominator (i.e., the predictor variance) is larger than the numerator (the predictor-return covariance), causing a downward bias in the regression slope. We label this a *persistence bias*, which also features in our unreported results for the long memory configuration $(d_1, d_2) = \{(0.00, 0.45), (0.45, 0.80)\}$, as it stems from the wedge between d_1 and d_2, not their absolute magnitudes. Similarly, for LCM based on $\widehat{\gamma}_x = \widehat{d}_2$, the lower power is caused by excessive fractional filtering. 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 LCM 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 and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002), and Andersen, Bollerslev, Diebold and 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 and 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 and Varneskov (2021*a*) and the exact local Whittle (ELW) estimator with a correction for the mean, or initial value, of Shimotsu and Phillips (2005) and Shimotsu (2010).[30] 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 nonstationary 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 with an explorative predictive return regression analysis using OLS, IVX and LCM. Specifically, we implement OLS and IVX as described in the simulation study, and the LCM procedure with $\widehat{\gamma}_x = \max_{i=2,\ldots,5} = \widehat{d}_i(\text{TELW}) = 1.0356$, since the latter is insignificantly different from one, and we adopt the tuning parameter configurations $\nu = \nu_G = 0.20$, $\kappa = \kappa_G = 0.70$. Importantly, as discussed in Section 4.3, these LCM selections deliver valid significance tests for the null hypothesis of return predictability as long as $d_1 \leq \gamma_x$ is satisfied. The results are reported in the top panel 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 LCM procedure, we find that the predictors are, indeed, jointly significant at a 1% level, yet DS emerges as the only individually significant predictor among the regressors, judging by the standard errors. Hence, the use of LCM seems to sharpen the test results.

Third, if we impose an assumption of the form $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$ on the system, e.g., pretend to have known ex-ante that DS is significant, then we may use the LCMB approach to construct an estimate of d_1. Hence, for the sake of illustration, we implement the LCMB persistence estimator using a univariate LCM regression with DS, the same tuning parameters and $\varkappa = 4/3$, similarly to the simulation

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

study, and obtain an estimate $\widehat{d}_1(\psi) = 0.7676$. This suggests that asset returns contain a fractionally integrated conditional mean component, which we cannot identify using standard univariate time series techniques. Specifically, the estimate implies a conditional mean process that is strongly persistent, nonstationary, yet mean-reverting. If we take the analysis one step further and implement a second-stage, multivariate LCM analysis with $\widehat{d}_1(\psi) = 0.7676$, $\kappa = \kappa_G = 0.69$ and the same trimming parameters, we obtain the parameter estimates and standard errors in the bottom panel of Table 2. Interestingly, whereas the point estimates and standard errors for RV and PE are similar to those obtained by OLS and IVX, indicating insignificance of the two variables, the corresponding results for DE and TB indicate significance of both. Moreover, 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). However, we have to be careful claiming significance of the regressors in this step, since we cannot necessarily rely on standard critical values and testing procedures without the assumption $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$, due to the sequential nature of the exploratory analysis, as discussed in Remark 8. This is also why we refrain from reporting a Wald statistic and associated $\mathbb{P}$-value in this case. Hence, our test results from LCM with $\widehat{\gamma}_x$ shows significance of DS, and those from $\widehat{d}_1(\psi)$ indicate that auxiliary predictive power may be generated by TB.

Despite these reservations, our findings are sharper than usual for predictive return regressions, especially at short horizons, e.g., Welch and Goyal (2008), Lettau and Ludvigson (2010), Campbell (2018, Chapter 5) and references therein. We attribute this to the advantages of our LCM procedure. First, uncovering the persistence of the conditional mean via our multivariate regression approach, we may adequately filter returns, reducing the impact of the "large" contemporaneous innovations, thus mitigating their asymptotic and finite sample effects. Second, by letting $\ell/m + m/n \to 0$ as $n \to \infty$, 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 and Stambaugh (2009) and Phillips and Lee (2013). These LCM features all mitigate critical attenuation biases, consistent with the larger coefficient estimates for DS and TB from the LCM procedure than from OLS and IVX in Table 2. In contrast, Andersen and Varneskov (2021*a*) 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 a large set of empirical studies, for which standard time series techniques 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 our LCM procedure is consistent and delivers asymptotically Gaussian inference, if we can estimate the persistence of the conditional mean return, d_1. In addition, we develop 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 apply even without an estimate for the return persistence. Simulations illustrate the theoretical arguments and demonstrate favorable finite-sample properties of our LCMB-augmented procedure. Our empirical application generates evidence for a (latent) fractionally integrated conditional mean component in the monthly S&P 500 returns. Moreover, our predictive regression procedure suggests that macroeconomic indicators like the default spread and treasury rates have predictive power for the future equity-index returns.

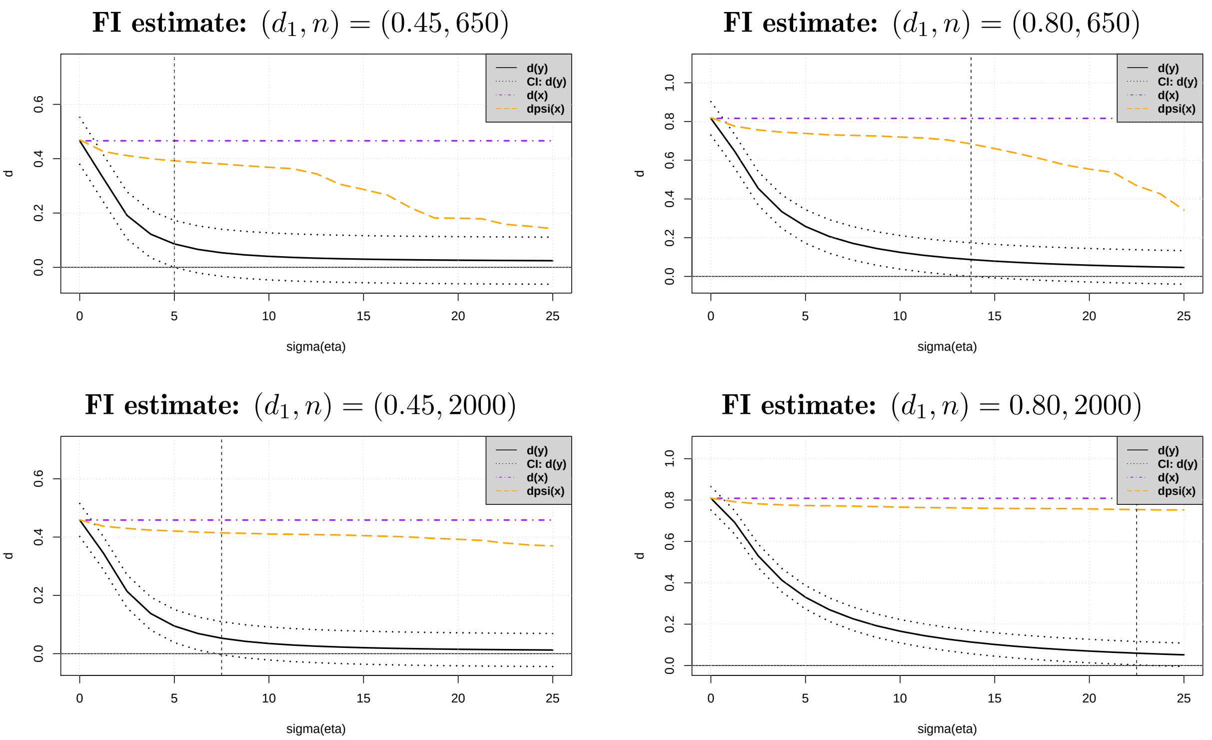

Figure 1: Fractional integration estimation. The panels provide estimates of d_1 and d_2 as a function of the standard deviation of the weakly dependent component of returns, $\sigma_\eta \in [0, 25]$, for different combinations of d_1 and n. Moreover, 95% confidence intervals are provided for $\widehat{d}_1$. The persistence 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, $\widehat{d}_1$, is no longer significantly different from zero. In addition, we include the corresponding LCMB estimate based on $\widehat{\gamma}_x = \widehat{d}_2$, as described in Sections 4-5, which uses a trimming parameter $\nu = \lfloor n^{0.20} \rfloor$ as well as a preliminary bandwidth $m_p = \lfloor n^{0.70} \rfloor$. This is indicated by "dpsi(x)" in the figures. Finally, the simulations are implemented with 1000 replications. The readers are referred to the online version for colored figures.

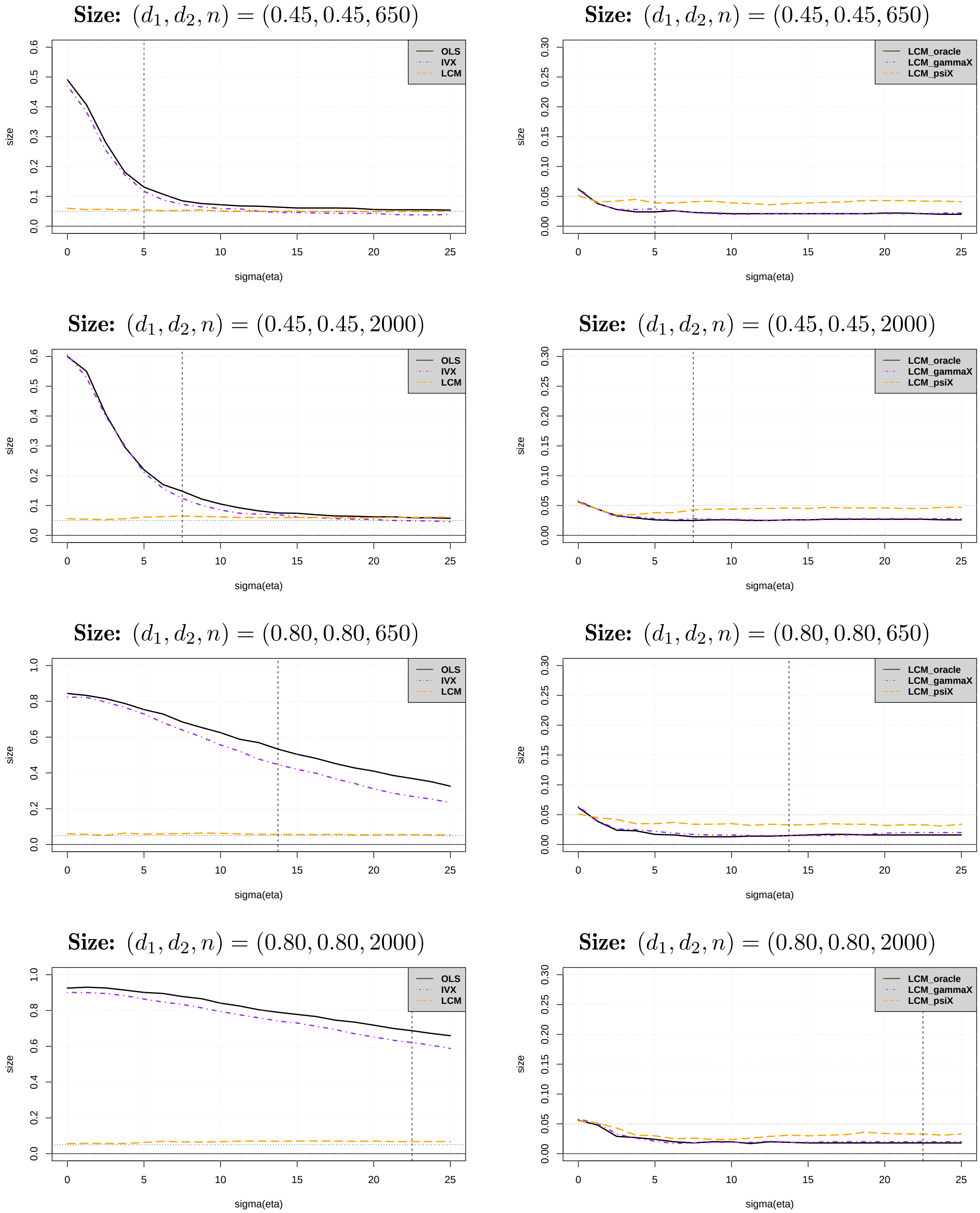

Figure 2: Size properties. The left panels provide the size of OLS, IVX and the standard LCM estimator. The right panels have results for an oracle version of LCM, where (d_1, d_2) is treated as known, LCM with $\widehat{\gamma}_x = \widehat{d}_2$ (LCM_gammaX), and LCM based on the LCMB estimate $\widehat{d}_1(\psi)$ (LCM_psiX), as described in Sections 4 and 5. Finally, we consider different combinations of (d_1, d_2, n), a 5% nominal test size and use 1000 replications. The readers are referred to the online version for colored figures.

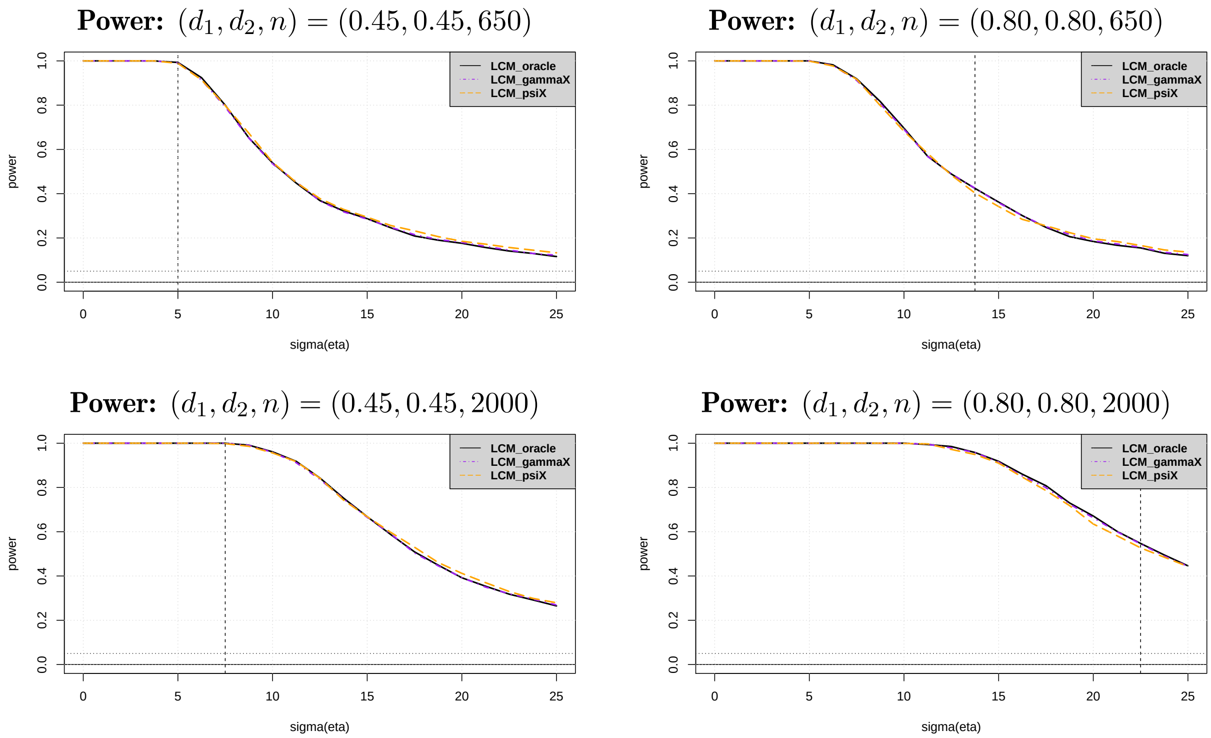

Figure 3: Power properties. The figures display rejection frequencies under the alternative hypothesis $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0.0)$ for an oracle version of LCM, where (d_1, d_2) is treated as known, LCM with $\widehat{\gamma}_x = \widehat{d}_2$ (LCM_gammaX), and LCM based on the LCMB estimate $\widehat{d}_1(\psi)$ (LCM_psiX), as described in Sections 4 and 5. Finally, we consider different combinations of (d_1, d_2, n), a 5% nominal test size and use 1000 replications. The readers are referred to the online version for colored figures.

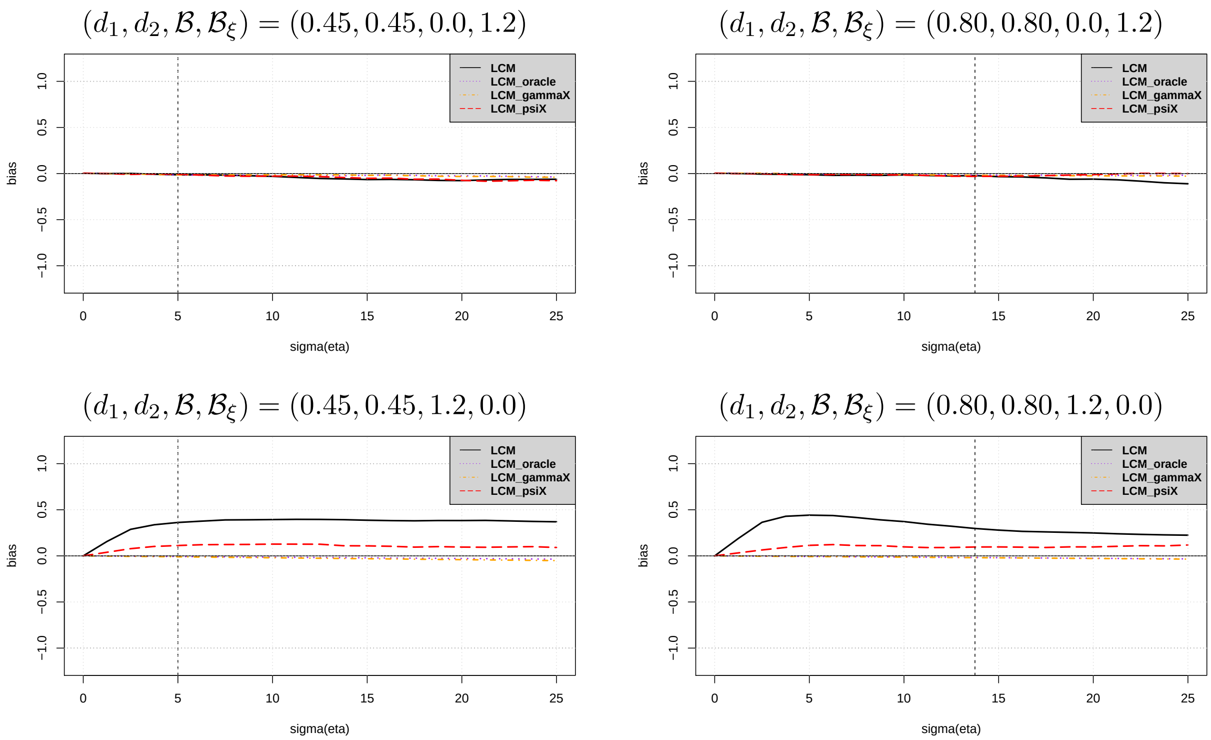

Figure 4: Bias properties. The figures display bias properties under the null hypothesis $(\mathcal{B}, \mathcal{B}_\xi) = (0.0, 1.2)$ and alternative hypothesis $(\mathcal{B}, \mathcal{B}_\xi) = (1.2, 0.0)$. Specifically, results are provided for the standard LCM estimator, an oracle version of LCM, where (d_1, d_2) is treated as known, LCM with $\widehat{\gamma}_x = \widehat{d}_2$ (LCM_gammaX), and LCM based on the LCMB estimate $\widehat{d}_1(\psi)$ (LCM_psiX), as described in Sections 4 and 5. Finally, we consider different combinations of persistence parameters (d_1, d_2), a sample size $n = 650$ and use 1000 replications. The readers are referred to the online version for colored figures.

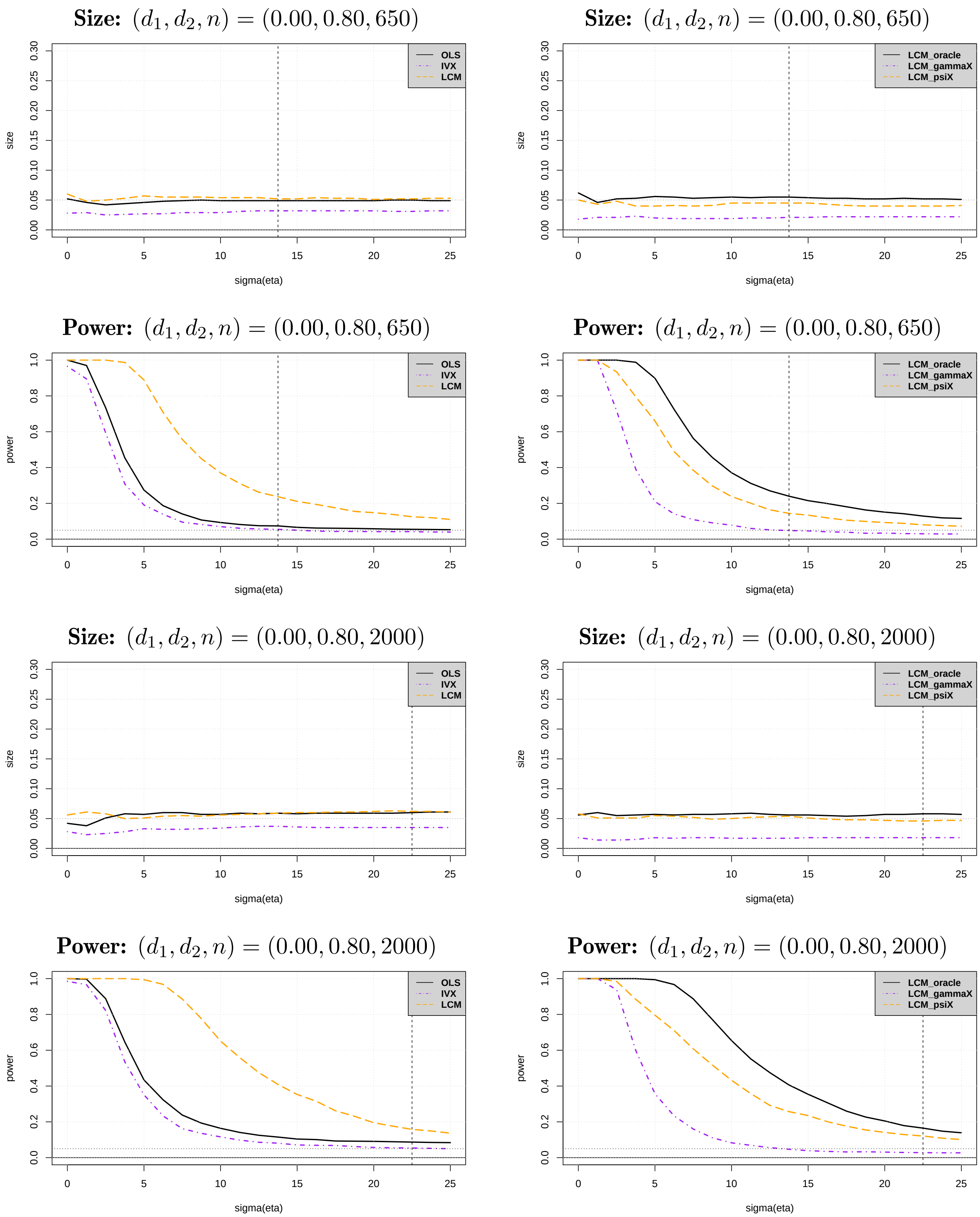

Figure 5: Size and power properties: Impact of balance. The left panels provide the size and power results (see also Figures 2-4 for details) for OLS, IVX and the standard LCM estimator. The right panels have corresponding results for an oracle version of LCM, where (d_1, d_2) is treated as known, LCM with $\widehat{\gamma}_x = \widehat{d}_2$ (LCM_gammaX), and LCM based on the LCMB estimate $\widehat{d}_1(\psi)$ (LCM_psiX), as described in Sections 4 and 5. Finally, we consider different combinations of (d_1, d_2, n), a 5% nominal test size and use 1000 replications. The readers are referred to the online version for colored figures.

Moments and Temporal Dependence					
	Returns_t	RV_{t-1}	DS_{t-1}	PE_{t-1}	TB_{t-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 moments 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 and 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						
Panel A	RV_{t-1}	DS_{t-1}	TB_{t-1}	PE_{t-1}	Wald	$\mathbb{P}$-Wald
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
LCM(γ_x)	0.1003 (0.0980)	0.2144 (0.1016)	−0.3679 (0.3021)	−2.8461 (1.5459)	17.0281	0.0019
Panel B	RV_{t-1}	DS_{t-1}	TB_{t-1}	PE_{t-1}	Wald	$\mathbb{P}$-Wald
LCM(d_ψ)	−0.1638 (0.1167)	0.3065 (0.1158)	−0.6771 (0.3352)	−1.2283 (1.7696)	-	-

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 $\widehat{\gamma}_x = \max_{i=2,\ldots,5} d_i = 1.0356$ from the TELW estimates in Table 1 for the return persistence, corresponding to the LCMB approach, or using $\widehat{d_1}(\psi) = 0.7676$, which is obtained under the ex-ante assumption that DS is significant. The two versions are indicated by γ_x and d_ψ parentheses, respectively. Moreover, 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 Eicker-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 ($n = 661$).

Appendix

A Additional Assumptions

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

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

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

B Proofs

This section provides proofs of the main asymptotic results in the paper. Before proceeding, however, we introduce some notation. For a generic vector $\boldsymbol{V}$, let $\boldsymbol{V}(i)$ index the ith element, and, similarly, for a matrix $\boldsymbol{M}$, let $\boldsymbol{M}(i, q)$ denote its (i, q)th element. Moreover, $K \in (0, \infty)$ denotes a generic constant, which may take different values from line to line or from (in)equality to (in)equality. Let $\mathfrak{X}_n$ denote a generic random variable, which depends on n. Then, we adopt the usual notation $\mathfrak{X}_n = O_p(1)$ and $\mathfrak{X}_n = o_p(1)$ to indicate stochastic boundedness and convergence to zero in probability, respectively. Similarly, we adopt the notation $\mathfrak{X}_n = O_p^+(1)$ and $\mathfrak{X}_n = o_p^+(1)$ when $\mathfrak{X}_n$ is a non-negative random variable. The latter is used to state stochastic bounds for inequalities. Sometimes these (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. The proofs of the technical lemmas are deferred to the supplementary material in the Online Appendix.

B.1 Auxiliary Lemmas for Section 3

This section provides two auxiliary lemmas that expand on Theorem 4 and Lemma A.4 in Andersen and Varneskov (2021*a*), 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,

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

$$\widehat{\boldsymbol{G}}^{c}_{\widehat{v}\widehat{v}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\widehat{v}\widehat{v}}(\ell_G, m_G), \quad \widehat{\boldsymbol{G}}^{(d_1,c)}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}^{(d_1)}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G), \tag{B.2}$$

where, as described in Section 3.2, $\widehat{\eta}_t^{(d_1,c)}$ 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 $\widehat{\boldsymbol{v}}_t^c$ rather than the unobservable $\widehat{\boldsymbol{v}}_t$ when calculating key measures and statistics, thus quantifying the impact of regressor endogeneity. The auxiliary lemma differs from AV (2021, Theorem 4) by allowing for cointegration, $d_1 > 0$, and imperfect regressors.

Lemma B.1. *Suppose Assumptions D1-D3, C, M, F, T and T-G hold. Then, for some arbitrarily small $\epsilon > 0$, it follows that,*

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

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

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

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

(e) $\widehat{\boldsymbol{G}}^{(d_1,c)}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) - \widehat{\boldsymbol{G}}^{(d_1)}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) \leq O_p^+((m_G/n)^{\underline{d}_x}/\ell_G^{1+\epsilon}) + O_p((m/n)^{\underline{d}_x}/\ell^{1+\epsilon})$.

The second auxiliary lemma establishes bounds and convergence results for $\widehat{\boldsymbol{G}}^{(d_1)}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G)$ in the cases with and without cointegration. To this end, recall from the proof of Lemma B.1 that we have the decomposition $\widehat{\eta}_t^{(d_1)} = \widehat{\eta}_t^{(d_1,1)} + \widehat{e}_t^{(2)}$ where $\widehat{\eta}_t^{(d_1,1)} = \widehat{e}_t^{(1)} - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)'\widehat{\boldsymbol{u}}_{t-1}$ and the terms $(\widehat{e}_t^{(1)}, \widehat{e}_t^{(2)})$ are defined as in the proof of Lemma B.1, given in the Online Appendix, as,

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

Hence, we can equivalently write,

$$\widehat{\boldsymbol{G}}^{(d_1)}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) = \widehat{\boldsymbol{G}}^{(d_1,1)}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}^{(2)}_{\widehat{e}\widehat{e}}(\ell_G, m_G) + 2\widehat{\boldsymbol{G}}^{(d_1,1,2)}_{\widehat{\eta}\widehat{e}}(\ell_G, m_G), \tag{B.4}$$

whose components will be analyzed separately in the following.

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

(a) $\lambda_{m_G}^{-2d_1}\Big(\widehat{\boldsymbol{G}}^{(2)}_{\widehat{e}\widehat{e}}(\ell_G, m_G) - G_{\eta\eta}/(1+2d_1)\Big) = o_p(1)$,

(b) The following convergence results hold,

$$\widehat{\boldsymbol{G}}^{(d_1,1)}_{\widehat{\eta}\widehat{\eta}}(\ell_G, m_G) = \begin{cases} G_{\xi\xi} + o_p(1), & \textit{under models (ii) and (iii)}, \\ o_p\big(\lambda_{m_G}^{2d_1}\big), & \textit{under model (iv)}. \end{cases}$$

(c) The following convergence results hold,

$$\begin{cases} \widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}^{(d_1)}(\ell_G, m_G) \xrightarrow{\mathbb{P}} G_{\xi\xi}, & \textit{under models (ii) and (iii)}, \\ \lambda_{m_G}^{-2d_1} \widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}^{(d_1)}(\ell_G, m_G) \xrightarrow{\mathbb{P}} G_{\eta\eta}/(1+2d_1), & \textit{under model (iv)}. \end{cases}$$

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

$$\check{e}_t = (1-L)^{\widehat{\gamma}_x}\alpha + \boldsymbol{\mathcal{B}}'(1-L)^{\widehat{\gamma}_x - d_1}\boldsymbol{u}_{t-1} + (1-L)^{\widehat{\gamma}_x - d_1}\xi_{t-1} + (1-L)^{\widehat{\gamma}_x}\eta_t \equiv \sum_{i=1}^{4} \check{e}_t^{(i)}. \tag{B.5}$$

The first lemma collects the DFT result for the four components in (B.5) from AVOA (2020).

Lemma B.3. *Suppose Assumptions D1-D3, C, M-d_1, F-d_1, T-d_1, T-G-d_1 hold. Then, for a sequence of integers $j = 1, \ldots, m$, it follows,*

(a) $w_{\check{e}}^{(1)}(\lambda_j) = O_p\big((j/n)^{\gamma_x} n^{1/2} j^{-1}\big)$.

(b) $w_{\check{e}}^{(2)}(\lambda_j) = \lambda_j^{\psi} e^{-(\pi/2)\psi \mathrm{i}} \boldsymbol{\mathcal{B}}' \boldsymbol{w}_u(\lambda_j) + O_p\big((j/n)^{\psi} n^{1/2} j^{-1}\big) + O_p\big(\ln(n) m_d^{-1/2} n^{1/2} j^{-1}\big)$.

(c) $w_{\check{e}}^{(3)}(\lambda_j) = \lambda_j^{\psi} e^{-(\pi/2)\psi \mathrm{i}} w_{\xi}(\lambda_j) + O_p\big((j/n)^{\psi} n^{1/2} j^{-1}\big) + O_p\big(\ln(n) m_d^{-1/2} n^{1/2} j^{-1}\big)$.

(d) $w_{\check{e}}^{(4)}(\lambda_j) = \lambda_j^{\gamma_x} e^{-(\pi/2)\psi \mathrm{i}} w_{\eta}(\lambda_j) + O_p\big((j/n)^{\gamma_x} n^{1/2} j^{-1}\big) + O_p\big(\ln(n) m_d^{-1/2} n^{1/2} j^{-1}\big)$.

The next lemma provides bounds on the endogenity 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 $\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{c}(\ell, m)$ and $\widehat{\boldsymbol{G}}_{\check{\eta}\check{\eta}}^{(b,c)}(\ell_G, m_G)$.

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

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

(b) $\widehat{\boldsymbol{G}}_{\check{\eta}\check{\eta}}^{(d_1,c)}(\ell_G, m_G) - \widehat{\boldsymbol{G}}_{\check{\eta}\check{\eta}}^{(d_1)}(\ell_G, m_G) \leq O_p^{+}\left((m/n)^{2\psi + \underline{d}_x}/\ell^{1+\epsilon}\right) + O_p^{+}\left((m/n)^{2\psi + \underline{d}_x} m^{\epsilon/2}/\ell_G^{(1+\epsilon)/2}\right)$.

The next lemma turns attention to the TDAC between $\widehat{\boldsymbol{u}}_{t-1}$ and $\check{e}_t$ without an endogenous component in the regressors. Specifically, using the decomposition in (B.5), we can write,

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

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

$$\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(2,1)}(\ell, m) = \frac{2\pi}{n} \sum_{j=\ell}^{m} \Re\Big(\boldsymbol{w}_{\widehat{u}}(\lambda_j) \overline{w}_{\check{e}}^{(2,1)}(\lambda_j)\Big), \quad w_{\check{e}}^{(2,1)}(\lambda_j) = \lambda_j^{\psi} e^{-(\pi/2)\psi \mathrm{i}} \boldsymbol{\mathcal{B}}' \boldsymbol{w}_u(\lambda_j),$$

$$\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(3,1)}(\ell,m)=\frac{2\pi}{n}\sum_{j=\ell}^{m}\Re\Big(\boldsymbol{w}_{\widehat{u}}(\lambda_j)\overline{w}_{\check{e}}^{(3,1)}(\lambda_j)\Big),\quad w_{\check{e}}^{(3,1)}(\lambda_j)=\lambda_j^{\psi}e^{-(\pi/2)\psi\mathrm{i}}w_{\xi}(\lambda_j),$$
$$\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(4,1)}(\ell,m)=\frac{2\pi}{n}\sum_{j=\ell}^{m}\Re\Big(\boldsymbol{w}_{\widehat{u}}(\lambda_j)\overline{w}_{\check{e}}^{(4,1)}(\lambda_j)\Big),\quad w_{\check{e}}^{(4,1)}(\lambda_j)=\lambda_j^{\gamma_x}e^{-(\pi/2)\gamma_x\mathrm{i}}w_{\eta}(\lambda_j),$$

corresponding, again, to the TDACs for the different components in (B.5), albeit without the (higher-order) terms that induce 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) $\lambda_m^{-1}\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(1)}(\ell,m)\le O_p^+\big((m/n)^{\gamma_x}n^{1/2}/(m^{1-\epsilon}\ell^{1+\epsilon})\big)+O_p^+\big((m/n)^{\gamma_x+\underline{d}_x}n/(m\,\ell^2)\big)$.

(b) $\lambda_m^{-1}\big(\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(2)}(\ell,m)-\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(2,1)}(\ell,m)\big)\le O_p^+\big((m/n)^{\psi}n^{1/2}/(m^{1-\epsilon}\ell^{1+\epsilon})\big)+O_p^+\big((m/n)^{\psi+\underline{d}_x}n/(m\,\ell^2)\big)$.

(c) $\lambda_m^{-1}\big(\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(3)}(\ell,m)-\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(3,1)}(\ell,m)\big)\le O_p^+\big((m/n)^{\psi}n^{1/2}/(m^{1-\epsilon}\ell^{1+\epsilon})\big)+O_p^+\big((m/n)^{\psi+\underline{d}_x}n/(m\,\ell^2)\big)$.

(d) $\lambda_m^{-1}\big(\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(4)}(\ell,m)-\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(4,1)}(\ell,m)\big)\le O_p^+\big((m/n)^{\gamma_x}n^{1/2}/(m^{1-\epsilon}\ell^{1+\epsilon})\big)+O_p^+\big((m/n)^{\gamma_x+\underline{d}_x}n/(m\,\ell^2)\big)$.

Next, let us define

$$\widetilde{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(2,1)}(\ell,m)\equiv\lambda_m^{\psi}\frac{\cos(\psi\pi/2)}{1+\psi}\widehat{\boldsymbol{F}}_{uu}(\ell,m)\boldsymbol{\mathcal{B}},\tag{B.6}$$

concentrate on the differences,

$$\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(2,1)}(\ell,m)-\widetilde{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(2,1)}(\ell,m),\quad\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(3,1)}(\ell,m)-\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(3,1)}(1,m),\quad\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(4,1)}(\ell,m)-\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(4,1)}(1,m),\tag{B.7}$$

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

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

(a) $\lambda_m^{-1-\psi}\big(\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(2,1)}(\ell,m)-\widetilde{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(2,1)}(\ell,m)\big)\le O_p^+\big(m^{-1}\big)+O_p^+\big(m^{-1/2}\lambda_m^{\varpi/2}\big)+O_p^+\big(\lambda_m^{\varpi}\big)$.

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

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

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

(a) $\sqrt{m}\lambda_m^{-1-\psi}\big(\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(3,1)}(\ell,m)-\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(3,1)}(1,m)\big)\le O_p^+\big(\ell^{1+\varpi}/(m^{1/2}n^{\varpi})(\ell/m)^{\psi}\big)$.

(b) $\sqrt{m}\lambda_m^{-1-\gamma_x}\big(\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(4,1)}(\ell,m)-\widehat{\boldsymbol{F}}_{\widehat{u}\check{e}}^{(4,1)}(1,m)\big)\le O_p^+\big(\ell^{1+\varpi}/(m^{1/2}n^{\varpi})(\ell/m)^{\gamma_x}\big)$.

Lemma B.8. *Suppose the conditions of Lemma B.3 hold. Moreover, assume that the conditions* $0<b=d_1<1$ *and* $\max(0,(1-3\kappa/2)/(1+\kappa/2),2d_1)<\varpi\le2$ *are satisfied, then,*

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

The next lemma establishes bounds and convergence results for $\widehat{\boldsymbol{G}}^{(d_1)}_{\breve{\eta}\breve{\eta}}(\ell_G, m_G)$ in the cases with and without cointegration, similarly to Lemma B.2. To this end, recall from equation (B.5) and the proof of Lemma B.4 that we have $\breve{\eta}^{(d_1)}_t = \breve{\eta}^{(d_1,1)}_t + \breve{e}^{(2)}_t$, where $\breve{\eta}^{(d_1,1)}_t = \breve{e}^{(1)}_t - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m, \widehat{\gamma}_x)'\widehat{\boldsymbol{u}}_{t-1}$ and the two terms $(\breve{e}^{(1)}_t, \breve{e}^{(2)}_t)$ are defined in equation (**??**). Hence, we can equivalently write,

$$\widehat{\boldsymbol{G}}^{(d_1)}_{\breve{\eta}\breve{\eta}}(\ell_G, m_G) = \widehat{\boldsymbol{G}}^{(d_1,1)}_{\breve{\eta}\breve{\eta}}(\ell_G, m_G) + \widehat{\boldsymbol{G}}^{(2)}_{\breve{e}\breve{e}}(\ell_G, m_G) + 2\widehat{\boldsymbol{G}}^{(d_1,1,2)}_{\breve{\eta}\breve{e}}(\ell_G, m_G), \tag{B.8}$$

whose components will be analyzed separately in the following.

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

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

(b) The following convergence results hold,

$$\begin{cases} \lambda_{m_G}^{-2\psi}\left(\widehat{\boldsymbol{G}}^{(d_1,1)}_{\breve{\eta}\breve{\eta}}(\ell_G, m_G) - G_{\xi\xi}/(1+2\psi)\right) = o_p(1), & \textit{in (ii)},\\ \lambda_{m_G}^{-2\psi}\left(\widehat{\boldsymbol{G}}^{(d_1,1)}_{\breve{\eta}\breve{\eta}}(\ell_G, m_G) - G_{\xi\xi}/(1+2\psi) - \big(1/(1+2\psi) - c(\psi)^2\big)\boldsymbol{\mathcal{B}}'\boldsymbol{G}_{uu}\boldsymbol{\mathcal{B}}\right) = o_p(1), & \textit{in (iii)},\\ \lambda_{m_G}^{-2\psi}\left(\widehat{\boldsymbol{G}}^{(d_1,1)}_{\breve{\eta}\breve{\eta}}(\ell_G, m_G) - \big(1/(1+2\psi) - c(\psi)^2\big)\boldsymbol{\mathcal{B}}'\boldsymbol{G}_{uu}\boldsymbol{\mathcal{B}}\right) = o_p(1), & \textit{in (iv)}. \end{cases}$$

(c) The following convergence results hold,

$$\lambda_{m_G}^{-2\psi}\widehat{\boldsymbol{G}}^{(d_1)}_{\breve{\eta}\breve{\eta}}(\ell_G, m_G) \xrightarrow{\mathbb{P}} \begin{cases} G_{\xi\xi}/(1+2\psi), & \textit{in model (ii)},\\ G_{\xi\xi}/(1+2\psi) + \big(1/(1+2\psi) - c(\psi)^2\big)\boldsymbol{\mathcal{B}}'\boldsymbol{G}_{uu}\boldsymbol{\mathcal{B}}, & \textit{in model (iii)},\\ \big(1/(1+2\psi) - c(\psi)^2\big)\boldsymbol{\mathcal{B}}'\boldsymbol{G}_{uu}\boldsymbol{\mathcal{B}}, & \textit{in model (iv)}. \end{cases}$$

B.3 Proof of Theorem 1

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

$$\sqrt{m}\lambda_m^{-d_1}\left(\widehat{\boldsymbol{\mathcal{B}}}_c(\ell, m) - \widehat{\boldsymbol{\mathcal{B}}}(\ell, m)\right) \leq O_p^+\left((m/n)^{\underline{d}_x - b}\sqrt{m}/\ell^{1+\epsilon}\right), \tag{B.9}$$

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

$$\widehat{\boldsymbol{\mathcal{B}}}(\ell, m) - \boldsymbol{\mathcal{B}} = \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1}\widehat{\boldsymbol{F}}^{(d_1)}_{u\eta}(1, m) + \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell, m)^{-1}\widehat{\boldsymbol{F}}_{u\xi}(1, m) - \boldsymbol{\mathcal{C}}_1 + \boldsymbol{\mathcal{C}}_2 + \boldsymbol{\mathcal{C}}^{(d_1)}_3 + \boldsymbol{\mathcal{C}}_4, \tag{B.10}$$

where the four error terms, $\boldsymbol{\mathcal{C}}_1$, $\boldsymbol{\mathcal{C}}_2$, $\boldsymbol{\mathcal{C}}_3^{(d_1)}$, and $\boldsymbol{\mathcal{C}}_4$ are defined as

$$\begin{aligned}
\boldsymbol{\mathcal{C}}_1 &\equiv \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{\widehat{u}\mu}^{(u)}(\ell,m)\boldsymbol{\mathcal{B}}, \qquad \boldsymbol{\mathcal{C}}_2 \equiv \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{\widehat{u}\mu}^{(e)}(\ell,m), \\
\boldsymbol{\mathcal{C}}_3^{(d_1)} &\equiv \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\left(\widehat{\boldsymbol{F}}_{\widehat{u}\eta}^{(d_1)}(\ell,m) - \widehat{\boldsymbol{F}}_{u\eta}^{(d_1)}(\ell,m) + \boldsymbol{\mathcal{D}}_1^{(d_1)}\right), \qquad \boldsymbol{\mathcal{D}}_1^{(d_1)} \equiv \widehat{\boldsymbol{F}}_{u\eta}^{(d_1)}(\ell,m) - \widehat{\boldsymbol{F}}_{u\eta}^{(d_1)}(1,m), \\
\boldsymbol{\mathcal{C}}_4 &\equiv \widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\left(\widehat{\boldsymbol{F}}_{\widehat{u}\xi}(\ell,m) - \widehat{\boldsymbol{F}}_{u\xi}(\ell,m) + \boldsymbol{\mathcal{D}}_2\right), \qquad \boldsymbol{\mathcal{D}}_2 \equiv \widehat{\boldsymbol{F}}_{u\xi}(\ell,m) - \widehat{\boldsymbol{F}}_{u\xi}(1,m),
\end{aligned}$$

with the superscripts indicating $\boldsymbol{\mu}_t^{(u)}$ and $\mu_t^{(e)}$, respectively. Whereas the asymptotic properties of the terms $\boldsymbol{\mathcal{C}}_1$, $\boldsymbol{\mathcal{C}}_2$, $\boldsymbol{\mathcal{C}}_3^{(d_1)}$ and $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{u\eta}^{(d_1)}(1,m)$ are the same irrespective of the models (ii)-(iii) and model (iv), the properties $\boldsymbol{\mathcal{C}}_4$ and $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{u\xi}(1,m)$ depend on the inference regime.

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

$$\sqrt{m}\lambda_m^{-d_1}\big(\boldsymbol{\mathcal{C}}_1 + \boldsymbol{\mathcal{C}}_2 + \boldsymbol{\mathcal{C}}_3^{(d_1)}\big) = o_p(1), \quad \lambda_m^{-1}\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{uu}. \tag{B.11}$$

The result, then, follows by applying AVOA (2020, Lemma A.3) to $\sqrt{m}\lambda_m^{-1-d_1}\widehat{\boldsymbol{F}}_{u\eta}^{(d_1)}(1,m)$ in conjunction with (B.11), the continuous mapping theorem and Slutsky's theorem.

Inference for models (ii) and (iii): Since $\widehat{\boldsymbol{F}}_{\widehat{u}\widehat{u}}(\ell,m)^{-1}\widehat{\boldsymbol{F}}_{u\eta}^{(d_1)}(1,m) = O_p(\lambda_m^{d_1}m^{-1/2})$, with $0 < b = d_1 \leq 1$, and we may use AVOA (2020, Lemma A.2) to show $\sqrt{m}\,\boldsymbol{\mathcal{C}}_4 = o_p(1)$, despite ξ_{t-1} being non-trivial, the central limit theory follows by applying AVOA (2020, Lemma A.3) to $\sqrt{m}\lambda_m^{-1}\widehat{\boldsymbol{F}}_{u\xi}(1,m)$ in conjunction with (B.11), 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 $d_1 = b = 0$ and, thus, applies to both inference scenarios. □

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 $\widehat{\boldsymbol{G}}_{\widehat{\eta}\widehat{\eta}}^{(d_1)}(\ell_G, m_G)$, AVOA (2020, Lemma A.4(a)) to establish an equivalent convergence result for $\widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}(\ell_G, m_G)$ and the continuous mapping theorem. □

B.5 Proof of Theorem 3

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

B.6 Proof of Theorem 4

First, define $\boldsymbol{\mathcal{B}}_m = \lambda_m^{\psi}c(\psi)\boldsymbol{\mathcal{B}}$ and the function $g(x_1,x_2) = x_1/x_2$, for $x_2 \neq 0$. Second, note that we can write $\mathcal{R}(\mathcal{L}) = g\big(\boldsymbol{\mathcal{L}}'\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,m,\widehat{\gamma}_x), \boldsymbol{\mathcal{L}}'\widehat{\boldsymbol{\mathcal{B}}}_c(\ell,\widetilde{m},\widehat{\gamma}_x)\big)$. Hence, using Theorem 3 and $\boldsymbol{\mathcal{L}}'\boldsymbol{\mathcal{B}} \neq 0$, 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,

$$g'(\mathcal{L}'\mathcal{B}_m, \mathcal{L}'\mathcal{B}_{\widetilde{m}}) = \frac{\lambda_{\widetilde{m}}^{-\psi}}{c(\psi)\mathcal{L}'\mathcal{B}} \begin{pmatrix} 1 \\ -\varkappa^{\psi} \end{pmatrix} \tag{B.12}$$

implying that one component of the asymptotic variance becomes,

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

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

$$\sqrt{m}\big(\mathcal{R}(\mathcal{L}) - \varkappa^{\psi}\big) \xrightarrow{\mathbb{D}} N\big(0, \varkappa^{2\psi}\mathcal{S}(\mathcal{L},\xi,\psi)\Theta(\varkappa,\psi,\psi)\big), \tag{B.13}$$

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,

$$\sqrt{m}\big(\widehat{\psi}(\mathcal{L}) - \psi\big) \xrightarrow{\mathbb{D}} N\big(0, \mathcal{S}(\mathcal{L},\xi,\psi)\Theta(\varkappa,\psi,\psi)/\ln(\varkappa)^2\big), \tag{B.14}$$

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

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 $\widehat{\boldsymbol{G}}^{(d_1)}_{\check{\eta}\check{\eta}}(\ell_G, m_G)$, AVOA (2020, Lemma A.4(a)) to establish an equivalent convergence result for $\widehat{\boldsymbol{G}}_{\widehat{u}\widehat{u}}(\ell_G, m_G)$ and the continuous mapping theorem. □

Online Supplementary Material

Torben G. Andersen and Rasmus T. Varneskov (June 12, 2022): Supplement to "Consistent Local Spectrum (LCM) Inference for Predictive Return Regressions", Econometric Theory Supplementary Material. To view, please visit:

References

Andersen, T. G. and Benzoni, L. (2009), Stochastic volatility, *in* R. A. Meyers, ed., 'Complex Systems in Finance and Econometrics', Springer-Verlag, pp. 694–726.

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

Andersen, T. G. and Varneskov, R. T. (2021*a*), 'Consistent inference for predictive regressions in persistent economic systems', *Journal of Econometrics* **224**, 215–244.

Andersen, T. G. and Varneskov, R. T. (2021*b*), 'Testing for parameter instability and structural change in persistent predictive regressions', *Journal of Econometrics* **forthcoming**.

Bansal, R., Kiku, D., Shaliastovich, I. and Yaron, A. (2014), 'Volatility, the macroeconomy and asset prices', *The Journal of Finance* **69**, 2471–2511.

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

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

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

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

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

Campbell, J. Y., Giglio, S., Polk, C. and Turley, R. (2018), 'An intertemporal CAPM with stochastic volatility', *Journal of Financial Economics* **128**, 207–233.

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

Cavanagh, C., Elliott, G. and 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. and 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. and 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. and Lütkepohl, H. (1996), 'Making Wald tests work for cointegrated VAR systems', *Econometric Reviews* **15**, 369–386.

Duffy, J. A. and Kasparis, I. (2021), 'Estimation and inference in the presence of fractional $d = 1/2$ and weakly nonstationary processes', *Annals of Statistics* **49**, 1195–1217.

Elliott, G., Müller, U. and 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. and 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. and Taylor, A. R. (2019), 'A bootstrap stationarity test for predictive regression invalidity', *Journal of Business & Economic Statistics* **37**, 528–541.

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

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

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

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

Johansen, S. and 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. and Stamatogiannis, M. P. (2015), 'Robust econometric inference for stock return predictability', *Review of Financial Studies* **28**, 1506–1553.

Leeb, H. and Pötscher, B. (2005), 'Model selection and inference: Fact and fiction', *Econometric Theory* **21**, 21–59.

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

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

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

Lin, Y. and Tu, Y. (2020), 'Robust inference for spurious regressions and cointegrations involving processes moderately deviated from a unit root', *Journal of Econometrics* **219**, 52–65.

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

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

Marriott, F. and 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. and Varneskov, R. T. (2021), 'Frequency dependent risk', *Journal of Financial Economics* **140**, 644–675.

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

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

Nielsen, M. O. and 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. and Tebaldi, C. (2013), 'Long-run risk and the persistence of consumption shocks', *Review of Financial Studies* **26**, 2876–2915.

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

Pastor, L. and 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. and 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. and Lee, J. H. (2016), 'Robust econometric inference with mixed integrated and mildly explosive regressors', *Journal of Econometrics* **192**, 433–450.

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

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

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

Ren, Y., Tu, Y. and 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. (2005), 'The distance between rival nonstationary fractional processes', *Journal of Econometrics* **128**, 283–300.

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

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

Robinson, P. M. and 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. and 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 unkown mean and time trend', *Econometric Theory* **26**, 501–540.

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

Tsay, W.-J. and 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.

Welch, I. and 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.