AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Christensen, B. J., & Varneskov, R. T. (2017). Medium Band Least Squares Estimation of Fractional Cointegration in the Presence of Low-Frequency Contamination. *Journal of Econometrics*, 197(2), 218-244.
<https://doi.org/10.1016/j.jeconom.2016.07.009>

Publication metadata

Title:	Medium Band Least Squares Estimation of Fractional Cointegration in the Presence of Low-Frequency Contamination
Author(s):	Christensen, B. J., & Varneskov, R. T.
Journal:	<i>Journal of Econometrics</i>
DOI/Link:	10.1016/j.jeconom.2016.07.009
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Medium Band Least Squares Estimation of Fractional Cointegration in the Presence of Low-Frequency Contamination*

Bent Jesper Christensen[†]

Aarhus University and CREATES

Rasmus Tangsgaard Varneskov[‡]

Aarhus University and CREATES

July 25, 2016

Abstract

This paper introduces a new estimator of the fractional cointegrating vector between stationary long memory processes that is robust to low-frequency contamination such as random level shifts, outliers, Markov switching means, and certain deterministic trends. In particular, the proposed medium band least squares (MBLS) estimator uses sample-size-dependent trimming of frequencies in the vicinity of the origin to account for such contamination. Consistency and asymptotic normality of the MBLS estimator are established, a feasible inference procedure is proposed, and rigorous tools for assessing the cointegration strength and testing MBLS against the existing narrow band least squares estimator are developed. Finally, the asymptotic framework for the MBLS estimator is used to provide new perspectives on volatility factors in an empirical application to long-span realized variance series for S&P 500 equities.

Keywords: Deterministic Trends, Factor Models, Fractional Cointegration, Long Memory, Realized Variance, Semiparametric Estimation, Structural Change.

JEL classification: C12, C14, C32, C58

*We are grateful to Richard Baillie, Andrew Harvey, Morten Nielsen, Pierre Perron, Peter Phillips, Gene Savin, Enrique Sentana, Robert Taylor, participants in the Conference on Recent Developments in Financial Econometrics and Empirical Finance at the University of Essex, June 2014, and in the 2015 SoFiE conference at Aarhus University as well as the Co-Managing Editor Yacine Aït-Sahalia and two anonymous referees for many useful comments and suggestions. Financial support from the Department of Economics and Business, Aarhus University, and from the Center for Research in Econometric Analysis of Time Series (CREATES), funded by the Danish National Research Foundation (DNRF78), is gratefully acknowledged.

[†]Department of Economics and Business Economics, Aarhus School of Business and Social Sciences, Aarhus University, 8210 Aarhus V., Denmark. Email: bjchristensen@creates.au.dk.

[‡]Corresponding author: Department of Economics and Business Economics, Aarhus School of Business and Social Sciences, Aarhus University, 8210 Aarhus V., Denmark. Email: rvarneskov@creates.au.dk.

1 Introduction

The introduction of cointegration by Granger (1981) and Engle & Granger (1987) has changed the way we think of, and draw inference on, co-movements among persistent time series. Motivated, in part, by a large number of applications in macroeconomics and finance, it has inspired decades of theoretical developments in the econometrics and time series literatures. While most analyses have been concerned with linear combinations of $I(1)$ variables that themselves are $I(0)$, the framework has later been generalized to fractionally integrated time series, i.e., to processes that are $I(d)$ where d is allowed to take non-integer values, nesting $d = 0$ and $d = 1$ as special cases.

For specificity, let v_t be an $I(0)$ process. Then z_t is said to be $I(d)$ if $(1 - L)^d z_t = v_t$ where L is the lag operator and $(1 - L)^d$ is a fractional filter, defined through the binomial expansion

$$(1 - L)^d = \sum_{i=0}^{\infty} \psi_i L^i, \quad \text{where} \quad \psi_i = \frac{\Gamma(i - d)}{\Gamma(-d)\Gamma(i + 1)}$$

with $\Gamma(\cdot)$ denoting the gamma function.¹ The properties of z_t depend critically on the value of d . In this paper, we shall mainly be concerned with the stationary case $0 \leq d < 1/2$, which exhibits long memory whenever $d > 0$ and is characterized by, among others, hyperbolically decaying autocovariances. However, we will also make reference to the nonstationary case, $d \geq 1/2$. For fractionally integrated processes, and as will be discussed in detail below, cointegration exists when a linear combination of $I(d)$ variables is fractionally integrated of strictly lower order than the original series, say $I(d_\epsilon)$, where $d > d_\epsilon$. This concept has been used to shed new light on topics ranging from purchasing power parity (Cheung & Lai 1993) to exchange rate dynamics (Baillie & Bollerslev 1994), bond rates (Dueker & Startz 1998), return volatility forecasting, see, e.g., Bandi & Perron (2006), Christensen & Nielsen (2006), and Andersen & Varneskov (2014), and risk-return modeling, see, for example, Christensen & Nielsen (2007) and Bollerslev, Osterrieder, Sizova & Tauchen (2013).

While early studies of fractional cointegration have been carried out using parametric techniques, Robinson (1994) shows that conventional estimators of the cointegrating vector, such as OLS, are inconsistent when $d_\epsilon > 0$, that is, when the cointegrating relation does not completely purge the regression errors of memory. As an alternative, he introduces the semiparametric narrow band least squares (NBLS) estimator and establishes its consistency in a framework that allows the regression errors to be endogenous, for example as a result of cointegration. Furthermore, all components of the system can exhibit arbitrary short memory dynamics, which may be left unspecified. The asymptotic central limit theory for the NBLS estimator has later been developed for sub-cases of nonstationary fractional cointegration, $d + d_\epsilon > 1/2$, in Robinson & Marinucci (2003), and for stationary fractional cointegration, $d + d_\epsilon < 1/2$, in Christensen & Nielsen (2006).² Recently, Johansen & Nielsen (2012)

¹Strictly speaking, this definition is only valid when $d < 1/2$. For $d \geq 1/2$, fractional integration is defined either by initialization or by considering partial sums. See, e.g., Robinson (2005) and Shimotsu & Phillips (2006) for discussions of the relative advantages and disadvantages of different fractional models.

²Other notable developments are Lobato (1997), who establishes consistency for the multiple regressor case, and Nielsen

have introduced the fractionally cointegrated VAR, or FCVAR, model as a fully parametric alternative to NBLS analysis. Moreover, they develop the necessary asymptotic central limit theory for maximum likelihood (ML) based inference such that the model may be formally analyzed.

Meanwhile, in the univariate case, a parallel literature has documented the presence of level shifts (or structural changes) and other deterministic trends in various time series. For example, Perron (1989) finds evidence of level shifts in the Nelson-Plosser data set, which is comprised of several macroeconomic indicators; Garcia & Perron (1996) find level, or regime, shifts in U.S. real interest rate series; and Qu (2011) rejects the null hypothesis of stationary short- or long memory dynamics against the alternative of level shifts or certain deterministic trends, possibly in conjunction with short memory dynamics, for a U.S. inflation rate series. Moreover, Granger & Hyung (2004), Mikosch & Stărică (2004), Stărică & Granger (2005), Lu & Perron (2010), Perron & Qu (2010), and Varneskov & Perron (2015), among others, provide evidence of level shifts in different asset return volatility series. Together with related work of, e.g., Bhattacharya, Gupta & Waymire (1983), Diebold & Inoue (2001), McCloskey & Perron (2013), and McCloskey & Hill (2015), this univariate literature demonstrates - empirically, theoretically and through simulations - that the presence of level shifts or deterministic trends in an otherwise stationary short memory process will bias log-periodogram and local Whittle estimates of the integration order upward, toward indicating fractional integration, and bias unit root tests toward non-rejection, in both cases introducing “spurious persistence” in the series.³ Similar results are obtained when fitting GARCH, stochastic volatility, or fractional ARIMA models to said processes. This effect arises because level shifts and deterministic trends induce hyperbolically decaying autocovariances and a spectral density pole in the vicinity of the origin, similar to that arising with long memory processes and hence “contaminating” the low-frequency signal in the series.

Despite the devastating impact of low-frequency contamination on various statistical tools that are frequently used for univariate time series analysis, the same problem has received scant attention in the multivariate literature.⁴ Hence, the present paper partially fills this void by analyzing estimation of the cointegrating vector for stationary fractionally cointegrated series that may be contaminated by a broad class of processes, covering, for example, level shifts, certain deterministic trends, Markov switching means, and outliers, and whose exact form need not be specified. Specifically, we introduce the medium band least squares (MBLS) estimator, which, similarly to NBLS, is a frequency domain least squares estimator using an asymptotically degenerating frequency band. However, unlike NBLS, the MBLS estimator utilizes sample-size-dependent trimming of the frequencies in the vicinity of the origin to asymptotically eliminate the impact of low-frequency contamination, exploiting the difference between the stochastic orders of the latent fractional signals and the contamination terms.

We establish consistency and asymptotic normality of the new MBLS estimator. In particular, the

& Frederiksen (2011), who relax an assumption of zero coherence between the regressors and regression errors in the vicinity of the origin and propose a bias-corrected NBLS estimator.

³See also the review Haldrup, Kruse, Teräsvirta & Varneskov (2013).

⁴Among the few exceptions are Campos, Ericsson & Hendry (1996), Gregory & Hansen (1996), and Johansen, Mosconi & Nielsen (2000), who analyze aspects of $I(1)$ - $I(0)$ cointegration, along with Chen & Hurvich (2003) and Robinson & Iacone (2005), who consider fractional cointegration analysis in the presence of polynomial trends.

asymptotic normal distribution is shown to be the same as that enjoyed by the NBLS estimator in the absence of contamination. In general, however, we show that NBLS is inconsistent in the presence of low-frequency contamination, measuring a linear “cointegrating” relation between the contaminating components rather than the fractional signals, see Theorem 2 below, and we conjecture that a similar result will hold for the FCVAR approach. As we will discuss in detail below, this distinction between cointegration among either the fractional signals or their low-frequency contamination is important for interpreting and drawing inference on equilibrium relations in economics and finance, as well as for hedging stock market risk. Hence, to consistently analyze such using the FCVAR model, one would have to correctly specify the low-frequency “contaminating” component, in addition to cointegrating relations and short-memory dynamics, which can be complicated even for simple specifications of deterministic trends. As a result, even researchers interested in parametric modeling of fractional cointegration may find our MBLS estimator useful as a complementary diagnostic tool to ensure low-frequency contamination-robust estimation of the cointegrating relations in the model.

Our solution is related to those in McCloskey & Perron (2013) and McCloskey & Hill (2015), who, in univariate settings, consider log-periodogram estimation of the fractional integration order and frequency domain quasi ML estimation of parametric short memory models, respectively, achieving robustness to low-frequency contamination using trimmed estimators. However, when taking the analysis to the multivariate framework, with the objective to analyze cointegration among stationary long memory processes, we find that the required tool, i.e., our proposed MBLS estimator, including the derived trimming conditions, is distinctively different from previous estimators in the literature. Furthermore, we develop a rigorous feasible inference procedure, along with two empirical diagnostics: A test of cointegration strength and a test of consistency of NBLS inference.

In addition to corroborating our theoretical findings in a simulation exercise, we illustrate the use of the MBLS approach by analyzing volatility factors for S&P 500 equities using a long-span data set of realized variance estimates. In particular, we specify a discrete time stochastic volatility model with two factors and an arbitrary vector of low-frequency contamination, encompassing many stochastic volatility models from different branches of the literature. Motivated by arbitrage pricing theory, we propose to link the most persistent volatility factor of the individual equities to the corresponding factor in market portfolio volatility through a linear fractional cointegrating relation. Using the MBLS approach, we uncover some novel results. First, we provide statistically significant evidence in favor of said specification, and demonstrate that the corresponding NBLS inference procedure fails to detect any significant cointegrating relation, thus “rejecting” the model. Second, we show that the estimated cointegrating vector is significantly larger for MBLS than for NBLS. Hence, only with the former can we adequately measure the degree of volatility dissemination from market volatility to individual equity volatility. Finally, using our empirical diagnostic tools, we formally show that the NBLS inference procedure is inconsistent in our empirical example.

The outline of the paper is as follows. Section 2 introduces the estimation problem and the notion of low-frequency contamination. Section 3 motivates the MBLS estimator and establishes its consistency

for the cointegrating vector. Section 4 provides a central limit theorem and develops feasible inference procedures, along with empirical diagnostic tools. Section 5 contains implementation details and the simulation study. Section 6 presents the empirical analysis. Finally, Section 7 concludes. The appendix contains additional theory and the proofs of the main asymptotic results in the paper.

2 The Estimation Problem

In this section, we introduce the stationary fractional cointegration setting that will be the focal point of our analysis. We highlight the estimation problem that arises when the fractional, or long memory, signal in the variables of interest is measured with low-frequency contamination. Furthermore, we clarify which processes classify as low-frequency contamination in our multivariate setting.

2.1 Fractionally Cointegrating Regressions

Suppose we observe a $p \times 1$ vector $\mathbf{z}_t = (y_t, \mathbf{x}_t')'$, $t = 1, \dots, n$, which is characterized by the decomposition

$$\mathbf{z}_t = \mathbf{c} + \mathbf{z}_t^* + \mathbf{u}_t, \quad (1)$$

where $\mathbf{c}$ is a $p \times 1$ vector of constants, $\mathbf{u}_t$ is a vector consisting of low-frequency contamination, and, finally, $\mathbf{z}_t^* \in I(d_1, \dots, d_p)$ is a vector of stationary fractionally integrated variables, i.e., $d_i \in (0, 1/2)$, $i = 1, \dots, p$, for which we wish to draw inference on linear cointegrating relations. We will adopt formal assumptions on both $\mathbf{u}_t$ and $\mathbf{z}_t^*$ below. For specificity, we are interested in inference on the $(p-1) \times 1$ vector $\boldsymbol{\beta}$ from the (latent) regression relation

$$y_t^* = \alpha + \boldsymbol{\beta}' \mathbf{x}_t^* + \epsilon_t^*, \quad (2)$$

where $\epsilon_t^* \in I(d_\epsilon)$ with $0 \leq d_\epsilon < \min_{i=1, \dots, p} d_i$, that is, elements of $\mathbf{z}_t^*$ are stationary fractionally cointegrated. However, the regression relation we immediately observe is, instead, of the form,

$$y_t = a + \mathbf{b}' \mathbf{x}_t + e_t. \quad (3)$$

As such, this looks similar to the classical errors-in-variables (EIV) estimation problem, but the present challenge is distinct and, in fact, more complicated, because the persistence of the “errors” affects estimates of the long-run relation between y_t^* and $\mathbf{x}_t^*$. In this situation, it is well-acknowledged that OLS is inconsistent. However, we emphasize the more subtle point that methods developed specifically for measuring and drawing inference on long-run relations among fractionally integrated processes, such as the narrow-band least squares (NBLS) estimator, cf. Robinson (1994), Robinson & Marinucci (2003), and Christensen & Nielsen (2006), will *not* work, either. This will be formally stated below. The specific problem of estimating cointegrating relations among stationary fractionally integrated processes in the presence of low-frequency contamination is what we address in this paper.

Remark 1. Beside the classical EIV problem, (2) and (3) also relate to the distinction between deterministic and stochastic cointegration, see Campbell & Perron (1991). Specifically, we allow the low-frequency contaminating components to include both deterministic and stochastic terms, and to cointegrate. While this may be of separate interest, the presence of contamination in the observable regression (3) prevents us from directly observing the stochastic cointegrating relation in (2), on which we seek to draw inference. The distinction between estimating cointegration in (2) and among the low-frequency contaminating processes is carefully discussed in Section 3 below.

2.2 Low-frequency Contamination

To characterize the low-frequency contaminating component, we use a multivariate extension of McCloskey & Perron (2013, Assumption A.1), see also McCloskey (2013) and McCloskey & Hill (2015). Before proceeding, however, let

$$\mathbf{w}_h(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \mathbf{h}_t e^{it\lambda_j}$$

be the discrete Fourier transform for a generic vector time series, $\mathbf{h}_t$, $t = 1, \dots, n$, where $i = \sqrt{-1}$ and $\lambda_j = 2\pi j/n$ are the Fourier frequencies. Furthermore, let $\mathbf{k}_t$ be another (compatible) vector time series. Then

$$\mathbf{I}_{hk}(\lambda_j) = \mathbf{w}_h(\lambda_j) \bar{\mathbf{w}}_k(\lambda_j) = \Re(\mathbf{I}_{hk}(\lambda_j)) + i\Im(\mathbf{I}_{hk}(\lambda_j))$$

defines the cross-periodogram, with $\bar{\mathbf{w}}_k(\cdot)$ indicating transposition and complex conjugation of $\mathbf{w}_k(\cdot)$, and $\Re(\cdot)$ and $\Im(\cdot)$ the real and imaginary parts, respectively. Let $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$, $o_p(\cdot)$ denote the usual (stochastic) orders of magnitude, sometimes referring to matrices, vectors, and scalars; $\|\cdot\|$ the Euclidean matrix norm; and let $\stackrel{\mathbb{D}}{\sim}$ mean distributed as. We then characterize the low-frequency contaminating component $\mathbf{u}_t$ in (1) using the following high-level assumption:

Assumption 1. $\mathbb{E}[\mathbf{I}_{uu}(\lambda_j)] = O(n/j^2)$, and $\lim_{n \rightarrow \infty} \mathbb{V}[(j^2/n)\|\mathbf{I}_{uu}(\lambda_j)\|] < \infty$, $\forall j \neq 0$.

Example 1 (Bivariate Random Level Shifts). Consider a bivariate random level shift specification $\mathbf{u}_t = (u_{1,t}, u_{2,t})'$, $t = 1, \dots, n$, which has individual elements defined by

$$u_{1,t} = \sum_{j=1}^t \pi_{n,j,1} \eta_{j,1} + \sum_{j=1}^t \pi_{n,j,3} \eta_{j,3}, \quad u_{2,t} = \sum_{j=1}^t \pi_{n,j,2} \eta_{j,2} + \sum_{j=1}^t \pi_{n,j,3} \eta_{j,3}$$

where, for $q = 1, 2, 3$, $\pi_{n,j,q} \stackrel{\mathbb{D}}{\sim}$ i.i.d. Bernoulli(p_q/n) with $0 < p_q < n$ fixed, $\eta_{j,q} \stackrel{\mathbb{D}}{\sim}$ i.i.d. $N(0, \sigma_{\eta,q}^2)$, and the components $\pi_{n,j,q}$ and $\eta_{j,q}$ are independent for all combinations of j and q .

Lemma 1. Suppose that $\mathbf{u}_t = (u_{1,t}, u_{2,t})'$, $t = 1, \dots, n$, obeys the bivariate data generating process in Example 1. Then, for frequency ordinates, λ_j , $j = 1, \dots, \lfloor n/2 \rfloor$, it follows that

$$(1) \lim_{n \rightarrow \infty} \mathbb{E}[(j^2/n)I_{u_1 u_1}(\lambda_j)] = (p_1 \sigma_{\eta,1}^2 + p_3 \sigma_{\eta,3}^2)/(4\pi^3) \text{ and } \lim_{n \rightarrow \infty} \mathbb{V}[(j^2/n)I_{u_1 u_1}(\lambda_j)] < \infty.$$

- (2) $\lim_{n \rightarrow \infty} \mathbb{E}[(j^2/n)I_{u_2 u_2}(\lambda_j)] = (p_2 \sigma_{\eta,2}^2 + p_3 \sigma_{\eta,3}^2)/(4\pi^3)$ and $\lim_{n \rightarrow \infty} \mathbb{V}[(j^2/n)I_{u_2 u_2}(\lambda_j)] < \infty$.
- (3) $\lim_{n \rightarrow \infty} \mathbb{E}[(j^2/n)I_{u_1 u_2}(\lambda_j)] \leq (p_1 \sigma_{\eta,1}^2 + p_3 \sigma_{\eta,3}^2)^{1/2} (p_2 \sigma_{\eta,2}^2 + p_3 \sigma_{\eta,3}^2)^{1/2} / (4\pi^3)$ and the limiting variance satisfies $\lim_{n \rightarrow \infty} \mathbb{V}[(j^2/n)I_{u_1 u_2}(\lambda_j)] < \infty$.

Proof. The results follow directly from repeated use of Perron & Qu (2010, Proposition 3) and the Cauchy-Schwarz inequality. $\square$

Example 1, along with Lemma 1, illustrates an important case that satisfies the high-level Assumption 1. In particular, it demonstrates that our framework includes the case of a bivariate low-frequency contamination process consisting of individual and common structural breaks, each of which is specified as in Perron & Qu (2010). Along the same lines, a Cauchy-Schwarz inequality argument in conjunction with related results from the univariate literature shows that Assumption 1 is satisfied when the low-frequency contaminating process contains deterministic level shifts (Mikosch & Stărică (2004) and McCloskey & Perron (2013)), a broad class of monotone and Lipschitz continuous deterministic trends (Qu 2011), fractional trends, and outliers (Iacone 2010).⁵ Formal definitions of such data generating processes for low-frequency contamination are provided in Appendix A.

Importantly, Assumption 1 does *not* impose restrictions on the conditional distribution of the contaminating process, nor on its dependence on the fractional signal, $\mathbf{z}_t^*$. This is in contrast to earlier work, e.g., McCloskey & Perron (2013) and McCloskey & Hill (2015), who assume independence between $\mathbf{u}_t$ and $\mathbf{z}_t^*$ in the univariate case. Hence, we allow $\mathbf{u}_t$ to be generated by a broader array of processes than those outlined above. For example, Assumption 1 accommodates Markov switching level shifts, as well as general compound Lévy processes where the jump, or break, intensity may display stochastic time-variation and the jump-size distribution need not be Gaussian. Furthermore, these level shift processes may be self-exciting or nonlinear functions of $\mathbf{z}_t^*$. Such features are important for our empirical analysis of asset return volatility factors in Section 6, since they allow the volatility process to have a break, or jump, intensity that may depend on other factors, consistently with, e.g., the option pricing models in Bates (2000) and Andersen, Fusari & Todorov (2015), or be of self-exciting nature, e.g., Aït-Sahalia, Cacho-Diaz & Laeven (2015). To concretize this discussion, and to highlight the incremental value of relaxing these assumptions, we provide examples of Markov switching level shift and compound Lévy processes in Appendix A, where we also formally establish that such processes indeed satisfy the high-level conditions in Assumption 1.

Finally, we note that Assumption 1 is sufficient, but not always necessary, for all forms of contamination mentioned above, and it may be relaxed for special cases. For example, from Iacone (2010), $\lim_{n \rightarrow \infty} (j^2/n) \mathbb{E}[\mathbf{I}_{uu}(\lambda_j)] = \mathbf{0}$ if $\mathbf{u}_t$ consists only of outliers (and $j/n \rightarrow 0$) and fractional trends. Its general semi-parametric form, however, allows us to remain agnostic regarding the specification of contamination in the series, e.g., different series may be afflicted by different forms of contamination, and it encompasses a broad array of contaminating processes.⁶

⁵Robinson (1997) shows that a similar assumption holds for a class of nonparametric mean functions.

⁶The periodogram conditions in Assumption 1, as well as the fractional trend specification in Appendix A, rule out certain

3 Consistency: MBLS versus NBLS

In this section, we introduce and motivate our medium band least squares estimator of the cointegrating vector β in the latent regression relation of interest (2), and we establish its consistency. Furthermore, we analyze the probability limit of NBLS in the present framework with low-frequency contamination. To this end, let “ $\sim$ ” signify that the ratio of the left- and right-hand side tends element-wise to one in the limit and, similarly, let “ $\xrightarrow{\mathbb{P}}$ ” and “ $\xrightarrow{\mathbb{D}}$ ” indicate convergence in probability and distribution, respectively. Moreover, we will assume equality of integration orders for the latent fractional signals, $\mathbf{z}_t^*$, that is, $d_i = d \in (0, 1/2)$, $i = 1, \dots, p$, throughout and subsequently discuss this condition. We introduce the following assumptions to derive the theoretical results.

Assumption 2. *The vector process $\mathbf{z}_t^*$, $t = 0, \pm 1, \dots$, is covariance stationary with spectral density matrix satisfying $\mathbf{f}_{zz}^*(\lambda) \sim \mathbf{\Lambda} \mathbf{G} \mathbf{\Lambda}$ as $\lambda \rightarrow 0^+$ where $\mathbf{G}$ is a $p \times p$ real symmetric matrix whose lower right $(p-1) \times (p-1)$ submatrix has full rank and $\mathbf{\Lambda} = \text{diag}[\lambda^{-d}, \dots, \lambda^{-d}]$ is $p \times p$. Moreover, there exists a $(p-1) \times 1$ vector $\beta \neq \mathbf{0}$, a constant $g \in (0, \infty)$, and $0 \leq d_\epsilon < d$ such that*

$$(1, -\beta') \mathbf{f}_{zz}^*(\lambda) (1, -\beta')' = f_\epsilon(\lambda) \sim g \lambda^{-2d_\epsilon}, \quad \text{as } \lambda \rightarrow 0^+.$$

Assumption 3. *$\mathbf{z}_t^*$ admits the Wold representation*

$$\mathbf{z}_t^* = \boldsymbol{\mu}_z^* + \sum_{j=0}^{\infty} \mathbf{A}_j^* \mathbf{v}_{t-j}, \quad \sum_{j=0}^{\infty} \|\mathbf{A}_j^*\|^2 < \infty,$$

where the innovations $\mathbf{v}_t$ satisfy $\mathbb{E}[\mathbf{v}_t | \mathcal{F}_{t-1}] = \mathbf{0}$, $\mathbb{E}[\mathbf{v}_t \mathbf{v}_t' | \mathcal{F}_{t-1}] = \boldsymbol{\Sigma}$, a.s., and $\mathbf{v}_t \mathbf{v}_t'$ are uniformly integrable. Here, $\boldsymbol{\mu}_z^* = \mathbb{E}[\mathbf{z}_0^*]$, $\boldsymbol{\Sigma}$ is a constant matrix of full rank, and $\mathcal{F}_t = \sigma(\mathbf{v}_s, s \leq t)$ is the σ -field generated by the innovations $\mathbf{v}_s, s \leq t$.

Assumptions 2 and 3 are similar to those provided for consistency in Robinson & Marinucci (2003, Assumptions 1 and 2) and Christensen & Nielsen (2006, Assumptions A and B), but they are stated in terms of the latent process of interest, $\mathbf{z}_t^*$, rather than the observable $\mathbf{z}_t$ in (1), consistent with the nature of the estimation problem at hand. The equality of memory assumption is a sufficient, but not necessary assumption for the consistency analysis below.⁷ However, it is needed to derive the central limit theory in Section 4. Hence, we impose it now for simplicity of exposition. We emphasize that assuming equality of integration order is common to the literature testing for fractional cointegration rank, see, e.g., Robinson & Yajima (2002) and Nielsen & Shimotsu (2007).

These assumptions allow the latent process of interest $\mathbf{z}_t^*$ to exhibit vector fractional ARIMA

polynomial trends in $\mathbf{u}_t$. The impact of the latter has already been studied by Chen & Hurvich (2003) and Robinson & Iacone (2005), in a fractional cointegration context. Their analyses, however, do not accommodate the broad array of contaminating processes discussed above.

⁷To verify this statement, we establish Theorem 1 (a) below, the key intermediate result in proving consistency of our medium band least squares estimator, without requiring equality of integration orders. Details are provided in the proof of Theorem 1 in Appendix B.1. The same comments apply to the (in)consistency analysis of NBLS.

dynamics with non-zero coherence between its individual elements at the origin. In our setting, however, the presence of low-frequency contamination in $\mathbf{z}_t$ introduces a bias of the (co-)periodogram in the *close* vicinity of the origin, which is highlighted by the decomposition

$$\mathbf{I}_{zz}(\lambda_j) = \mathbf{I}_{zz}^*(\lambda_j) + \mathbf{I}_{uu}(\lambda_j) + 2\mathbf{I}_{zu}^*(\lambda_j) = O_p\left((n/j)^{2d}\right) + O_p\left(n/j^2\right) + O_p\left((n/j)^{1+d}n^{-1/2}\right) \quad (4)$$

as $\lambda_j \rightarrow 0^+$, where the $(*)$ superscript is used as generic notation for a (co-)periodogram or spectral density for the latent $\mathbf{z}_t^*$. The component $\mathbf{u}_t$, thus, dominates the periodogram for frequencies λ_j such that $j = o(n^{(1-2d)/(2-2d)})$ and vice versa for $jn^{(2d-1)/(2-2d)} \rightarrow \infty$, see also the corresponding discussion in McCloskey & Perron (2013). It is important to note that since the impact of low-frequency contamination tapers off more rapidly as j increases than the signal from the latent fractionally integrated variables of interest, we may identify and estimate the cointegration vector β in (2) by carefully selecting which frequency ordinates to include in the analysis.

3.1 Consistency of MBLS

The decomposition in (4) shows that the bias due to low-frequency contamination dominates the periodogram $\mathbf{I}_{zz}(\lambda_j)$ at frequency ordinates in the close vicinity of the origin. This motivates us to introduce the trimmed discretely averaged co-periodogram,

$$\hat{\mathbf{F}}_{hk}(\ell, m) = \frac{2\pi}{n} \sum_{j=\ell}^m \Re(\mathbf{I}_{hk}(\lambda_j)), \quad 1 \leq \ell \leq m \leq n, \quad (5)$$

which is used to define our trimmed NBLS estimator, labelled the medium band least squares (MBLS) estimator, as

$$\hat{\beta}(\ell, m) = \hat{\mathbf{F}}_{xx}(\ell, m)^{-1} \hat{\mathbf{F}}_{xy}(\ell, m). \quad (6)$$

The trimming function, ℓ , is crucial for selecting the frequency range where $\mathbf{I}_{zz}(\lambda_j)$ is asymptotically free from low-frequency contamination bias, such that we may consistently estimate β , and form asymptotically normal inference. The design of the MBLS estimator in (6) makes it flexible enough to nest earlier estimators in the literature. Specifically, NBLS is recovered as the special case $\ell = 1$, that is, as $\hat{\beta}(1, m)$, and OLS is algebraically equivalent to $\hat{\beta}(1, n-1)$. In general, however, more sophisticated conditions on the trimming and bandwidth functions are required to obtain consistency and asymptotic normality. These will be stated formally below and discussed accordingly.

As a final step before stating consistency of the MBLS estimator, define the continuously averaged spectral density matrix as $\mathbf{F}_{zz}^*(\lambda) = \int_0^\lambda \Re(\mathbf{f}_{zz}^*(s))ds$. The latter is not only important for establishing our asymptotic results, a scaled version of it also appears in the asymptotic variance of our central limit theorem for the MBLS estimator in Section 4. Moreover, Assumptions 2 and 3 imply that the local approximation

$$\mathbf{F}_{zz}^*(\lambda) \sim \mathbf{G}(\lambda) \quad \text{as } \lambda \rightarrow 0^+$$

holds, where the right-hand-side limit is the $p \times p$ matrix $\mathbf{G}(\lambda) = \lambda^{1-2d} \mathbf{G}/(1-2d)$.

Theorem 1. *Under Assumptions 1-3, and with $m = m(n)$ and $\ell = \ell(n)$ positive sequences of integers satisfying $1/m + m/n + \ell/m \rightarrow 0$ as $n \rightarrow \infty$, we have*

$$(a) \quad |\hat{\mathbf{F}}_{zz}(\ell, m) - \mathbf{G}(\lambda_m)| \leq O_p((\ell/n)^{1-2d}) + O_p(\ell^{-2}) + O_p(\ell^{-(1+d)} n^{d-1/2}) + o_p((m/n)^{1-2d}).$$

$$(b) \quad \text{If additionally } 1/\ell + (n/m)^{1-2d} \ell^{-2} \rightarrow 0, \text{ then } \hat{\boldsymbol{\beta}}(\ell, m) \xrightarrow{\mathbb{P}} \boldsymbol{\beta}.$$

Proof. See Appendix B.1. □

Theorem 1 shows that if the trimming function $\ell \rightarrow \infty$, more slowly than the bandwidth, m , but still such that a mild minimum trimming rate requirement $(n/m)^{1-2d} \ell^{-2} \rightarrow 0$ is satisfied, then the MBLS estimator is consistent for $\boldsymbol{\beta}$. In particular, if we assume polynomial functions $m \propto n^\kappa$ and $\ell \propto n^\nu$, then Theorem 1 (b) requires $0 < (1 - \kappa)(1 - 2d)/2 < \nu < \kappa < 1$, whose bounds on ν are naturally related to the decomposition of errors in Theorem 1 (a). Specifically, the first error term measures the loss of information when trimming frequencies, providing the upper bound $\nu < \kappa$. The second error term reflects the bias in the periodogram (4) from low-frequency contamination, leading to the lower bound $(1 - \kappa)(1 - 2d)/2 < \nu$. The latter, intuitively, decreases in both κ and d . If d is large and close to $1/2$, the contaminating component, $\mathbf{u}_t$, will only dominate (4) at few frequency ordinates near the origin, reducing the trimming required to eliminate its bias, and vice versa if d is small. Similarly, if κ is large, the impact of low-frequency contamination on the discretely averaged periodogram (5) is, to some extent, averaged out, again, reducing the lower bound on ν . The fact that $0 < (1 - \kappa)(1 - 2d)/2$, however, shows that we always need $\ell \rightarrow \infty$ for consistency.

The third error term captures cross-products between $\mathbf{u}_t$ and $\mathbf{z}_t^*$, whose presence follows immediately from the fact that Assumption 1 allows for general forms of dependence between the two, thus accommodating a broader array of low-frequency contaminating processes. Finally, the last term mirrors the error in the uncontaminated case, considered by Robinson & Marinucci (2003).

Remark 2. *The fact that we need the trimming parameter ℓ to diverge is consistent with the findings of Iacone (2010), McCloskey (2013), McCloskey & Perron (2013), and McCloskey & Hill (2015), who consider estimation of the fractional integration order, parametric short memory models, and long memory stochastic volatility models in univariate settings with low-frequency contamination. However, the lower bound on trimming, $(1 - \kappa)(1 - 2d)/2 < \nu$, is unique to the present estimation problem.*

3.2 Inconsistency of NBLS

The importance of trimming cannot be exaggerated, as highlighted by the following theorem, which studies the asymptotic behavior of $\hat{\boldsymbol{\beta}}(\ell, m)$ when the trimming function $\ell = \ell_c$ is fixed.

Theorem 2. *Under Assumptions 1-3, and with $1 \leq \ell = \ell_c < m$ a fixed positive integer, and $m = m(n)$ a sequence of integers satisfying $1/m + m/n \rightarrow 0$ as $n \rightarrow \infty$, we have*

(a) $\hat{F}_{uu}(\ell_c, m) \xrightarrow{\mathbb{P}} \mathbf{K}$, where $\mathbf{K}$ is positive semi-definite and $\|\mathbf{K}\| < \infty$. Moreover, uniformly,

$$\hat{F}_{zz}(\ell_c, m) = \mathbf{K} + O_p\left((m/n)^{1-2d}\right) + O_p\left(n^{d-1/2}\right).$$

(b) If we additionally let $\mathbf{K}_{xx}$ be the $(p-1) \times (p-1)$ lower right submatrix of $\mathbf{K}$, $\mathbf{K}_{xy}$ the $(p-1) \times 1$ vector containing the last $(p-1)$ elements of the first column of $\mathbf{K}$, and assume that $\mathbf{K}_{xx}$ is strictly positive definite, then $\hat{\beta}(\ell_c, m) \xrightarrow{\mathbb{P}} \mathbf{K}_{xx}^{-1} \mathbf{K}_{xy}$.

(c) Under the conditions for (b), if $\beta \neq \mathbf{K}_{xx}^{-1} \mathbf{K}_{xy}$, then $\hat{\beta}(\ell_c, m)$ is inconsistent for β .

Proof. See Appendix B.2. □

Theorem 2 highlights the general asymptotic behavior of frequency domain least squares estimators of the cointegrating vector β when variables are observed with low-frequency contamination and trimming is fixed, thus including the NBLS estimator. In particular, when $\mathbf{K}_{xx}$ is positive definite, implying that the order conditions in Assumption 1 are exact, and $\beta \neq \mathbf{K}_{xx}^{-1} \mathbf{K}_{xy}$, ruling out the special case where the low-frequency contamination is “cointegrated” in exactly the same way as the latent fractional signals, these estimators are inconsistent for β . In fact, they do not provide any information about the latent cointegrating relation in (2), asymptotically. Even though the positive definiteness assumption on $\mathbf{K}_{xx}$ rules out certain forms of low-frequency contamination, such as outliers, fractional trends, and its own absence, this is a general and striking result. In particular, and as alluded to in Remark 1, this difference between β and $\beta_{LF} \equiv \mathbf{K}_{xx}^{-1} \mathbf{K}_{xy}$ and, hence, MBLS and NBLS estimates carries important implications for how empirical cointegrating relations should be interpreted.

MBLS allows measuring the equilibrium relation between stationary, albeit latent, variables, whose shocks may cause temporary deviations from β , but will revert back in the long run. If these latent variables are observed with Lipschitz continuous or monotone trends, then MBLS robustly estimates the trend-stationary equilibrium relation. If, on the other hand, they are observed with regime changes in means, MBLS will consistently estimate the long-run relation that prevails within each regime. For example, the international Fisher equation specifies a linear long-run equilibrium between nominal and real interest rates and expected inflation. If these persistent variables contain *unexpected* level shifts, MBLS will remain consistent for the *expected* relation of interest. Given the compelling empirical evidence in, e.g., Garcia & Perron (1996), Baillie & Kapetanios (2007), and Qu (2011) that the long memory properties of such series are generated by stochastic level shift processes, a rigorous assessment of this important relation in international finance will need to account for low-frequency contamination, e.g., using MBLS. Similar considerations need to be taken into account when analyzing long-run relations in, e.g., purchasing power parity (PPP), as, again, Baillie & Kapetanios (2007) find evidence that the persistence in real exchange rate series is generated by shifting processes. We provide a final example in Section 6, showing that robust estimation of β using MBLS is important for disentangling risk sources when hedging stock market volatility exposure.

In comparison, if one applies NBLS and thus estimates β_{LF} , this will either measure a relation between deterministic components or between stochastic processes that need not be stationary, e.g., the random level shift specification in Example 1. Hence, its use may generate misleading conclusions about important relations in economic and finance, such as when testing long-run equilibria for the international Fisher equation and PPP. Only by using a low-frequency contamination robust procedure, such as MBLS, can we be certain that our estimates and tests are consistent.

Remark 3. *Theorem 2 characterizes the probability limit of NBLS, which does not depend on the fractional signal, $\mathbf{z}_t^*$. In finite samples, however, NBLS may retain some information about $\mathbf{z}_t^*$, and we can make general predictions about the impact of low-frequency contamination. Let u_1 denote the first element of $\mathbf{u}$ and, similarly, $\mathbf{u}_2$ the remaining $(p-1) \times 1$ vector, then NBLS decomposes as*

$$\hat{\beta}(1, m) = \beta + \hat{\mathbf{F}}_{xx}^{-1}(1, m) \left[\hat{\mathbf{F}}_{x\epsilon^*}(1, m) + \hat{\mathbf{F}}_{xu_1}(1, m) - \hat{\mathbf{F}}_{xu_2}(1, m)\beta \right] \quad (7)$$

where $\hat{\mathbf{F}}_{xx}(1, m) = \hat{\mathbf{F}}_{xx}^*(1, m) + \hat{\mathbf{F}}_{u_2u_2}(1, m) + 2\hat{\mathbf{F}}_{xu_2}^*(1, m)$. For simplicity of argument, assume here that $\mathbf{u}$ and $\mathbf{z}_t^*$ are independent. First, if there is only contamination in y , then $\hat{\mathbf{F}}_{xx}(1, m) = \hat{\mathbf{F}}_{xx}^*(1, m)$ and the only potential bias contributor inside the brackets is $\hat{\mathbf{F}}_{xu_1}(1, m)$. However, since $\mathbf{u}$ and $\mathbf{z}_t^*$ are independent, $\hat{\mathbf{F}}_{xu_1}(1, m)$ will not generate bias, but rather inflate the variance of the estimation procedure. Second, if there is only contamination in $\mathbf{x}$, the last term inside the brackets will bias $\hat{\beta}(1, m)$ toward $\mathbf{0}$. Finally, if there is contamination in both y and $\mathbf{x}$, the last term will, again, introduce a downward bias. This bias may, however, be partially offset by the second term if u_1 and $\mathbf{u}_2$ are positively correlated. Note that in the last two cases, $\hat{\mathbf{F}}_{xx}^{-1}(\ell, m)\hat{\mathbf{F}}_{xu_1}(\ell, m)$ determines the probability limit of NBLS. We illustrate these patterns in our simulation study in Section 5.

Remark 4. *Not only do Theorems 1 and 2 demonstrate the consistency of MBLS and inconsistency NBLS, respectively, they also show that the band spectrum regression estimators in Engle (1974) and Yamamoto & Perron (2013) cannot be applied, either. For the former, this is evident as $\ell = O(1)$ and $m/n = O(1)$. The trimmed band spectrum regression estimator in the latter, on the other hand, allows for trimming $1/\ell \rightarrow 0$ and $\ell/m \rightarrow 0$, but also requires $m/n = O(1)$ within each frequency range for its asymptotic results to hold. Furthermore, neither of the two frameworks accommodates the presence of fractional integration in the latent process, $\mathbf{z}_t^*$, nor cointegration among its elements.*

4 Asymptotic Normality

We proceed with our analysis and establish asymptotic normality of our MBLS estimator of the cointegrating vector, β , in the presence of low-frequency contamination. As in the frameworks without contamination, e.g., Lobato (1999), Robinson & Marinucci (2003), and Christensen & Nielsen (2006), we will strengthen the assumptions on the latent variables and restate them in terms of $\mathbf{w}_t^* = ((\mathbf{x}_t^*)', \epsilon_t^*)'$ instead of $\mathbf{z}_t^*$. We define the collective memory of the system as $d_c = d + d_\epsilon$, which plays an important role in the derivation of the asymptotic results.

Assumption 2'. The vector process $\mathbf{w}_t^*$, $t = 0, \pm 1, \dots$, is covariance stationary with spectral density matrix satisfying $\mathbf{f}_{ww}^*(\lambda) \sim \mathbf{\Lambda} \mathbf{G} \mathbf{\Lambda}$ as $\lambda \rightarrow 0^+$. In particular, $\mathbf{\Lambda} = \text{diag}[\lambda^{-d}, \dots, \lambda^{-d}, \lambda^{-d_\epsilon}]$ is $p \times p$ with $0 \leq d_\epsilon < d$, there exists a $\varphi \in (0, 2]$ such that $|\mathbf{f}_{ww}^*(\lambda, i, k) - G_{i,k} \lambda^{-2d}| = O(\lambda^{\varphi-2d})$ as $\lambda \rightarrow 0^+$ for $i, k = 1, \dots, p-1$, where $\mathbf{f}_{ww}^*(\lambda, i, k)$ is the (i, k) -th element of $\mathbf{f}_{ww}^*(\lambda)$, $|\mathbf{f}_{ww}^*(\lambda, p, p) - G_{p,p} \lambda^{-2d_\epsilon}| = O(\lambda^{\varphi-2d_\epsilon})$ as $\lambda \rightarrow 0^+$, and $|\mathbf{f}_{ww}^*(\lambda, i, p) - G_{i,p} \lambda^{-d_\epsilon}| = O(\lambda^{\varphi-d_\epsilon})$ as $\lambda \rightarrow 0^+$. Finally, $G_{i,k}$ corresponds to the (i, k) -element of $\mathbf{G}$, which has the same properties as in Assumption 2, and $G_{i,p} = G_{p,i} = 0$.

Assumption 3'. $\mathbf{w}_t^*$ is a linear process, $\mathbf{w}_t^* = \boldsymbol{\mu}_w + \sum_{j=0}^{\infty} \mathbf{A}_j^* \mathbf{v}_{t-j}$, with square summable coefficients, $\sum_{j=0}^{\infty} \|\mathbf{A}_j^*\|^2 < \infty$, the innovations satisfy, almost surely, $\mathbb{E}[\mathbf{v}_t | \mathcal{F}_{t-1}] = \mathbf{0}$, $\mathbb{E}[\mathbf{v}_t \mathbf{v}_t' | \mathcal{F}_{t-1}] = \mathbf{I}_p$, and the matrices $\mathbb{E}[\mathbf{v}_t \otimes \mathbf{v}_t \mathbf{v}_t' | \mathcal{F}_{t-1}]$ and $\mathbb{E}[\mathbf{v}_t \mathbf{v}_t' \otimes \mathbf{v}_t \mathbf{v}_t' | \mathcal{F}_{t-1}]$ are nonstochastic, finite, and do not depend on t , with $\mathcal{F}_t = \sigma(\mathbf{v}_s, s \leq t)$. There exists a random variable ζ such that $\mathbb{E}[\zeta^2] < \infty$ and for all c and some C , $\mathbb{P}[\|\mathbf{w}_t^*\| > c] \leq C \mathbb{P}[\|\zeta\| > c]$. Finally, let the periodogram of $\mathbf{v}_t$ be denoted by $\mathbf{J}(\lambda)$.

Assumption 4. For $\mathbf{A}^*(\lambda, i)$, the i -th row of $\mathbf{A}^*(\lambda) = \sum_{j=0}^{\infty} \mathbf{A}_j^* e^{ij\lambda}$, we have $\|\partial \mathbf{A}^*(\lambda, i) / \partial \lambda\| = O(\lambda^{-1} \|\mathbf{A}^*(\lambda, i)\|)$ as $\lambda \rightarrow 0^+$, for $i = 1, \dots, p$.

Assumption 5. The bandwidth m and trimming ℓ satisfy

$$\frac{1}{m} + \frac{m^{1+2\varphi}}{n^{2\varphi}} + \frac{\ell}{m} + \frac{n^{1-d_\epsilon}}{m^{1/2-d_\epsilon}} \frac{1}{\ell^2} + \frac{\ell^{1+d_\epsilon+\varphi}}{m^{1/2+d_\epsilon} n^\varphi} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Assumptions 2', 3', and 4 resemble the assumptions in Christensen & Nielsen (2006), and they share similarities with those in Robinson (1995a) and Lobato (1999), who analyze uni- and multivariate local Whittle estimation of the fractional integration order. Whereas the linearity, martingale and moment conditions on $\mathbf{w}_t^*$, required by Assumption 3', replace the corresponding Assumption 3, Assumptions 2' and 4 strengthen Assumption 2 by imposing a rate of convergence on the *latent* spectral density matrix, $\mathbf{f}_{ww}^*(\lambda)$, depending on $\varphi \in (0, 2]$, as well as a smoothness condition. We will require slightly stronger conditions on the lower bound for φ below, but these never exclude the empirically relevant vector fractional ARIMA process, for which Assumption 2' holds with $\varphi = 2$. Moreover, the zero-coherence condition $G_{i,p} = G_{p,i} = 0$ in Assumption 2' is important for our central limit theory, and this is the reason we need equality of integration orders, as discussed in Shimotsu (2007) and Nielsen & Frederiksen (2011).⁸ Intuitively, the condition requires the *latent* fractional signals of the regressors and cointegration residuals to be uncorrelated at frequencies in the vicinity of the origin. However, we do allow them to correlate at medium- and short-run frequencies, and the individual latent regressors may correlate at all frequencies. Furthermore, we allow the *observable* regressors and cointegration errors to exhibit co-dependencies at the long-run frequencies through the low-frequency contamination component, e.g., through common breaks (see Example 1), common Lipschitz continuous and monotone trends (see Appendix A), and similar behavior of the observed series, which by Theorem 2 leads to inconsistency of standard estimators of the fractional cointegration vector.

⁸In short, if equality of integration orders is not imposed, then Assumption 2' with $G_{i,p} = G_{p,i} = 0$ will result in a real-valued cross spectral density at the origin, which is incompatible with the one-sided moving average representation of $\mathbf{w}_t^*$ in Assumption 3'. However, these assumptions are made mutually consistent by assuming equality of memory.

While the first two terms in Assumption 5 are well-known from NBLS estimation, the last three terms are caused by the added challenge of asymptotically eliminating the impact of low-frequency contamination, and are thus specific to the MBLS estimator. Similarly to Theorem 1 (a), the fourth term reflects the bias from low-frequency contamination, imposing a lower bound on the required rate at which $1/\ell \rightarrow 0$, and the fifth term the loss of information from trimming frequencies. If we, again, assume polynomial functions $m \propto n^\kappa$ and $\ell \propto n^\nu$, then the trimming conditions in Assumption 5 are satisfied if

$$(1 - d_c)(1 - \kappa)/2 + \kappa/4 < \nu < \min \{(\varphi + \kappa(1/2 + d_c))/(1 + d_c + \varphi), \kappa\}, \quad (8)$$

for which the determinants of the lower bound on ν are similar to Theorem 1 (b). We restrict attention to the stationary fractional cointegration case $0 \leq d_\epsilon < d_c < 1/2$, implying that the lower bound is decreasing in the collective memory of the system, d_c , as well as κ . Whereas the intuition behind the lower bound is unchanged, $(\varphi + \kappa(1/2 + d_c))/(1 + d_c + \varphi)$, the additional determinant of the upper bound, is increasing in both d_c and φ , which has the following interpretation: When d_c and φ are large, less information about the latent fractional signal is lost by trimming the frequencies closest to the origin, since large φ implies that $\mathbf{f}_{ww}^*(\lambda)$ converges faster, making the signal less “noisy” at higher frequencies, and, moreover, since large d_c guarantees that the signal dominates low-frequency contamination of the periodogram (4) for a larger range of frequencies. Hence, we can accommodate higher trimming, despite not necessarily needing it as the lower bound also decreases with d_c , and still extract sufficient information to identify the latent cointegrating vector, β . To ensure consistency between our Assumption 5, (8), and $0 \leq d_\epsilon < d_c < 1/2$, we need to complement Assumption 2’ with a lower bound restriction $\max(0, (\nu - \kappa/2)/(1 - \nu)) < \varphi$. However, assumptions on ν and κ guarantee that Assumption 2’ is always satisfied for some φ since $\max(0, (\nu - \kappa/2)/(1 - \nu)) < 2$. Note that a corresponding condition is not needed for NBLS estimation in the absence of low-frequency contamination. Similarly, we impose one more mild technical lower bound condition on the trimming rate,

$$\frac{1/2 - \kappa(1/2 - d) - (1 - \kappa)d_\epsilon}{1 + d} < \nu, \quad (9)$$

which suffices to eliminate cross-products between the fractional signals and low-frequency contamination. If the latter are independent, this may be disregarded. In our empirical application, we consider two trimming specifications with $\nu > 0.35$ and $\kappa = 3/5$, for which (9) is always satisfied.

Before proceeding to the result, let us write $\mathbf{G}_{xx} = \{G_{i,k}\}_{i,k=1}^{p-1}$ for the $(p-1) \times (p-1)$ leading submatrix of $\mathbf{G}$ from Assumption 2’ and define the transformations

$$\mathbf{H}_{xx}(\lambda) = \frac{\lambda^{1-2d}}{1-2d} \mathbf{G}_{xx}, \quad \mathbf{H}_{xx} = \mathbf{H}_{xx}(\lambda) \lambda^{2d-1}, \quad \mathbf{J}_{xx} = \frac{1}{2(1-2d_c)} \mathbf{G}_{xx}. \quad (10)$$

Theorem 3. *Let Assumptions 1, 2’, 3’, 4, and 5 hold, let the bandwidth and trimming functions have polynomial rates of the form $m \propto n^\kappa$ and $\ell \propto n^\nu$, and suppose that $0 \leq d_\epsilon < d_c < 1/2$, that*

$\max(0, (\nu - \kappa/2)/(1 - \nu)) < \varphi \leq 2$, and that (9) holds. Then

$$\sqrt{m}\lambda_m^{d_\epsilon - d} \left(\hat{\beta}(\ell, m) - \beta \right) \xrightarrow{\mathbb{D}} N \left(\mathbf{0}, G_{p,p} \mathbf{H}_{xx}^{-1} \mathbf{J}_{xx} \mathbf{H}_{xx}^{-1} \right).$$

Proof. See Appendix B.3. □

Theorem 3 demonstrates that in the presence of low-frequency contamination, and under mild conditions on ℓ and m , the MBLS estimator enjoys the same asymptotic normal distribution that, in the absence of contamination, is achieved by the NBLS estimator, see Christensen & Nielsen (2006, Theorem 2). We note that the absence of contamination is a special case of the current framework, and this is generally the only case in which the constant α in the latent regression relation (2) may be identified. Indeed, the limiting distribution does not depend on ℓ , which is consistent with related univariate studies, e.g., McCloskey & Perron (2013) and McCloskey & Hill (2015). This appealing feature implies that our MBLS inference procedure is asymptotically invariant to the choice of this tuning parameter, as long as it satisfies mild rate restrictions. In finite samples, however, the choice of trimming may impact the bias and variance of MBLS. Depending on whether y , $\mathbf{x}$, or both are low-frequency contaminated, Remark 3 shows that insufficient trimming will result in either a variance inflation of the estimator, a bias, or both. On the other hand, excessive trimming may lead to increased variance of the estimator, and render it more sensitive to short-run dynamics. We will, however, defer further discussion of the selection of ℓ and m to Section 5.1.

Remark 5. *The rate requirement for Theorem 3 differs quite suggestively from corresponding rates in the literature. To see this, compare (8) with the corresponding bounds in McCloskey (2013) and McCloskey & Hill (2015), which are $\nu \in (1/2, 4/5)$ and $\nu \in (1/2, 1)$, respectively. While the upper bounds demonstrate the difference between estimation of parametric models in the literature and the semiparametric estimation of the cointegration vector in the present setting, the derived trimming requirement for the MBLS estimator is particularly illustrative of the problem at hand, since it depends not only on the fractional integration order of the input variables of the multivariate system, but also on the cointegration strength, both of which are reflected in the collective memory parameter d_c .*

Remark 6. *It is instructive to rewrite the scaled asymptotic variance of the MBLS estimator as*

$$G_{p,p} \mathbf{H}_{xx}^{-1} \mathbf{J}_{xx} \mathbf{H}_{xx}^{-1} = \frac{(1 - 2d)^2}{2(1 - 2d_c)} G_{p,p} \mathbf{G}_{xx}^{-1},$$

since this illustrates how the accuracy of the MBLS inference procedure depends on the ratio of the latent long-run variance of the regression error, $(1 - L)^{d_\epsilon} \epsilon_t^$, to the latent long-run covariance matrix of regression signal, $(1 - L)^d \mathbf{x}_t^*$. Apart from the scalar $(1 - 2d)^2/2(1 - 2d_c)$, the asymptotic variance resembles that derived by Brillinger (1981, Chapter 8) for a different class of band spectrum estimators. The latter, however, is established under much stricter assumptions than in the present paper, for example, without consideration of low-frequency contamination in the observed series, and accommodating neither fractional integration nor cointegration.*

4.1 Robust Feasible Inference

Until now, we have assumed d and d_ϵ known. However, knowledge of the lower and upper bounds on the rate of required trimming, as well as feasible inference on the MBLs estimator in Theorem 3, depend on consistent estimates of these fractional integration orders. For this purpose, we apply the low-frequency robust trimmed local Whittle (LW) estimator, introduced by Iacone (2010) and refined in McCloskey & Perron (2013), to each element of the multivariate process $\mathbf{z}_t$ separately, to obtain individual estimates $\hat{d}_i$, $i = 1, \dots, p$, and then use $\hat{d} = p^{-1} \sum_{i=1}^p \hat{d}_i$ as a global estimate, following Robinson & Yajima (2002) and Nielsen & Shimotsu (2007). In particular, for a generic univariate series z_t in the vector $\mathbf{z}_t$, where $z_t^* \in I(d)$, the trimmed LW estimator is defined by

$$d_{\text{LW}}(z, \ell_1, m_1) = \underset{\delta \in [0, 1/2)}{\operatorname{argmin}} \log \hat{G}(\delta, z, \ell_1, m_1) - \frac{2\delta}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \log \lambda_j, \quad (11)$$

where

$$\hat{G}(\delta, z, \ell_1, m_1) = \frac{1}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \lambda_j^{2\delta} \Re(I_{zz}(\lambda_j)), \quad (12)$$

writing m_1 and ℓ_1 for the bandwidth and trimming parameters used for the LW estimator to distinguish them from the corresponding m and ℓ used for our MBLs estimator. Furthermore, to ease exposition and clarify the role of the different bandwidth and trimming parameters introduced in this as well as the previous and following sections, an overview of them all is provided in Table 1, including sufficient conditions for both consistency and asymptotic normality of the estimators.

4.1.1 Estimation of d

Despite our assumptions being stated in a multivariate framework, careful inspection shows that they imply that each element of $\mathbf{z}_t$ satisfies the conditions of Iacone (2010, Theorems 2 and 3), suggesting that, under suitable conditions on m_1 and ℓ_1 , we may apply the trimmed LW estimator, along with its asymptotic results, directly in our framework.⁹ Formally,

Assumption 5-LW. *The bandwidth $m_1 \propto n^{\kappa_1}$ and trimming $\ell_1 \propto n^{\nu_1}$ satisfy either*

- (a) $0 < \nu_1 < \kappa_1 < 1$ and $1 - 2d < \kappa_1/(1 - \nu_1)$, or
- (b) $0 < \nu_1 < \kappa_1 < 2\varphi/(1 + 2\varphi)$ and $2(1 - 2d) < \kappa_1/(1 - \nu_1)$.

Lemma 2. *Under Assumptions 1, 2', 3', and 4, and if additionally*

- (a) *Assumption 5-LW (a) or (b) holds, then $d_{\text{LW}}(z, \ell_1, m_1) \xrightarrow{\mathbb{P}} d$.*

⁹One difference between our assumptions and the corresponding ones in Iacone (2010) is that we allow the low-frequency contamination, $\mathbf{u}_t$, to be stochastic and correlated with the latent fractional signal, $\mathbf{z}_t^*$. The order conditions on the periodogram, however, are similar. As we argue in Appendix B.4, the proofs in Iacone (2010) nowhere rely on independence between $\mathbf{u}_t$ and $\mathbf{z}_t^*$, as they use Cauchy-Schwarz arguments to show that cross-product terms are asymptotically negligible. Hence, his results do indeed carry over to our framework.

(b) *Assumption 5-LW (b) holds, then $m_1^{1/2} (d_{\text{LW}}(z, \ell_1, m_1) - d) \xrightarrow{\mathbb{D}} N(0, 1/4)$.*

Proof. See Appendix B.4. □

In Assumption 5-LW, we state conditions on ℓ_1 and m_1 in polynomial terms for comparability with Iacone (2010). Notice, in particular, that the required trimming decreases in d , similarly to the trimming for MBLS in Theorems 1 and 3, since the signal becomes more detectable the higher its fractional integration order, as highlighted by (4). From Lemma 2, the trimmed LW estimator is both consistent and asymptotically normal under mild conditions. While we only require consistency for feasible inference, we note that asymptotic normality is important for our derivation of a test for fractional cointegration strength that is robust to low-frequency contamination, and another test for (in)consistency of the NBLS inference procedure, both in Section 4.2 below.

4.1.2 Estimation of d_ϵ

In addition to estimating d , we also need a consistent estimate of d_ϵ . This is more challenging, as the observable residuals from (3),

$$\hat{e}_t = y_t - \hat{\beta}(\ell, m)' \mathbf{x}_t, \quad (13)$$

differ from the true errors, ϵ_t^* , not only due to low-frequency contamination in y_t and $\mathbf{x}_t$, but also since the MBLS estimate, $\hat{\beta}(\ell, m)$, may differ from the true β in finite samples. The same problem is studied by Nielsen & Frederiksen (2011) in the absence of low-frequency contamination, and their analysis motivates the additional restrictions we impose on the tuning parameters.

Assumption 5-LW'. *Whereas the bandwidth m and trimming ℓ for the MBLS estimator satisfy Assumption 5 and (9), the bandwidth $m_1 \propto n^{\kappa_1}$ and trimming $\ell_1 \propto n^{\nu_1}$ for the LW estimator applied to the observable residuals (13) satisfy either*

(a) *Assumption 5-LW (a) and $(\log n)^4 (\log m_1) (m/m_1)^{d-d_\epsilon} \rightarrow 0$, or*

(b) *Assumption 5-LW (b), (a), and $(m/m_1)^{2(d-d_\epsilon)} m_1^{1/2}/m \rightarrow 0$, as $n \rightarrow \infty$.*

Since $d > d_\epsilon$, Assumption 5-LW' (a) is essentially satisfied if the bandwidth for the LW estimate, m_1 , diverges faster than the corresponding bandwidth for the MBLS estimator, m , as $n \rightarrow \infty$. Assumption 5-LW' (b) slightly strengthens the requirement on the rate at which this occurs. In particular, it imposes an upper bound on κ_1 , which, however, as explicated in Table 1, is rarely binding in practice.

Theorem 4. *Under Assumptions 1, 2', 3', and 4, and if additionally*

(a) *Assumption 5-LW' (a) or (b) holds, then $d_{\text{LW}}(\hat{e}, \ell_1, m_1) \xrightarrow{\mathbb{P}} d_\epsilon$,*

(b) *Assumption 5-LW' (b) holds, then $m_1^{1/2} (d_{\text{LW}}(\hat{e}, \ell_1, m_1) - d_\epsilon) \xrightarrow{\mathbb{D}} N(0, 1/4)$.*

Proof. See Appendix B.5. □

Theorem 4 shows that the results of Lemma 2, i.e., of Iacone (2010), carry over to the case where the input series is the residuals from a fractionally cointegrated relation where the cointegration vector is consistently pre-estimated. In comparison with Nielsen & Frederiksen (2011), we show that the fractional integration order of the residuals from a consistent first-stage frequency domain least squares-type analysis can be estimated consistently and with an asymptotic normal distribution in the presence of low-frequency contamination. As such, Theorem 4 may seem like a small extension of Lemma 2 on the surface. However, we emphasize that exactly this result allows us to introduce a testing procedure for whether standard NBLS inference is inconsistent in Section 4.2.2 below.

4.1.3 Estimation of $\mathbf{G}_{xx}$ and $G_{p,p}$

As a final step in developing feasible inference procedures for Theorem 3, we also need consistent estimates of $\mathbf{G}_{xx}$ and $G_{p,p}$. Hence, we generalize the representation in (12) to accommodate a generic vector time series, $\mathbf{h}_t$, $t = 1, \dots, n$, with fractional integration order $d_h \in [0, 1/2)$, i.e.,

$$\hat{\mathbf{G}}(d_h, \mathbf{h}, \ell_2, m_2) = \frac{1}{m_2 - \ell_2 + 1} \sum_{j=\ell_2}^{m_2} \lambda_j^{2d_h} \Re(\mathbf{I}_{hh}(\lambda_j)), \quad (14)$$

where we write m_2 and ℓ_2 for the bandwidth and trimming parameters to distinguish them from those used for the MBLS and LW estimators. Again, we refer to Table 1 for an overview of the tuning parameters, their restrictions, and cross-restrictions.

Assumption 5-G. *Whereas the bandwidth m and trimming ℓ for MBLS satisfy Assumption 5 and (9), the bandwidth $1/m_2 \rightarrow 0$ and trimming $1/\ell_2 \rightarrow 0$ for the estimator (14) satisfy either*

- (a) $(\log m_2)^2/m_2 + (m_2/n)^\varphi(\log m_2) + \ell_2/m_2 + ((m_2/n)^{d_h}n^{1/2} + (n/\ell_2)^{1-2d_h})(\ell_2 m_2)^{-1} \rightarrow 0$, or
- (b) $(\log m_2)^2/m_2 + (m_2/n)^\varphi(\log m_2) + \ell_2/m_2 + ((m_2/n)^{d_\epsilon}n^{1/2} + (n/\ell_2)^{1-2d_\epsilon})(\ell_2 m_2)^{-1} \rightarrow 0$.

Assumption 5-G imposes mild restrictions on the required trimming, ℓ_2 , which, again, depends on the memory of the underlying process. Interestingly, while we require the bandwidth and trimming functions for MBLS to satisfy the conditions for Theorem 3, similarly to Assumption LW', we do not need to impose cross-restrictions between them and the corresponding m_2 and ℓ_2 when establishing consistency of the estimator $\hat{\mathbf{G}}(d_\epsilon, \hat{\epsilon}, \ell_2, m_2)$ for $G_{p,p}$, even though the observable residuals, $\hat{\epsilon}$, depend on first-stage estimation errors for MBLS. The latter simply vanish sufficiently fast.

Theorem 5. *Under Assumptions 1, 2', 3', and 4, and if additionally*

- (a) *Assumption 5-G (a) holds, then $\hat{\mathbf{G}}(d, \mathbf{x}, \ell_2, m_2) \xrightarrow{\mathbb{P}} \mathbf{G}_{xx}$.*
- (b) *Assumption 5-G (b) holds, then $\hat{G}(d_\epsilon, \hat{\epsilon}, \ell_2, m_2) \xrightarrow{\mathbb{P}} G_{p,p}$.*

Proof. See Appendix B.6. □

Hence, we can estimate d and d_ϵ consistently by Lemma 2 and Theorem 4, along with $\mathbf{G}_{xx}$ and $G_{p,p}$ using Theorem 5, and thereby obtain consistent estimates of $\mathbf{H}_{xx}$ and $\mathbf{J}_{xx}$ in (10) by the continuous mapping theorem and Slutsky's theorem. This enables us to perform feasible inference for the MBLS estimator in the presence of low-frequency contamination using Theorem 3.

4.2 Empirical Diagnostic Tools

In developing the low-frequency contamination robust MBLS methodology described above, we have assumed the existence of a stationary fractional cointegrated equilibrium. In this section, we present new diagnostic tools, not only for assessing this assumption, but also for testing whether NBLS estimation is, in fact, affected by low-frequency contamination.

4.2.1 Testing Cointegration Strength

We are interested in assessing the basic fractional cointegration assumption that a linear combination of stochastic processes is of strictly lower order of fractional integration than the processes themselves. In our setting, this amounts to testing a null hypothesis of the form

$$\mathcal{H}_0 : d - d_\epsilon = \theta \quad \text{for } \theta \in (0, d],$$

where, again, $\mathbf{z}_t^* \in I(d)$ and $\epsilon_t^* \in I(d_\epsilon)$, and the test should account for the fact that these are latent processes. When testing $\mathcal{H}_0$, two possible outcomes would be indicative of fractional cointegration. First, failure to reject is consistent with fractional cointegration of strength $\theta > 0$. Second, rejection in favor of $d - d_\epsilon > \theta > 0$ indicates a stronger fractionally cointegrated relation. Ideally, we would like to test $\mathcal{H}_0$ for the case $\theta = 0$, that is, the case of no cointegration. However, our proposed testing procedure below relies on consistency of MBLS at an appropriate rate, as provided by the CLT in Theorem 3, and orthogonality of $\mathbf{x}_t^*$ and ϵ_t^* , which hold only under fractional cointegration, and thus require $\theta > 0$ (Assumption 2' would be violated for $\theta = 0$). Informally, failure to reject for a very small value of θ could provide indicative evidence against cointegration.

For ease of exposition, we introduce the shorthand notation $\hat{d}_M = d_{\text{LW}}(z, \ell_1, m_1)$ and $\hat{d}_{\epsilon, M} = d_{\text{LW}}(\hat{\epsilon}, \ell_1, m_1)$ for the trimmed LW estimators of the fractional integration orders of $\mathbf{z}_t^*$ and ϵ_t^* , respectively, using the MBLS residuals. By Lemma 2 and Theorem 4, both are robust to low-frequency contamination. Hence, we may introduce the following test.

Corollary 1. *Under Assumptions 1, 2', 3', 4, 5-LW', and $\mathcal{H}_0$,*

$$t_\theta = (2m_1)^{1/2}(\hat{d}_M - \hat{d}_{\epsilon, M} - \theta) \xrightarrow{\mathbb{D}} N(0, 1). \quad (15)$$

We assess the power properties of this test in Section 5.3. Furthermore, we illustrate how the test may be applied as a diagnostic tool in our empirical analysis in Section 6.2 below.

4.2.2 Testing Consistency of NBLS

As motivation for our simple Hausman-type test, recall the fundamental cointegrating condition from Assumption 2,

$$(1, -\beta') \mathbf{f}_{zz}^*(\lambda) (1, -\beta')' = f_\epsilon(\lambda) \sim g\lambda^{-2d_\epsilon}, \quad \text{as } \lambda \rightarrow 0^+, \quad (16)$$

for $0 \leq d_\epsilon < d$. Since $\mathbf{f}_{zz}^*$ is positive definite and the lower right $(p-1) \times (p-1)$ submatrix of $\mathbf{G}$ is of full rank, the full reduction in memory from d to d_ϵ is achieved by a consistent estimate of the unique cointegrating vector β . An inconsistent estimator of β will, thus, lead to an upward bias in the estimate of d_ϵ . Therefore we can reduce the multi-dimensional problem of testing consistency of the NBLS estimator of β to a univariate problem based solely on the estimated fractional integration order of the residuals. Since MBLS is consistent both with and without low-frequency contamination, and NBLS is consistent in the latter case, the residual memory estimators from the two methods are expected to be close in the absence of low-frequency contamination. In fact, under such a null hypothesis, they are perfectly correlated, asymptotically. In the presence of contamination, however, the residual memory parameter estimate from NBLS is expected to be higher than for MBLS due to inconsistency of the procedure, thus suggesting a one-sided test.

Next, to obtain a non-degenerate test statistic, we propose implementing the LW estimator with a bandwidth cm_1 , for some constant $c \in (0, 1)$, when estimating the fractional integration order of the NBLS residuals, writing $\hat{d}_{\epsilon,N}(1, c)$ for the resulting estimator. Meanwhile, we keep the bandwidth m_1 in the trimmed LW estimator applied to the MBLS residuals unchanged.

Corollary 2. *Under Assumptions 2', 3', 4, 5-LW' and $\mathbf{u}_t = \mathbf{0}$, $t = 1, \dots, n$,*

$$t_d(c) = m_1^{1/2}(\hat{d}_{\epsilon,N}(1, c) - \hat{d}_{\epsilon,M}) \xrightarrow{\mathbb{D}} N\left(0, (1/c^{1/2} - 1)^2/4\right).$$

Corollary 2 highlights the importance of slowing down the rate of convergence of $\hat{d}_{\epsilon,N}(1, c)$ by using a fixed proportion of the band to obtain a non-degenerate test statistic. In the presence of low-frequency contamination, however, $t_d(c)$ diverges due to the inconsistency of the NBLS methodology. From equation (16), one source of inconsistency arises from NBLS being inconsistent for β . Another source is error in LW estimation of the residual memory parameter in the presence of low-frequency contamination. While the former induces a positive bias, the direction of bias stemming from the latter is not necessarily clear, as it depends on the specific form of low-frequency contamination. For example, random and deterministic level shifts are known to bias the LW estimator upward, see, e.g., Mikosch & Stărică (2004) and Perron & Qu (2010), while outliers are known to generate the opposite effect, see Haldrup & Nielsen (2007). Hence, to segregate the two sources of inconsistency, we also implement the testing procedure using a trimmed LW estimator to determine the fractional integration order of the NBLS residuals, denoted $\hat{d}_{\epsilon,N}(\ell_1, c)$, in place of $\hat{d}_{\epsilon,N}(1, c)$, as follows.

Corollary 3. *Under Assumptions 2', 3', 4, 5-LW' and $\mathbf{u}_t = \mathbf{0}$, $t = 1, \dots, n$,*

$$\tilde{t}_d(c) = m_1^{1/2}(\hat{d}_{\epsilon, N}(\ell_1, c) - \hat{d}_{\epsilon, M}) \xrightarrow{\mathbb{D}} N\left(0, (1/c^{1/2} - 1)^2/4\right).$$

Finally, notice the size-power tradeoff that comes with the selection of c ; a high value of c leads to high power of the tests, but also to size distortions, and vice versa for a small value of c , though this relation is not linear. A detailed assessment of the tests, however, is deferred to Section 5.3.

5 Implementation and Simulations

In this section, we first discuss implementation details for the MBLS framework, including specific choices of tuning parameters. We then study the finite sample performance of MBLS and NBLS in fractional cointegration settings with low-frequency contamination. Finally, we gauge the size and power properties of the testing procedures introduced in Section 4.2

5.1 Discussion of Tuning Parameters

We have provided sufficient conditions on the divergence rates of the bandwidth and trimming functions for the asymptotic theory developed in the previous sections to hold. These are summarized in Table 1. However, to implement the MBLS estimator, the trimmed local Whittle estimator, and the spectrum estimator in (14), we need to provide specific guidelines for their selection.

First, let $m = n^\kappa$, $m_1 = n^{\kappa_1}$, and $m_2 = n^{\kappa_2}$. That is, all three bandwidths are assumed to diverge at polynomial rates as $n \rightarrow \infty$. For simplicity of exposition, we let $\kappa_1 = \kappa_2$. The asymptotic theory, cf. Assumption 5-LW' for Theorem 4, then dictates the bound $\kappa < \kappa_1 < 2\varphi/(1+2\varphi)$. The latter simplifies by assuming that the latent process of interest $\mathbf{w}_t^*$, whose properties are formalized in Assumptions 2', 3' and 4, follows a vector fractional ARIMA process, since this implies a spectral density smoothness of $\varphi = 2$ in the vicinity of the origin and, hence, $2\varphi/(1+2\varphi) = 4/5$.

Second, let similarly $\ell = n^\nu$, $\ell_1 = n^{\nu_1}$, and $\ell_2 = n^{\nu_2}$ for the three trimming functions, respectively, and simplify by taking $\nu_1 = \nu_2$. We then need $\nu < \kappa$ and $\nu_1 < \kappa_1$, with both ν and ν_1 satisfying additional restrictions, depending on d and d_ϵ either explicitly, or implicitly through d_c . We select a trimming rate for the local Whittle estimator similar to the simple, and conservative, rule-of-thumb $\nu_1 = 1/2 + \zeta$, $\zeta = 0.005$, proposed and extensively studied by McCloskey & Perron (2013), which is valid irrespective of the value of $d \in (0, 1/2)$. This implies $\kappa_1 \in (\kappa \wedge 0.505, 4/5)$.

Finally, we note that for $\varphi = 2$, the upper bound on ν is always determined by κ .¹⁰ The lower bound, on the other hand, is strictly decreasing in both κ and d_c . For the boundary cases $d_c = 0$ and $d_c = 1/2$, it is equal to $(1 - \kappa/2)/2$ and $1/4$, respectively. If, for example, we select $\kappa = 3/5$, as we will do later in the empirical analysis, this implies a conservative restriction $\nu > 0.35$.¹¹

¹⁰This follows since $\frac{2+\kappa(1/2+d_c)}{3+d_c} > \frac{2+\kappa/2}{3} > \kappa$ whenever $\kappa < 4/5$, as required by the bandwidth restriction when $\varphi = 2$.

¹¹The cross-product restriction for asymptotic normality of MBLS in (9) is always satisfied for $\nu > 0.35$ and $\kappa = 3/5$.

5.2 Simulation Study: Evidence in Favor of MBLS

We proceed by studying the finite sample properties of MBLS and NBLs, as well as the proposed diagnostic tools, when the observable series are cloaked with low-frequency contamination. To this end, we simulate a latent cointegrating relation and add various forms of contamination to the series. Specifically, **DGP 1** models x_t^* as a fractional noise process with $d = 0.3$ and i.i.d.N(0, 1) innovations, and it sets $y_t^* = \beta x_t^* + \epsilon_t^*$ with $\beta = 0.7$ and ϵ_t^* similarly being i.i.d.N(0, 1). This baseline data generating process (DGP) resembles our empirical findings below and allows us to assess the finite sample efficiency loss, if any, MBLS may incur when trimming frequencies. Next, three different forms of low-frequency contamination are sequentially added to either y_t^* , x_t^* or both. These are random level shifts (RLS), a Lipschitz continuous deterministic trend (DT), or Markov switching means (MS):

- **RLS.** Let $u_{j,t} = u_{j,t-1} + \pi_{n,t,j}\eta_{t,j}$ for $j = \{1, 2, 3\}$ where $\pi_{n,t,j} \sim \text{i.i.d. Bernoulli}(p/n)$, $p = \{2, 5\}$, and $\eta_{t,j} \sim \text{i.i.d. N}(0, 1)$. Then, **DGP 2 - RLS** has $y_t = y_t^* + u_{1,t}$, **DGP 3 - RLS** has $x_t = x_t^* + u_{2,t}$, and **DGP 4 - RLS** has $y_t = y_t^* + u_{1,t} + u_{3,t}$ along with $x_t = x_t^* + u_{2,t} + u_{3,t}$.
- **DT.** Let $d_t = \sin(4\pi t/n)$ and $\eta_{t,j} \sim \text{i.i.d. N}(0, 1/2)$ for $j = \{1, 2, 3\}$. Then, **DGP 2 - DT** has $y_t = y_t^* + d_t + \eta_{1,t}$, **DGP 3 - DT** has $x_t = x_t^* + d_t + \eta_{2,t}$, and **DGP 4 - DT** has $y_t = y_t^* + d_t + \eta_{1,t} + \eta_{3,t}$ along with $x_t = x_t^* + d_t + \eta_{2,t} + \eta_{3,t}$.
- **MS.** For $j = \{1, 2, 3\}$, let $\eta_{j,t,0} \sim \text{i.i.d. N}(1/2, 1/2)$, $\eta_{j,t,1} \sim \text{i.i.d. N}(-1/2, 1/2)$, and the state variables $s_{j,t} \in \{0, 1\}$ have switching probabilities $p_{01,n} = p_{10,n} = p/n$ where $p = 4$. Moreover, $u_{j,t} = \eta_{j,t,0}\mathbf{1}\{s_{j,t} = 0\} + \eta_{j,t,1}\mathbf{1}\{s_{j,t} = 1\}$. Then, **DGP 2 - MS** has $y_t = y_t^* + u_{1,t}$, **DGP 3 - MS** has $x_t = x_t^* + u_{2,t}$, and **DGP 4 - MS** has $y_t = y_t^* + u_{1,t} + u_{3,t}$ along with $x_t = x_t^* + u_{2,t} + u_{3,t}$.

The design of the bivariate RLS specification follows directly from Example 1, which is a special case of the compound Lévy processes considered in Appendix A. The DT and MS processes, on the other hand, are motivated by the univariate simulation studies of Ohanissian, Russell & Tsay (2008) and Qu (2011) as well as Lemma 3 in Appendix A, which establishes that such MS models satisfy the high-level Assumption 1. We consider three sample sizes $n \in \{200, 500, 2000\}$, where the last two correspond well with the monthly, respectively, weekly realized variance series in the empirical study in Section 6 below. We implement both MBLS and NBLs with $m = n^\kappa$, $\kappa = 0.6$, and MBLS using trimming $\ell = n^\nu$, $\nu = 0.35 + \zeta$ with $\zeta = 0.005$.¹² This trimming specification trivially satisfies the conservative restriction $\nu > 0.35$. The biases and RMSEs of MBLS and NBLs are presented for DGP 1 and the RLS contaminated processes in Table 2, and, similarly, for the DT and MS processes in Table 3.

There are several interesting takeaways from Table 2. First, when there is no low-frequency contamination, both MBLS and NBLs provide unbiased estimates of the cointegrating vector. MBLS exhibits a slight efficiency loss from trimming frequencies for small samples, but this disappears as the sample size increases. Second, if there is contamination in y_t , MBLS is largely unaffected by this,

¹²We also considered trimming with $\nu = 0.40 + \zeta$, which produces highly similar results. Moreover, we will introduce an adaptive trimming procedure in the empirical study in Section 6.

while the variance of the NBLs estimates increases dramatically for all sample sizes considered. Third, if there is contamination in x_t , NBLs suffers from a pronounced downward bias, in addition to large RMSEs. Moreover, the bias and RMSE grow even worse for larger samples. MBLS, on the other hand, almost removes the entire bias and converges to β as the sample size increases. Fourth, when there are low-frequency contamination in both, the results are similar to those for **DGP 3 - RLS**, albeit with NBLs displaying slightly smaller bias and slightly higher RMSEs. These results exactly coincide with the predictions of Theorems 1-2 and Remark 3. In particular, when there is contamination in x_t , NBLs is no longer consistent and, in fact, converges to 0 as $n \rightarrow \infty$. If there is low-frequency contamination in both y_t and x_t , the probability limit is determined by the linear “cointegrating” relation between the contaminating components, partly offsetting this downward bias. MBLS, however, remains consistent in all cases. Finally, the respective patterns for DT and MS contamination in Table 3 are identical.

5.3 Size and Power of Empirical Diagnostic Tools

The finite sample properties of the empirical diagnostic tools are examined using the setup described above, with particular emphasis on the test for consistency of NBLs inference. For the latter, we showed in Section 4.2.2 that when slowing down the bandwidth of the NBLs residual memory parameter estimate with a constant $c \in (0, 1)$, the proposed test-statistic, $t_d(c)$, has a non-generate distribution. However, unreported simulations demonstrate that $t_d(c)$ has pronounced size distortions, in particular for c close to 1, and large samples ($n \geq 10,000$) are needed to make the Gaussian approximation in Corollary 2 appropriate. Hence, we propose using the following adjusted test statistic,

$$t_d^a(c) = \frac{2t_d(c)}{(1/c + 1)^{1/2}} = \frac{2m_1^{1/2}}{(1/c + 1)^{1/2}} \left(\hat{d}_{\epsilon,N}(1, c) - \hat{d}_{\epsilon,M} \right), \quad (17)$$

and apply simulated critical values to determine significance. To motivate this adjustment, write the variance of $t_d(c)$ as $(1/c + 1 - 2c^{-1/2}\rho)/4$ where ρ is the correlation between $\hat{d}_{\epsilon,N}(1, c)$ and $\hat{d}_{\epsilon,M}$, which is 1, asymptotically, when there is no low-frequency contamination in the series. Hence, when c is close to 1, the variance is small and large samples are needed to make the Gaussian approximation suitable. The adjusted statistic (17) simply assumes $\rho = 0$ in the standardization to make the Gaussian approximation better for smaller samples. However, the test no longer has unit variance.¹³

To corroborate these remarks, we implement $t_d^a(c)$ using $c \in \{0.1, 0.2\}$ and the configurations for the local Whittle estimator outlined in Section 5.1, with $\kappa_1 = 0.7$ fixed. We initially consider **DGP 1** and $n \in \{500, 2000\}$ to study the (size) properties of the test under the null hypothesis for sample sizes that mimic our empirical study. To this end, Figure 1 illustrates the empirical distribution of the statistic (17) for the chosen combinations of c and n . In general, the tests are seen to display fatter tails than comparable normal distributions with identical first two moments. When the sample size is

¹³We note that the qualitative conclusions from the empirical study are identical when using the standardization from Corollary 2 to construct the tests and, similarly, apply simulated critical values to determine significance. However, since this test is much further from Gaussianity in finite samples, we prefer using the test statistic (17).

increased from $n = 500$ to $n = 2000$, the tests provide better approximations to a standard Gaussian distribution, albeit not perfect, as expected.

The graphical analysis of $t_d^a(c)$ is elaborated upon by simulating its upper quantiles for **DGP 1**, which are used to determine significance levels for the test, going forward. These are provided in Table 4, where we also report the bias of $\hat{d}_{\epsilon,N}(1, 1) \equiv \hat{d}_{\epsilon,N}$, and $\hat{d}_{\epsilon,M}$ as well as the raw rejection rates (RRs) for $t_d^a(c)$ under the null hypothesis (size), and the adjusted RRs under the alternative DGPs with low-frequency contamination in the series (size-adjusted power), using one-sided tests with a 5% significance level. Moreover, for all DGPs, the RRs of the proposed test for cointegration strength, based on Corollary 1, with $\theta = 0.05$ fixed are reported to gauge the detectability of a significant cointegrating relation.

Table 4 features several interesting results. First, the simulated upper quantiles for $t_d^a(c)$ are above those for a standard Gaussian distribution, but closer for $n = 2000$. Hence, and not surprisingly, the test is over-sized compared to a nominal 5% level. Specifically, it has 18-19% size for $n = 500$, and though it decreases to 10-12% when $n = 2000$, such distortions are non-negligible, and we apply the simulated critical values to correct them. Second, whereas $\hat{d}_{\epsilon,N}$ displays large upward biases for all DGPs with low-frequency contamination, $\hat{d}_{\epsilon,M}$ is much less distorted, showing average estimates fairly close to 0, albeit with a small bias for some RLS specifications. Third, in the present setup, it is easier to detect inconsistency of the NBLS inference when there is low-frequency contamination in y_t than when it contaminates x_t . This is partly due to $\beta < 1$, thus scaling its impact down, and partly due to first-stage estimation errors for MBLS and NBLS.¹⁴ Fourth, the RRs are lowest for **DGP 4 - DT**, that is, when y_t and x_t are contaminated by the same deterministic trend. In this case, the results in Table 3 show that NBLS is upward biased toward 1, implying that said deterministic trend will cancel. Hence, these RRs speak directly to the inconsistency of the NBLS point estimates, which is slightly harder to detect than the joint inconsistency of the point estimates and the feasible inference procedure. Indeed, unreported RRs for a similarly adjusted test $\tilde{t}_d^a(c)$ where $\hat{d}_{\epsilon,N}(\ell_1, c)$ is used in place of $\hat{d}_{\epsilon,N}(1, c)$, see also Corollary 3, are of similar magnitude across all forms of low-frequency contamination. Finally, the test for cointegration strength, t_θ , has decent power for both sample sizes and all forms of low-frequency contamination.

6 Empirical Analysis

This section introduces an alternative volatility factor model for individual assets where the most persistent factor is fractionally integrated and linked to the most persistent factor of market volatility. Moreover, it contains an empirical analysis where we study such factor linkages by applying MBLS to long-span realized variance series for S&P 500 equities.

¹⁴Unreported simulations show that the RRs increase considerably if $\beta = 1.7$ and x_t is contaminated.

6.1 A Market Volatility Factor Model

The presence of long memory in stock market volatility has long been recognized in the financial econometrics literature, see, e.g., Baillie, Bollerslev & Mikkelsen (1996), Granger & Ding (1996), and Andersen, Bollerslev, Diebold & Ebens (2001) for a few early references. Moreover, volatility is often suggested to have two, or more, factors driving its dynamics, to capture both persistent movements and instantaneous innovations to its path.¹⁵ A similar factor decomposition arises from a different literature, which considers estimation of continuous time diffusion models based on discrete time observations.¹⁶ However, as outlined in the introduction, an increasing body of work suggests that persistence in asset return volatility may alternatively be explained by random level shifts, or other forms of discontinuities, in the process, that is, by a type of low-frequency contamination.

With $V_{i,t}$ denoting the log-variance for asset i , we may thus write

$$V_{i,t} = c_i + V_{i,t,1} + V_{i,t,2} + \boldsymbol{\iota}' \mathbf{u}_{i,t} \quad (18)$$

as an encompassing discrete time stochastic volatility model. The model has a two-factor structure, with $V_{i,t,1} \in I(d_i)$ being the most persistent factor (hence, for $V_{i,t,2} \in I(d_{i,2})$, we have $d_i > d_{i,2}$), and it includes $\mathbf{u}_{i,t}$ as a $\tau \times 1$ vector containing possibly different forms of low-frequency contamination. Finally, $\boldsymbol{\iota}$ is a $\tau \times 1$ vector of ones. If we consider the range of processes satisfying Assumption 1, like the various univariate processes outlined in Example 1 and Appendix A, we see that $\mathbf{u}_{i,t}$ not only accommodates features such as random level shifts and deterministic trends in the log-variance, but also a broad array of finite activity Lévy specifications, which have been shown to improve the volatility model fit for, e.g., option pricing, see Bates (2000) and Andersen et al. (2015).

Next, let us combine the insights from (18) with those from an arbitrage pricing theory (APT) factor model in the spirit of Ross (1976), which suggests writing the returns on asset i as $r_{i,t} = \alpha_i + \lambda_i r_{M,t} + \boldsymbol{\Lambda}_i' \mathbf{f}_t + \eta_{i,t}$, with $r_{M,t}$ the excess return on the market portfolio, $\mathbf{f}_t$ a $k \times 1$ vector of additional pricing factors, $\eta_{i,t}$ the idiosyncratic shock for asset i , and $(\alpha_i, \lambda_i, \boldsymbol{\Lambda}_i)'$ the corresponding $(k+2) \times 1$ parameter vector. Hence, the variance of $r_{i,t}$ is expected to be a function of the factor- and idiosyncratic variances, and their covariance if mutually dependent. This decomposition in terms of factor variances motivates the alternative volatility model¹⁷

$$V_{i,t} = c_i + \beta_i V_{M,t,1} + V_{i,t,2} + \boldsymbol{\iota}' \mathbf{u}_{i,t}, \quad (19)$$

¹⁵This is the case when fitting fractional ARIMA models to the volatility series, e.g., Andersen, Bollerslev, Diebold & Labys (2003), Christensen & Nielsen (2007), Varneskov & Voev (2013), and Andersen & Varneskov (2014), when fitting HAR models, e.g., Corsi (2009), Andersen, Bollerslev & Diebold (2007), and Busch, Christensen & Nielsen (2011), or when fitting discrete time stochastic volatility models with long memory, e.g., Breidt, Crato & de Lima (1998), Hurvich, Moulines & Soulier (2005), and Deo, Hurvich & Lu (2006).

¹⁶See, among others, Gallant, Hsu & Tauchen (1999), Bates (2000), Chernov, Gallant, Ghysels & Tauchen (2003), Andersen et al. (2015), and many references therein.

¹⁷We can add cross-products between the most persistent factor and the low-frequency contaminating terms since their mutual independence is not required to identify β_i . These are left out for simplicity.

where $V_{M,t,1}$ is the most persistent volatility factor of the market portfolio returns, and $V_{i,t,2}$ captures the idiosyncratic volatility for asset i . The parameter β_i is of particular interest, as it measures how much persistent innovations to market volatility disseminate to the volatility of individual equities, that is, how much they are affected by prolonged periods of high, or low, market volatility. Beside the affine dependence on $V_{M,t,1}$, the model (19) shares all the generic features of (18). Moreover, the framework suggests that $V_{i,t,1}$ and $\beta_i V_{M,t,1}$ form a fractionally cointegrating relation, where β_i may be estimated consistently by a linear projection of $V_{i,t}$ on $V_{M,t}$ if the selected estimator is able to accommodate a second volatility factor and low-frequency contamination in both series. While MBLS satisfies this requirement, Theorem 2 shows that NBLS fails to recover β_i , estimating instead a linear combination of low-frequency contaminating terms if these are present in the series, as suggested by an abundance of evidence. Hence, by using existing tools, we risk drawing inference on a completely different, and unknown, risk source.

6.2 Empirical Analysis of Volatility Factors

We analyze MBLS and NBLS estimation of β_i in (19) using a long-span data set of daily observations from 1975 through August 2014 on International Business Machines Corporations (IBM), Coca Cola Company (KO), Walmart Inc. (WMT), and the S&P 500 as a proxy for the market portfolio. From the daily log-returns we construct weekly and monthly (four weeks) realized variance series, then log-transform them, thus generating two different samples of sizes $n = 2068$ and $n = 516$, respectively. Despite being comprised of fewer observations, the monthly realized variance series has the advantage of being less prone to measurement errors. As a result, we will focus on this series throughout, and include the weekly series as a robustness check.

Initially, we gauge the properties of the monthly realized variance series by depicting them and their autocorrelation functions (ACFs) in Figure 2. The series for the three individual stocks are seen to exhibit strong co-movements with the corresponding S&P 500 series, in addition to outliers and prolonged periods of high and low volatility. As emphasized above, however, the latter may be caused by either a persistent long memory component, a combination of structural breaks (or different forms of low-frequency contamination) and short memory dynamics, or both. Hence, it is not surprising that all series display slowly decaying ACFs. We proceed by considering standard descriptive statistics for the four series, provided in Table 5. These demonstrate that the logarithmic transformation has alleviated the large skewness and excess kurtosis usually found in realized variance series in levels. In addition, Table 5 displays the results from applying LW and trimmed LW estimators to the series, using bandwidth configurations $\kappa_1 \in \{0.7, 0.75\}$. The resulting estimates illustrate that each volatility series contains a fractionally integrated component and, for most series, that reliance on the (untrimmed) LW estimator will overstate its persistence due to the presence of low-frequency contamination. In particular, when choosing $\kappa_1 = 0.7$, thereby reducing the impact of short memory dynamics, we find d to be in the interval 0.25-0.35 rather than $d \approx 0.45$, which is usually found in the literature (and is also obtained by the LW estimator). These results corroborate the findings in McCloskey & Perron

(2013) and Varneskov & Perron (2015) for realized variance-style series, and they illustrate the need for a low-frequency contamination-robust estimator of the cointegrating vector.

Next, we estimate the dissemination parameter, β_i , from (19) for each of the three stocks, using both NBLs with a bandwidth $\kappa = 0.6$ and MBLS with the additional trimming parameter, ν . Specifically, we consider three different trimming selections, $\nu \in \{0.35 + \zeta, 0.4 + \zeta, \hat{\nu}_a + \zeta\}$, where $\zeta = 0.005$ and the last $\hat{\nu}_a = (1 - \hat{d}_c)(1 - \kappa)/2 + \kappa/4$ is an adaptive trimming based on $\hat{d}(\ell_1, m_1)$ and $\hat{d}_c(\ell_1, m_1)$ from a first-stage MBLS analysis using $\nu = 0.35 + \zeta$.¹⁸ The latter is motivated by the lower bound restriction in Table 1. Moreover, we set $\kappa_1 = 0.7$ and implement the feasible inference procedure for both MBLS and NBLs, using the global estimate $\hat{d} = p^{-1} \sum_{i=1}^p \hat{d}_i$ for each bivariate pairing with the S&P 500 log-realized variance series. Whereas feasible inference for the MBLS estimates is carried out as described in Section 4.1, the corresponding procedure for NBLs is implemented without any trimming of frequencies in the vicinity of the origin, i.e., with $\nu = \nu_1 = \nu_2 = 0$. Similarly, the tests of cointegration strength and consistency of NBLs inference are also implemented using global estimates of d and without trimming for NBLs. To avoid problems with multiple testing across a large set of values for θ when implementing the test of cointegration strength, t_θ , we utilize the empirically estimated cointegration strength, $\hat{d} - \hat{d}_c$, and the asymptotic distribution from Corollary 1 to back out the upper (right-tail) critical values of θ at the 10%, 5%, and 1% significance levels. A critical value greater than zero indicates fractional cointegration at the associated significance level. The results of this exercise are reported in Tables 6 and 7 for the monthly and weekly series, respectively.

Table 6 contains several interesting results. First, the MBLS estimates of β_i are uniformly above the corresponding NBLs estimates for all three equities, and by more than two standard errors, judging by the MBLS inference.¹⁹ The presence of low-frequency contamination introduces a downward bias in the inconsistent NBLs estimator, making it significantly understate the magnitude of market volatility spillover into the individual equities. This may have dramatic consequences for an investor holding a portfolio of individual equities and seeking to hedge its volatility exposure against persistent movements in market volatility. In this case, she will not only be under-hedged, but also base her investment decision on the coherence between different and unspecified risk sources, the low-frequency contamination, see Theorem 2. Of course, the hedging of such risk sources may be of separate interest. However, for an investor expecting to be protected against persistent movements in market volatility, failure to hedge such consistently may lead to deterioration of portfolio performance. As an analogy, we may think of equity option surfaces whose skew, or strike dimension, and term structure, or maturity dimension, reflect different risk sources. The hedging of term structure risk using skew positions would, then, obviously leave term structure exposure open, and vice versa.

Second, we find substantial differences in the estimates of the residual integration order and, as a consequence, the estimated cointegration strength. Whereas the latter is large and positive based

¹⁸Empirically, we suggest restricting $0 \leq \hat{d}_c \leq 1/2$ when implementing the adaptive trimming procedure. Alternatively, one may define an adaptive procedure that also takes the cross-product trimming condition for MBLS in (9) into account. Such a procedure yields results similar to those presented below and is, thus, omitted for brevity.

¹⁹We have also run OLS regressions. The results are highly similar to those for NBLs and omitted for brevity.

on MBLS, indicating the presence of cointegration, it is small and close to zero when using NBLs, suggesting the opposite conclusion. Such differences are exactly to be expected if there are low-frequency contamination in series, see Table 4. The estimates also show that $\hat{d} + \hat{d}_\epsilon < 1/2$ is violated for NBLs, which is the reason for not reporting standard errors with the NBLs estimates of β_i . Third, the critical values of θ , backed out from the corresponding empirical estimates, corroborate the claim of cointegration based on the MBLS results, as they are positive at the 5% significance level for IBM and KO, and despite falling slightly below 0 at a 10% level for WMT, the point estimates of θ are well above 0. Failure to formally detect significance in the latter case may be explained by the test having only moderate power in smaller samples, see Table 4. Fourth, we strongly reject consistency of the NBLs inference procedure at all conventional significance levels, the adjusted test $t_d^a(c)$ showing significance at a 5% level in all cases. Finally, when decomposing the source of NBLs inconsistency using the trimmed LW estimator to estimate the respective fractional integration orders for $\tilde{t}_d^a(c)$, as described in Corollary 3 and Section 5.3, we still reject consistency of the NBLs point estimates in many cases, despite the procedure lacking power in smaller samples.²⁰ In fact, for the corresponding tests using the weekly realized variance series, we are able to formally reject for all equities.

In sum, we find compelling evidence of fractional cointegration between the realized variance of individual equities and the S&P 500 using the MBLS approach. Furthermore, we show that this is not detectable using the existing NBLs inference procedure, where point estimates are downward biased and residual-based tests fail to detect cointegration. In contrast, our MBLS inference suggests that the dissemination of market volatility to individual equity volatility is significantly stronger than suggested by the inconsistent NBLs estimator. Hence, our analysis of the proposed volatility model (19) provides new perspectives on the (co-)persistence of volatility factors, which may have important implications for risk management. Moreover, these are only detectable using a low-frequency robust approach to fractional cointegration such as MBLS. Finally, in Table 7, we obtain similar results for the weekly realized variance series, adding robustness to the conclusions. The most noteworthy discrepancies between the results for the weekly and monthly series are that the fractional integration order seems to differ slightly more across assets in the weekly case, and that here, the memory condition for inference, $\hat{d} + \hat{d}_\epsilon < 1/2$, is violated for IBM and KO. Hence, no standard errors are reported. However, this has no impact on consistency of MBLS. Furthermore, based on the point estimates and the tests, all qualitative conclusions pertain to the weekly series, as well.

7 Conclusion

This paper introduces the medium band least squares (MBLS) estimator of the cointegrating vector between stationary long memory processes. The estimator is robust to low-frequency contamination such as level shifts and certain deterministic trends. It uses sample-size-dependent trimming of frequencies

²⁰Strictly speaking, we implement $\tilde{t}_d^a(c)$ with $\hat{d}_{\epsilon,N}(c\ell_1, c)$ to ensure that $cm_1 > c\ell_1$. The critical values for this procedure have also been simulated. They are $\{3.57, 3.34\}$ for a 10% significance level, $c \in \{0.1, 0.2\}$, and $n = 500$, and similarly $\{2.82, 3.39\}$ for $n = 2000$. The power of the tests resemble that of $t_d^a(c)$ for **DGP 4 - DT** in Table 4.

in the vicinity of the origin to asymptotically eliminate the impact from low-frequency contamination. We show that the MBLS estimator is consistent and enjoys an asymptotic normal distribution mirroring that achieved by the narrow band least squares (NBLS) estimator in the absence of contamination. Furthermore, we show that NBLS is inconsistent in the presence of low-frequency contamination, measuring a linear “cointegrating” relation among contaminating terms instead of the fractional signals in the limit. Finally, we develop a rigorous feasible inference procedure and two empirical diagnostic tools; a test of cointegration strength and a test of inconsistency of NBLS inference.

In addition to corroborating our theoretical findings in a simulation exercise, we illustrate the use of the MBLS approach by analyzing volatility factors for S&P 500 equities using a long-span data set of realized variance estimates. We show that *only* by using the low-frequency contamination robust MBLS approach can we uncover a fractional cointegrating relation between the most persistent volatility factor in market volatility and individual equity volatility. Moreover, the MBLS estimator is required in order to adequately measure the resulting degree of volatility dissemination, or spillover, from market volatility into individual equity volatility. The use of traditional estimators in this empirical example, such as OLS or NBLS, leads to inconsistent inference.

Parameter Restrictions: Consistency				
	Specification	Lower Bound	Upper Bound	Cross Estimator
MBLS bandwidth	$m = n^\kappa$	$0 < \nu < \kappa$	$\kappa < 1$	-
MBLS trimming	$\ell = n^\nu$	$\frac{(1-\kappa)(1-2d)}{2} < \nu$	$\nu < \kappa$	-
LW' bandwidth	$m_1 = n^{\kappa_1}$	$0 < \nu_1 < \kappa_1$	$\kappa_1 < 1$	$\kappa < \kappa_1$
LW' trimming	$\ell_1 = n^{\nu_1}$	$\max\left\{0, 1 - \frac{\kappa_1}{1-2d}\right\} < \nu_1$	$\nu_1 < \kappa_1$	-
$\mathbf{G}_{xx}$ bandwidth	$m_2 = n^{\kappa_2}$	$0 < \nu_2 < \kappa_2$	$\kappa_2 < 1$	-
$\mathbf{G}_{xx}$ trimming	$\ell_2 = n^{\nu_2}$	$\max\{0, f_G(d, \kappa_2)\} < \nu_2$	$\nu_2 < \kappa_2$	-
$G_{p,p}$ bandwidth	$m_2 = n^{\kappa_2}$	$0 < \nu_2 < \kappa_2$	$\kappa_2 < 1$	-
$G_{p,p}$ trimming	$\ell_2 = n^{\nu_2}$	$\max\{0, f_G(d_\epsilon, \kappa_2)\} < \nu_2$	$\nu_2 < \kappa_2$	-
Parameter Restrictions: Asymptotic Normality				
	Specification	Lower Bound	Upper Bound	Cross Estimator
MBLS bandwidth	$m = n^\kappa$	$0 < \nu < \kappa$	$\kappa < 2\varphi/(1+2\varphi)$	-
MBLS trimming	$\ell = n^\nu$	$\frac{(1-d_c)(1-\kappa)}{2} + \frac{\kappa}{4} < \nu$	$\nu < \min\left\{\frac{\varphi+\kappa(1/2+d_c)}{1+d_c+\varphi}, \kappa\right\}$	-
MBLS CP trimming	$\ell = n^\nu$	$\frac{1/2-\kappa(1/2-d)-(1-\kappa)d_\epsilon}{1+d} < \nu$	-	-
LW' bandwidth	$m_1 = n^{\kappa_1}$	$0 < \nu_1 < \kappa_1$	$\kappa_1 < 2\varphi/(1+2\varphi)$	$\kappa < \kappa_1$
LW' trimming	$\ell_1 = n^{\nu_1}$	$\max\left\{0, 1 - \frac{\kappa_1}{2(1-2d)}\right\} < \nu_1$	$\nu_1 < \kappa_1$	-

Table 1: Parameter restrictions for a full MBLS analysis. This table provides an overview of the parameter restrictions required on the bandwidths and trimming functions for consistency, asymptotic normality, and feasible inference on the cointegrating vector, β , all robust to low-frequency contamination. Here, the conditions implied by Assumption LW' are explicated instead of those for Assumption LW since the former are slightly stronger. Moreover, $f_G(d, \kappa_2) = \max\{\frac{1-2d-\kappa_2}{2(1-d)}, (d-1)\kappa_2 + 1/2 - d\}$, with $f_G(d_\epsilon, \kappa_2)$ defined analogously. Recall that φ measures the smoothness of the spectral density matrix of the latent signal, $\mathbf{w}_t^*$, in the vicinity of the origin, see Assumption 2', and $d_c = d + d_\epsilon$ is the collective fractional integration order of the system. Note that Assumption 5-LW' imposes one additional cross estimator restriction on the bandwidth parameters κ and κ_1 for asymptotic normality; $(\kappa - \kappa_1)(d - d_\epsilon - 1/2) < \kappa_1/4$. However, this is left out of the table for ease of exposition since it is never binding for realistic values of κ , κ_1 , d and d_ϵ . For example, in our empirical section, we typically have $\kappa - \kappa_1 = 6/10 - 7/10 = -1/10$, $d = 3/10$, and $d_\epsilon = 0$, leaving the left-hand-side of the inequality approximately of size $1/50$, which is much smaller than the right-hand-side term. Finally, MBLS CP trimming is the additional lower bound condition on ν in (9) to eliminate cross-products between the fractional signals and low-frequency contamination. If the latter are independent, this may be disregarded.

Simulation Results: Random Level Shifts														
$n =$	DGP 1		DGP 2 - RLS			DGP 3 - RLS			DGP 4 - RLS					
	$p = 0$		$p = 2$		$p = 5$	$p = 2$		$p = 5$		$p = 2$		$p = 5$		
	MBLS	NBLS	MBLS	NBLS	MBLS	NBLS	MBLS	NBLS	MBLS	NBLS	MBLS	NBLS	MBLS	NBLS
Bias														
200	-0.0012	-0.0035	0.0007	0.0101	0.0020	-0.0046	-0.0492	-0.1678	-0.1105	-0.3196	-0.0233	-0.0674	-0.0544	-0.1349
500	-0.0023	-0.0006	-0.0016	0.0051	-0.0038	-0.0027	-0.0375	-0.1781	-0.0926	-0.3383	-0.0201	-0.0745	-0.0435	-0.1237
2000	-0.0017	-0.0003	-0.0010	0.0078	-0.0036	0.0109	-0.0312	-0.1981	-0.0704	-0.3557	-0.0159	-0.0833	-0.0377	-0.1218
RMSE														
200	0.1283	0.0979	0.1434	0.2193	0.1567	0.3107	0.1498	0.2529	0.1901	0.3726	0.1663	0.3306	0.1985	0.4027
500	0.0831	0.0630	0.0930	0.1955	0.1047	0.3020	0.1019	0.2550	0.1403	0.3836	0.1093	0.3174	0.1528	0.4300
2000	0.0431	0.0336	0.0486	0.1392	0.0552	0.2075	0.0645	0.2699	0.1014	0.3974	0.0675	0.3291	0.1051	0.4073

Table 2: Simulations for RLS contamination. This table provides bias and RMSE results for the MBLS and NBLS estimators. Whereas both estimators are implemented with bandwidth $m = n^\kappa$, $\kappa = 0.6$, MBLS uses trimming $\ell = n^\nu$, $\nu = 0.35 + \zeta$ with $\zeta = 0.005$, and NBLS correspondingly uses $\ell = 1$. **DGP 1** is the case of no low-frequency contamination where x_t^* is simulated as a fractional noise process with $d = 0.3$ and $i.i.d.N(0, 1)$ innovations, and $y_t^* = \beta x_t^* + \epsilon_t^*$ with $\beta = 0.7$ and ϵ_t^* similarly being $i.i.d.N(0, 1)$ innovations. **DGP 2 - RLS** augments **DGP 1** with random level shifts in y_t , that is, where $y_t = y_t^* + u_{1,t}$ and $u_{1,t}$ is a random level shift process $u_{1,t} = u_{1,t-1} + \pi_{n,t,1}\eta_{t,1}$ with $\pi_{n,t,1}$ being $i.i.d.Bernoulli(p_1/n)$, $p_1 = p$, and $\eta_{t,1}$ is $i.i.d.N(0, 1)$. **DGP 3 - RLS** augments **DGP 1** with random level shifts in x_t , that is, where $x_t = x_t^* + u_{2,t}$ and $u_{2,t}$ is a random level shift process $u_{2,t} = u_{2,t-1} + \pi_{n,t,2}\eta_{t,2}$ with $\pi_{n,t,2}$ being $i.i.d.Bernoulli(p_2/n)$, $p_2 = p$, and $\eta_{t,2}$ is $i.i.d.N(0, 1)$. **DGP 4 - RLS** augments **DGP 1** with individual random level shifts in both y_t and x_t as well as common shifts, similar to the case considered in Example 1, where the individual and common random level shift processes are simulated as for **DGP 2 - RLS** and **DGP 3 - RLS**. Finally, the results are provided for three sample sizes $n \in \{200, 500, 2000\}$ and two random level shift frequencies $p \in \{2, 5\}$, using 1000 replications.

Simulation Results: Deterministic Trend and Markov Switching									
	DGP 2 - DT			DGP 3 - DT			DGP 4 - DT		
$n =$	200	500	2000	200	500	2000	200	500	2000
Bias									
MBLS	-0.0015	-0.0021	-0.0028	-0.0896	-0.0731	-0.0514	-0.0452	-0.0371	-0.0286
NBLS	0.0055	0.0037	-0.0010	-0.3533	-0.3602	-0.3824	0.1047	0.1220	0.1427
RMSE									
MBLS	0.1428	0.0946	0.0477	0.1557	0.1115	0.0672	0.1410	0.0986	0.0547
NBLS	0.2346	0.2014	0.1778	0.3629	0.3651	0.3840	0.1309	0.1329	0.1454
	DGP 2 - MS			DGP 3 - MS			DGP 4 - MS		
$n =$	200	500	2000	200	500	2000	200	500	2000
Bias									
MBLS	0.0004	-0.0000	-0.0016	-0.1391	-0.1113	-0.0823	-0.0578	-0.0479	-0.0396
NBLS	-0.0018	-0.0013	-0.0027	-0.2124	-0.2083	-0.2252	-0.0795	-0.0733	-0.0820
RMSE									
MBLS	0.1532	0.1004	0.0531	0.1876	0.1399	0.0951	0.1481	0.1053	0.0678
NBLS	0.1621	0.1405	0.1088	0.2371	0.2249	0.2356	0.1709	0.1607	0.1573

Table 3: Simulations for deterministic trend (DT) and Markov switching (MS). This table provides bias and RMSE results for the MBLS and NBLS estimators. Whereas both estimators are implemented with bandwidth $m = n^\kappa$, $\kappa = 0.6$, MBLS uses trimming $\ell = n^\nu$, $\nu = 0.35 + \zeta$ with $\zeta = 0.005$, and NBLS correspondingly uses $\ell = 1$. The latent fractional signals are simulated using **DGP 1**, see also Table 2. **DGP 2 - DT** augments **DGP 1** with a Lipschitz continuous trend plus noise in y_t , that is, where $y_t = y_t^* + u_{1,t}$ and $u_{1,t} = d_t + \eta_{t,1}$ with $d_t = \sin(4\pi t/n)$ and $\eta_{t,1}$ is i.i.d.N(0, 1/2). **DGP 2 - DT** augments **DGP 1** with a Lipschitz continuous trend plus noise in x_t , that is, where $x_t = x_t^* + u_{2,t}$ and $u_{2,t} = d_t + \eta_{t,2}$ with $d_t = \sin(4\pi t/n)$ and $\eta_{t,2}$ is i.i.d.N(0, 1/2). **DGP 4 - DT** augments **DGP 1** with a Lipschitz continuous trend plus noise in both y_t and x_t as well as common noise, that is where $u_{j,t} = d_t + \eta_{t,j} + \eta_{t,3}$ for $j = \{1, 2\}$ where $\eta_{t,3}$ is i.i.d.N(0, 1/2). Similarly, for Markov switching and $j = \{1, 2, 3\}$, $\eta_{j,t,0}$ is i.i.d.N(1/2, 1/2), $\eta_{j,t,1}$ is i.i.d.N(-1/2, 1/2), and the state variables $s_{j,t} \in \{0, 1\}$ all have switching probabilities $p_{01,n} = p_{10,n} = p/n$ where $p = 4$. Moreover, let $u_{j,t} = \eta_{j,t,0}\mathbf{1}\{s_{j,t} = 0\} + \eta_{j,t,1}\mathbf{1}\{s_{j,t} = 1\}$. **DGP 2 - MS** augments **DGP 1** with a Markov switching mean for y_t , that is, where $y_t = y_t^* + u_{1,t}$. **DGP 3 - MS** augments **DGP 1** with a Markov switching mean for x_t , that is, where $x_t = x_t^* + u_{2,t}$. **DGP 4 - MS** augments **DGP 1** with both individual and common Markov switching in y_t and x_t , that is, where $y_t = y_t^* + u_{1,t} + u_{3,t}$ and $x_t = x_t^* + u_{2,t} + u_{3,t}$. Finally, the results are provided for three sample sizes $n \in \{200, 500, 2000\}$ and 1000 replications.

Figure 1: Empirical distribution of $t_d^a(c)$. We depict the empirical distribution (line) of $t_d^a(c)$ for **DGP 1** and sample sizes $n \in \{500, 2000\}$, when implemented with $c \in \{0.1, 0.2\}$. These are supplemented by corresponding normal distributions that have the same respective means and variances (dotted). Details on **DGP 1** are provided in Table 2. Finally, the empirical distributions are based on 1000 replications.

Simulations Results: Empirical Diagnostic Tools										
DGP 1	$n = 500$					$n = 2000$				
	$\hat{d}_{\epsilon,M}$	$\hat{d}_{\epsilon,N}$	$t_d^a(0.1)$	$t_d^a(0.2)$	$t_\theta(0.05)$	$\hat{d}_{\epsilon,M}$	$\hat{d}_{\epsilon,N}$	$t_d^a(0.1)$	$t_d^a(0.2)$	$t_\theta(0.05)$
$\mathcal{RV}_{90}(t_d^a)$	-	-	2.3063	2.1661	-	-	-	1.7269	1.8310	-
$\mathcal{RV}_{95}(t_d^a)$	-	-	2.8428	2.7136	-	-	-	2.3320	2.4266	-
$\mathcal{RV}_{99}(t_d^a)$	-	-	3.9703	4.2175	-	-	-	3.1340	3.5120	-
Bias & RR	-0.0116	-0.0119	0.1900	0.1750	0.5440	-0.0070	-0.0057	0.1090	0.1240	0.7970
Bias & RR	$n = 500$					$n = 2000$				
	$\hat{d}_{\epsilon,M}$	$\hat{d}_{\epsilon,N}$	$t_d^a(0.1)$	$t_d^a(0.2)$	$t_\theta(0.05)$	$\hat{d}_{\epsilon,M}$	$\hat{d}_{\epsilon,N}$	$t_d^a(0.1)$	$t_d^a(0.2)$	$t_\theta(0.05)$
RLS										
DGP 2	0.0567	0.3485	0.6400	0.7440	0.4440	0.0415	0.3635	0.9570	0.9630	0.6650
DGP 3	0.0148	0.1482	0.2960	0.3230	0.5280	0.0156	0.1744	0.7490	0.7620	0.7590
DGP 4	0.0818	0.3795	0.5460	0.7260	0.4660	0.0629	0.3898	0.9780	0.9870	0.6840
DT										
DGP 2	-0.0075	0.3624	0.9990	0.9790	0.5160	-0.0064	0.3612	1.0000	1.0000	0.7860
DGP 3	-0.0070	0.1683	0.5210	0.4750	0.5010	-0.0060	0.1948	0.9480	0.9160	0.7520
DGP 4	-0.0037	0.0432	0.1580	0.1340	0.4370	-0.0066	0.0662	0.4960	0.3870	0.6650
MS										
DGP 2	0.0050	0.2202	0.4440	0.5410	0.5140	0.0138	0.2570	0.9310	0.9490	0.7350
DGP 3	0.0025	0.1130	0.2400	0.2690	0.4960	0.0053	0.1530	0.7930	0.7890	0.7120
DGP 4	0.0150	0.2507	0.4300	0.5830	0.4500	0.0216	0.2840	0.9640	0.9870	0.6160

Table 4: Finite sample results for empirical diagnostics. This table provides various summary statistics for the proposed empirical diagnostic tools across **DGP 1** and all three forms of low-frequency contamination considered, see Tables 2-3 for details. In particular, these are the bias of $\hat{d}_{\epsilon,M}$ and $\hat{d}_{\epsilon,N} \equiv \hat{d}_{\epsilon,N}(1, 1)$; the empirical rejection rates (RRs) of t_θ with $\theta = 0.05$; and the simulated upper quantiles of $t_d^a(c)$, $c \in \{0.1, 0.2\}$, which are denoted by $\mathcal{RV}_{(1-\alpha)\%}(t_d^a)$ for $\alpha \in \{0.1, 0.05, 0.01\}$. Moreover, it shows the raw RRs of $t_d^a(c)$ for **DGP 1**, i.e., its size, and the adjusted RRs for the alternative DGPs with low-frequency contamination, that is, its size-adjusted power. Note that the random level shift specification in **DGP 2** is implemented with $p = 5$. Finally, the results are provided for sample sizes $n \in \{500, 2000\}$ and 1000 replications.

Figure 2: Realized variance series and ACFs. We depict the monthly log-realized variance series for S&P 500 alone and together with the corresponding series for the three equities IBM, KO, and WMT, using data from 1975 through August 2014, amounting to $n = 516$ observations. Furthermore, we illustrate the respective empirical autocorrelation functions (ACFs) for the first 250 lags.

Full Sample Summary Statistics								
	Mean	Std. Dev.	Skewness	Ekurtosis	$\hat{d}(0, 0.7)$	$\hat{d}(\nu_1, 0.7)$	$\hat{d}(0, 0.75)$	$\hat{d}(\nu_1, 0.75)$
Monthly RV								
IBM	-3.0558	0.9489	1.3746	5.6017	0.4660 (0.0563)	0.3297 (0.0662)	0.4009 (0.0481)	0.2001 (0.0539)
KO	-3.2228	0.8698	0.4843	2.5914	0.4768 (0.0563)	0.3983 (0.0662)	0.4570 (0.0481)	0.3789 (0.0539)
WMT	-2.7493	1.0603	1.0988	4.2249	0.4588 (0.0563)	0.2552 (0.0662)	0.4471 (0.0481)	0.3200 (0.0539)
S&P 500	-4.0114	0.8587	0.8332	1.6812	0.4636 (0.0563)	0.3760 (0.0662)	0.4898 (0.0481)	0.5159 (0.0539)
Weekly RV								
IBM	-3.3095	1.1389	0.3138	2.6419	0.4178 (0.0346)	0.3086 (0.0392)	0.3879 (0.0286)	0.2830 (0.0310)
KO	-3.4729	1.1268	-0.0398	0.9382	0.4442 (0.0346)	0.4108 (0.0392)	0.4051 (0.0286)	0.3242 (0.0310)
WMT	-3.0378	1.2301	0.0983	4.0948	0.4255 (0.0346)	0.2524 (0.0392)	0.3742 (0.0286)	0.1957 (0.0310)
S&P 500	-4.2070	1.0808	0.1874	1.1449	0.4947 (0.0346)	0.5076 (0.0392)	0.4644 (0.0286)	0.4238 (0.0310)

Table 5: Full sample summary statistics for RV measures. We report summary statistics for the full samples of monthly and weekly log-realized variance estimates of sizes $n = 516$ and $n = 2068$, respectively. Here, “Ekurtosis” measures the excess kurtosis relative to three. Furthermore, $\hat{d}(0, 0.7)$ and $\hat{d}(\nu_1, 0.7)$ indicate the LW and trimmed LW estimator, respectively, which are implemented with a bandwidth $m = n^{\kappa_1}$, for $\kappa_1 = 0.7$, and for the latter with a trimming function $\ell_1 = n^{\nu_1}$ for $\nu_1 = 1/2 + \zeta$, $\zeta = 0.005$.

Parameter Estimates for MBLS and NBLS: Monthly RV											
	IBM, $\nu =$			KO, $\nu =$			WMT, $\nu =$				
	0.35 + ζ	0.4 + ζ	$\hat{\nu}_a + \zeta$	NLBS	0.35 + ζ	0.4 + ζ	$\hat{\nu}_a + \zeta$	NBLS	0.35 + ζ	0.4 + ζ	$\hat{\nu}_a + \zeta$
$\hat{\beta}(\ell, m)$	0.6843 (0.0500)	0.6905 (0.0497)	0.7049 (0.0489)	0.5711 (-)	0.7806 (0.0275)	0.7028 (0.0294)	0.7653 (0.0277)	0.5970 (-)	0.8106 (0.1211)	0.7279 (0.1158)	0.7547 (0.1169)
$\hat{d}_e(\ell_1, m_1)$	0.0462 (0.0662)	0.0445 (0.0662)	0.0409 (0.0662)	0.4377 (0.0563)	-0.0504 (0.0662)	-0.0250 (0.0662)	-0.0468 (0.0662)	0.4280 (0.0563)	0.1391 (0.0662)	0.1366 (0.0662)	0.1370 (0.0662)
$\hat{\theta}(\ell_1, m_1)$	0.3066	0.3083	0.3119	0.0271	0.4376	0.4121	0.4340	0.0422	0.1765	0.1790	0.1786
$\mathcal{CV}_{90}^-(t_\theta)$	0.1177	0.1193	0.1230	-	0.2486	0.2232	0.2450	-	-0.0125	-0.0100	-0.0103
$\mathcal{CV}_{95}^-(t_\theta)$	0.0641	0.0658	0.0694	-	0.1951	0.1696	0.1914	-	-0.0660	-0.0635	-0.0639
$\mathcal{CV}_{99}^-(t_\theta)$	-0.0364	-0.0347	-0.0311	-	0.0946	0.0691	0.0909	-	-0.1665	-0.1640	-0.1644
$t_d^a(0.1)$	3.179**	3.187**	3.205**	-	4.823***	4.700***	4.805***	-	2.846**	2.857**	2.856**
$t_d^a(0.2)$	4.425***	4.436***	4.460***	-	4.615***	4.448***	4.591***	-	3.071**	3.087**	3.084**
$\tilde{t}_d^a(0.1)$	3.912*	3.400	3.417	-	4.544**	4.421**	4.526**	-	3.296	3.308	3.306
$\tilde{t}_d^a(0.2)$	3.489*	3.500*	3.524*	-	0.329	0.1632	0.306	-	2.525	2.541	2.538

Table 6: Fractional cointegration analysis using MBLS and NBLS. We report full sample MBLS estimates for a bandwidth $m = n^\kappa$, $\kappa = 0.6$, along with trimming function $\ell = n^\nu$ for three different values $\nu \in \{0.35 + \zeta, 0.4 + \zeta, \hat{\nu}_a + \zeta\}$ where $\zeta = 0.005$ and $\hat{\nu}_a = (1 - \hat{d}_c)(1 - \kappa)/2 + \kappa/4$ is an adaptive trimming based on $\hat{d}(\ell_1, m_1)$ and $\hat{d}_c(\ell_1, m_1)$ from a first-stage MBLS analysis using $\nu = 0.35 + \zeta$. The (log-)realized variance series are monthly (20 days) from January 1975 through August 2014, resulting in $n = 516$ observations for each series. The standard errors are based on Theorems 3-5 and Lemma 2, which are implemented using bandwidths $m_1 = n^{\kappa_1}$ and $m_2 = n^{\kappa_2}$ along with trimming functions $\ell_1 = n^{\nu_1}$ and $\ell_2 = n^{\nu_2}$ where $\kappa_1 = \kappa_2 = 0.7$ and $\nu_1 = \nu_2 = 0.5 + \zeta$. Note that the estimated fractional integration order, $\hat{d}(\ell_1, m_1)$, is computed as the average from both the left- and right-side variables. The NBLS estimate and inference procedure are implemented with the same bandwidths along with $\nu = \nu_1 = \nu_2 = 0$. The estimated fractional integration order of the residuals, $\hat{d}_e(\ell_1, m_1)$, is implemented with $\kappa_1 = 0.7$ along with $\nu_1 = 0.5 + \zeta$ for MBLS and $\nu_1 = 0$ for NBLS. The estimated cointegration strength is defined as $\theta(\ell_1, m_1) = \hat{d}(\ell_1, m_1) - \hat{d}_e(\ell_1, m_1)$. The values $\mathcal{CV}_{(1-\alpha)\%}^-(t_\theta)$ for $\alpha \in \{0.1, 0.05, 0.01\}$ are the upper critical values of θ using a significance level $1 - \alpha$ implied by the test for cointegration strength in Corollary 1. Furthermore, $t_d^a(c)$, for $c \in \{0.1, 0.2\}$, denotes the adjusted Hausman-type t -test for consistency of NBLS inference, see Section 5.3, based on the estimates $\hat{d}_e(\ell_1, m_1)$ and the simulated critical values from Table 4. The test $\tilde{t}_d^a(c)$ is similar, but with the fractional integration order of the residuals estimated using the trimmed LW with $\nu_1 = 0.5 + \zeta$ for both MBLS and NBLS, but with ℓ_1 scaled by c for NBLS. The critical values for $\tilde{t}_d^a(c)$ are simulated as for $t_d^a(c)$, see the main text for details. (*), (**), and (***) denote rejection at a 10%, 5% and 1% significance level, respectively.

Parameter Estimates for MBLS and NBLS: Weekly RV												
	IBM			KO			WMT					
	0.35 + ζ	0.4 + ζ	$\hat{\nu}_a + \zeta$	NBLS	0.35 + ζ	0.4 + ζ	$\hat{\nu}_a + \zeta$	NBLS	0.35 + ζ			
$\hat{\beta}(\ell, m)$	0.6523 (-)	0.6431 (-)	0.6815 (-)	0.5873 (-)	0.6456 (-)	0.6353 (-)	0.6950 (-)	0.5900 (-)	0.5910 (0.0374)	0.5705 (0.0380)	0.6556 (0.0362)	0.4528 (-)
$\hat{d}_e(\ell_1, m_1)$	0.1154 (0.0392)	0.1160 (0.0392)	0.1143 (0.0392)	0.3639 (0.0346)	0.2186 (0.0392)	0.2200 (0.0392)	0.2129 (0.0392)	0.3467 (0.0346)	0.0324 (0.0392)	0.0358 (0.0392)	0.0252 (0.0392)	0.3852 (0.0346)
$\hat{\theta}(\ell_1, m_1)$	0.2927	0.2921	0.2938	0.0923	0.2407	0.2392	0.2464	0.1228	0.3476	0.3442	0.3548	0.0749
$\mathcal{CV}_{90}^-(t_\theta)$	0.1605	0.1599	0.1617	-	0.1085	0.1070	0.1142	-	0.2154	0.2120	0.2226	-
$\mathcal{CV}_{95}^-(t_\theta)$	0.1231	0.1225	0.1242	-	0.0710	0.0695	0.0767	-	0.1779	0.1746	0.1852	-
$\mathcal{CV}_{99}^-(t_\theta)$	0.0528	0.0522	0.0539	-	0.0007	-0.0007	0.0064	-	0.1076	0.1043	0.1149	-
$t_d^a(0.1)$	5.611***	5.606***	5.620***	-	3.649***	3.637***	3.697***	-	5.248***	5.220***	5.309***	-
$t_d^a(0.2)$	5.693***	5.687***	5.706***	-	4.507***	4.490***	4.572***	-	8.425***	8.387***	8.580***	-
$\tilde{t}_d^a(0.1)$	5.977***	5.972***	5.987***	-	1.014	1.002	1.062	-	5.011**	4.982**	5.072**	-
$\tilde{t}_d^a(0.2)$	2.916	2.910	2.929	-	4.848**	4.831**	4.912**	-	10.96***	10.92***	11.05***	-

Table 7: Fractional cointegration analysis using MBLS and NBLS. We report full sample MBLS estimates for a bandwidth $m = n^\kappa$, $\kappa = 0.6$, along with trimming function $\ell = n^\nu$ for three different values $\nu \in \{0.35 + \zeta, 0.4 + \zeta, \hat{\nu}_a + \zeta\}$ where $\zeta = 0.005$ and $\hat{\nu}_a = (1 - \hat{d}_c)(1 - \kappa)/2 + \kappa/4$ is an adaptive trimming based on $\hat{d}(\ell_1, m_1)$ and $\hat{d}_c(\ell_1, m_1)$ from a first-stage MBLS analysis using $\nu = 0.35 + \zeta$. The (log-)realized variance series are weekly (5 days) from January 1975 through August 2014, resulting in $n = 2068$ observations for each series. The standard errors are based on Theorems 3-5 and Lemma 2, which are implemented using bandwidths $m_1 = n^{\kappa_1}$ and $m_2 = n^{\kappa_2}$ along with trimming functions $\ell_1 = n^{\nu_1}$ and $\ell_2 = n^{\nu_2}$ where $\kappa_1 = \kappa_2 = 0.7$ and $\nu_1 = \nu_2 = 0.5 + \zeta$. Note that the estimated fractional integration order, $\hat{d}(\ell_1, m_1)$, is computed as the average from both the left- and right-side variables. The NBLS estimate and inference procedure are implemented with the same bandwidths along with $\nu = \nu_1 = \nu_2 = 0$. The estimated fractional integration order of the residuals, $\hat{d}_e(\ell_1, m_1)$, is implemented with $\kappa_1 = 0.7$ along with $\nu_1 = 0.5 + \zeta$ for MBLS and $\nu_1 = 0$ for NBLS. The estimated cointegration strength is defined as $\theta(\ell_1, m_1) = \hat{d}(\ell_1, m_1) - \hat{d}_e(\ell_1, m_1)$. The values $\mathcal{CV}_{(1-\alpha)\%}^-(t_\theta)$ for $\alpha \in \{0.1, 0.05, 0.01\}$ are the upper critical values of θ using a significance level $1 - \alpha$ implied by the test for cointegration strength in Corollary 1. Furthermore, $\tilde{t}_d^a(c)$, for $c \in \{0.1, 0.2\}$, denotes the adjusted Hausman-type t -test for consistency of NBLS inference, see Section 5.3, based on the estimates $\hat{d}_e(\ell_1, m_1)$ and the simulated critical values from Table 4. The test $\tilde{t}_d^a(c)$ is similar, but with the fractional integration order of the residuals estimated using the trimmed LW with $\nu_1 = 0.5 + \zeta$ for both MBLS and NBLS, but with ℓ_1 scaled by c for NBLS. The critical values for $\tilde{t}_d^a(c)$ are simulated as for $t_d^a(c)$, see the main text for details. (*), (**), and (***) denote rejection at a 10%, 5% and 1% significance level, respectively.

References

- Aït-Sahalia, Y., Cacho-Diaz, J. & Laeven, R. J. A. (2015), ‘Modeling financial contagion using mutually exciting jump processes’, *Journal of Financial Economics* **117**, 585–606.
- Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), ‘Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility’, *The Review of Economics and Statistics* **89**, 701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), ‘The distribution of realized stock return volatility’, *Journal of Financial Economics* **61**, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), ‘Modeling and forecasting realized volatility’, *Econometrica* **71**, 579–625.
- Andersen, T. G., Fusari, N. & Todorov, V. (2015), ‘Parametric inference and dynamic state recovery from option panels’, *Econometrica* **83**, 1081–1145.
- Andersen, T. G. & Varneskov, R. T. (2014), On the informational efficiency of option-implied and time series forecasts of realized volatility. Unpublished Manuscript, Northwestern University.
- Baillie, R., Bollerslev, T. & Mikkelsen, H. (1996), ‘Fractionally integrated generalized autoregressive conditional heteroskedasticity’, *Journal of Econometrics* **74**, 3–30.
- Baillie, R. T. & Bollerslev, T. (1994), ‘Cointegration, fractional cointegration, and exchange rate dynamics’, *Journal of Finance* **49**, 737–745.
- Baillie, R. T. & Kapetanios, G. (2007), ‘Testing for neglected nonlinearity in long-memory models’, *Journal of Business and Economic Statistics* **25**, 447–461.
- Bandi, F. M. & Perron, B. (2006), ‘Long memory and the relation between implied and realized volatility’, *Journal of Financial Econometrics* **4**(4), 636–670.
- Bates, D. S. (2000), ‘Post-‘87 crash fears in the S&P 500 futures option market’, *Journal of Econometrics* **94**, 181.
- Bhattacharya, R., Gupta, V. & Waymire, E. (1983), ‘The Hurst effect under trends’, *Journal of Applied Probability* **20**, 649–662.
- Bollerslev, T., Osterrieder, D., Sizova, N. & Tauchen, G. (2013), ‘Risk and return: Long-run relationships, fractional cointegration, and return predictability’, *Journal of Financial Economics* **108**(2), 409–424.
- Breidt, F. J., Crato, N. & de Lima, P. (1998), ‘The detection and estimation of long memory in stochastic volatility’, *Journal of Econometrics* **83**, 325–348.

- Brillinger, D. R. (1981), *Time Series. Data Analysis and Theory*, Siam: Classics in Applied Mathematics.
- Busch, T., Christensen, B. J. & Nielsen, M. Ø. (2011), ‘The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets’, *Journal of Econometrics* **160**, 48–57.
- Campbell, J. Y. & Perron, P. (1991), Pitfalls and opportunities: What macroeconomists should know about unit roots, in O. J. Blanchard & S. Fisher, eds, ‘NBER Macroeconomics Annual’, Vol. 6, MIT Press.
- Campos, J., Ericsson, N. R. & Hendry, D. F. (1996), ‘Cointegration tests in the presence of structural breaks’, *Journal of Econometrics* **70**, 187–220.
- Chen, W. W. & Hurvich, C. M. (2003), ‘Estimating fictional cointegration in the presence of polynomial trends’, *Journal of Econometrics* **117**, 95–121.
- Chernov, M., Gallant, R., Ghysels, E. & Tauchen, G. (2003), ‘Alternative models for stock price dynamics’, *Journal of Econometrics* **116**, 225–257.
- Cheung, Y. E. & Lai, K. S. (1993), ‘A fractional cointegration analysis of purchasing power parity’, *Journal of Business and Economic Statistics* **11**, 103–122.
- Christensen, B. J. & Nielsen, M. O. (2006), ‘Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting’, *Journal of Econometrics* **133**, 343–371.
- Christensen, B. J. & Nielsen, M. O. (2007), ‘The effect of long memory in volatility on stock market fluctuations’, *Review of Economics and Statistics* **89**, 684–700.
- Corsi, F. (2009), ‘A simple approximate long-memory model of realized volatility’, *Journal of Financial Econometrics* **7**, 174–196.
- Davidson, J. (2002), *Stochastic Limit Theory*, Oxford University Press.
- Deo, R. S., Hurvich, C. M. & Lu, Y. (2006), ‘Forecasting realized volatility using a long memory stochastic volatility model: Estimation, prediction, and seasonal adjustment’, *Journal of Econometrics* **131**, 29–58.
- Diebold, F. X. & Inoue, A. (2001), ‘Long memory and regime switching’, *Journal of Econometrics* **105**, 131–159.
- Dueker, M. & Startz, R. (1998), ‘Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates’, *The Review of Economics and Statistics* **83**, 420–426.

- Engle, R. F. (1974), ‘Band spectrum regression’, *International Economic Review* **15**, 1–11.
- Engle, R. F. & Granger, C. W. J. (1987), ‘Co-integration and error correction: Representation, estimation, and testing’, *Econometrica* **55**, 251–276.
- Gallant, R., Hsu, C.-T. & Tauchen, G. (1999), ‘Using daily range to calibrate volatility diffusions and extract the forward integrated variance’, *The Review of Economics and Statistics* **81**, 617–631.
- Garcia, R. & Perron, P. (1996), ‘An analysis of the real interest rate under regime shifts’, *The Review of Economics and Statistics* **78**, 111–125.
- Granger, C. W. J. (1981), ‘Some properties of time series data and their use in econometric model specification’, *Journal of Econometrics* **16**, 121–130.
- Granger, C. W. J. & Ding, Z. (1996), ‘Varieties of long memory models’, *Journal of Econometrics* **73**, 61–77.
- Granger, C. W. J. & Hyung, N. (2004), ‘Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns’, *Journal of Empirical Finance* **11**, 399–421.
- Gregory, A. W. & Hansen, B. E. (1996), ‘Residual-based tests for cointegration in models with regime shifts’, *Journal of Econometrics* **70**, 99–126.
- Haldrup, N., Kruse, R., Teräsvirta, T. & Varneskov, R. T. (2013), Unit roots, nonlinearities and structural breaks, in N. Hashimzade & M. Thornton, eds, ‘Handbook of Research Methods and Applications in Empirical Macroeconomics’, Handbook of Research Methods and Application Series, Edward Elgar Publishing Inc., pp. 61–94.
- Haldrup, N. & Nielsen, M. O. (2007), ‘Estimation of fractional integration in the presence of data noise’, *Computational Statistics and Data Analysis* **51**, 3100–3114.
- Hurvich, C. M., Moulines, E. & Soulier, P. (2005), ‘Estimating long memory in volatility’, *Econometrica* **78**, 1283–1328.
- Iacone, F. (2010), ‘Local whittle estimation of the memory parameter in the presence of deterministic components’, *Journal of Time Series Analysis* **31**, 37–49.
- Johansen, S., Mosconi, R. & Nielsen, B. (2000), ‘Cointegration analysis in the presence of structural breaks in the deterministic trend’, *The Econometrics Journal* **3**, 216–249.
- Johansen, S. & Nielsen, M. (2012), ‘Likelihood inference for a fractionally cointegrated vector autoregressive model’, *Econometrica* **80**, 2667–2732.
- Lobato, I. (1997), ‘Consistency of averaged cross-periodogram in long memory series’, *Journal of Time Series Analysis* **18**, 137–155.

- Lobato, I. (1999), ‘A semiparametric two-step estimator in a multivariate long memory model’, *Journal of Econometrics* **90**, 129–155.
- Lu, Y. K. & Perron, P. (2010), ‘Modeling and forecasting stock return volatility using a random level shift model’, *Journal of Empirical Finance* **17**, 138–156.
- McCloskey, A. (2013), ‘Estimation of the long-memory stochastic volatility model parameters that is robust to level shifts and deterministic trends’, *Journal of Time Series Analysis* **34**, 285–301.
- McCloskey, A. & Hill, J. B. (2015), ‘Parameter estimation robust to low-frequency contamination’, *Journal of Business and Economic Statistics* **forthcoming**.
- McCloskey, A. & Perron, P. (2013), ‘Memory parameter estimation in the presence of level shifts and deterministic trends’, *Econometric Theory* **29**, 1196–1237.
- Mikosch, T. & Stărică, C. (2004), ‘Nonstationarities in financial time series, the long range dependence, and the IGARCH effects’, *Review of Economics and Statistics* **86**, 378–390.
- Nielsen, M. O. & Frederiksen, P. (2011), ‘Fully modified narrow-band least squares estimation of weak fractional cointegration’, *The Econometrics Journal* **14**, 77–120.
- Nielsen, M. O. & Shimotsu, K. (2007), ‘Determining the cointegrating rank in nonstationary fractional systems by the exact local whittle approach’, *Journal of Econometrics* **141**, 574–596.
- Ohanissian, A., Russell, J. R. & Tsay, R. S. (2008), ‘True or spurious long memory? A new test’, *Journal of Business and Economic Statistics* **26**, 161–175.
- Perron, P. (1989), ‘The great crash, the oil price shock and the unit root hypothesis’, *Econometrica* **57**, 1361–1401.
- Perron, P. & Qu, Z. (2010), ‘Long memory and level shifts in the volatility of stock market return indices’, *Journal of Business and Economic Statistics* **28**, 275–290.
- Qu, Z. (2011), ‘A test against spurious long memory’, *Journal of Business and Economic Statistics* **29**, 423–438.
- Robinson, P. M. (1994), ‘Semiparametric analysis of long-memory time series’, *The Annals of Statistics* **22**, 515–539.
- Robinson, P. M. (1995a), ‘Gaussian semiparametric estimation of long range dependence’, *The Annals of Statistics* **23**, 1630–1661.
- Robinson, P. M. (1995b), ‘Log-periodogram regression of time series with long range dependence’, *The Annals of Statistics* **23**, 1048–1072.

- Robinson, P. M. (1997), ‘Large-sample inference for nonparametric regression with dependent errors’, *The Annals of Statistics* **25**, 2054–2083.
- Robinson, P. M. (2005), ‘The distance between rival nonstationary fractional processes’, *Journal of Econometrics* **128**, 283–300.
- Robinson, P. M. (2008), ‘Multiple local Whittle estimation in stationary systems’, *Annals of Statistics* **36**, 2508–2530.
- Robinson, P. M. & Iacone, F. (2005), ‘Cointegration in fractional systems with deterministic trends’, *Journal of Econometrics* **129**, 263–298.
- Robinson, P. M. & Marinucci, D. (2003), ‘Semiparametric frequency domain analysis of fractional cointegration’. In: Robinson, P.M. (Ed.), *Time Series with Long Memory*. Oxford University Press, Oxford, pp. 334–373.
- Robinson, P. M. & Yajima, Y. (2002), ‘Determination of cointegrating rank in fractional systems’, *Journal of Econometrics* **106**, 217–241.
- Ross, S. A. (1976), ‘The arbitrage theory of capital asset pricing’, *Journal of Economic Theory* **13**(3), 341 – 360.
- Shimotsu, K. (2007), ‘Gaussian semiparametric estimation of multivariate fractionally integrated processes’, *Journal of Econometrics* **137**, 277–310.
- Shimotsu, K. & Phillips, P. C. B. (2006), ‘Local whittle estimation of fractional integration and some of its variants’, *Journal of Econometrics* **130**, 209–233.
- Stărică, C. & Granger, C. W. J. (2005), ‘Nonstationarities in stock returns’, *The Review of Economics and Statistics* **87**, 503–522.
- Varneskov, R. T. (2015), Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels. Unpublished Manuscript, Aarhus University.
- Varneskov, R. T. & Perron, P. (2015), Combining long memory and level shifts in modeling and forecasting the volatility of asset returns. Unpublished Manuscript, Boston University.
- Varneskov, R. T. & Voev, V. (2013), ‘The role of realized ex-post covariance measures and dynamic model choice on the quality of covariance forecasts’, *Journal of Empirical Finance* **20**, 83–95.
- Yamamoto, Y. & Perron, P. (2013), ‘Estimating and testing multiple structural changes in linear models using band spectral regressions’, *The Econometrics Journal* **16**, 400–429.

A Examples of Low-Frequency Contamination

To complement Example 1 and illustrate the array of low-frequency contaminating processes that satisfy the multivariate high-level Assumption 1, we continue with the bivariate case $(u_{1,t}, u_{2,t})$, similar to the random level shift specification, and consider general linear processes of the form

$$u_{1,t} = c_{1,1}v_{1,t} + c_{1,2}v_{3,t} \quad \text{and} \quad u_{2,t} = c_{2,1}v_{2,t} + c_{2,2}v_{3,t} \quad (\text{A.1})$$

where $(c_{1,1}, c_{1,2}, c_{2,1}, c_{2,2})$ are finite constants, loading on the individual and common low-frequency contaminating terms $v_{i,t}$ with $i = \{1, 2, 3\}$. Dropping the i subscript for notational simplicity, each univariate v_t may be generated according to the following data generating processes:

Example 2 (Deterministic Level Shifts). *For some fixed positive integer B and deterministic partition of the sample $t_0 = 1, t_B = n, t_0 < t_1 < \dots < t_B, v_t = \sum_{j=1}^B \tilde{c}_j \mathbf{1}\{t_{j-1} < t \leq t_j\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function, $|\tilde{c}_j| < \infty$ and $t_j/n \rightarrow \tau_j \in (0, 1)$ for $j = 1, \dots, B$.*

Example 3 (Deterministic Trends). *$v_t = \tilde{f}(t/n)$, where $\tilde{f}(\cdot)$ is a deterministic non-constant function on $[0, 1]$ that is either Lipschitz continuous or monotone with $\tilde{f}(1) = 0$.*

Example 4 (Fractional Trends). *For some $\tilde{\varphi} \in (-1/2, 1/2)$, $v_t = O((t+1)^{\tilde{\varphi}-1/2})$, and it satisfies the two conditions $v_0 = 0$ and $|v_{t+1} - v_t| = O(|v_t|/t)$.*

Example 5 (Outliers). *For some fixed positive integer C and partition $1 < t_1 < \dots < t_C \leq n$, write the deterministic outliers as $v_t = \sum_{j=1}^C \tilde{g}_j \mathbf{1}\{t = t_j\}$ where $|\tilde{g}_j| < \infty$ for $j = 1, \dots, C$.*

As emphasized in Section 2.2, univariate results demonstrating that the data generating processes in Examples 2-5 satisfy Assumption 1 have already been provided in the extant literature, and we refer to the respective papers for details. Hence, similarly to Lemma 1, the corresponding results for the bivariate low-frequency contamination model in (A.1) and each of the examples above readily follow by the respective univariate results in conjunction with the Cauchy-Schwarz inequality.

A different approach to modeling structural breaks in the levels of the series is to assume that the shifts obey a Markov switching mechanism. Such a data generating process has been considered in the simulation studies of, e.g., Ohanissian et al. (2008) and Qu (2011), without, however, formally analyzing its asymptotic properties. Hence, we complement their work by formally showing that a general class of Markov switching processes also satisfies the high-level Assumption 1:

Example 6 (Markov Switching). *$v_t = \eta_{t,0} \mathbf{1}\{s_t = 0\} + \eta_{t,1} \mathbf{1}\{s_t = 1\}$ for some state variable s_t taking values $\{0, 1\}$, and $\eta_{t,i} \stackrel{\mathbb{D}}{\sim} \text{i.i.d.} \mathcal{N}(\mu_i, \sigma_{\eta,i}^2)$ for $i = 0, 1$ where $\eta_{t,0}$ and $\eta_{t,1}$ are independent. Moreover, the state variable behaves according to the transition matrix*

$$\mathcal{P}_n \equiv \begin{pmatrix} \mathbb{P}[s_t = 0 | s_{t-1} = 0] & \mathbb{P}[s_t = 1 | s_{t-1} = 0] \\ \mathbb{P}[s_t = 0 | s_{t-1} = 1] & \mathbb{P}[s_t = 1 | s_{t-1} = 1] \end{pmatrix} = \begin{pmatrix} p_{00,n} & p_{01,n} \\ p_{10,n} & p_{11,n} \end{pmatrix}$$

where sample size dependent switching probabilities $p_{01,n}$ and $p_{10,n}$ satisfy $p_{01,n} = p_{01}/n$ and $p_{10,n} = p_{10}/n$ with $0 < p_{01}, p_{10} < n$ fixed as well as $p_{00,n} + p_{01,n} = 1$ and $p_{10,n} + p_{11,n} = 1$.

The Markov switching process, v_t , has time t transition probabilities between states that are given by p_{01}/n and p_{10}/n for $s_{t-1} = 0$ and $s_{t-1} = 1$, respectively, which both share the generic form of the random level shift probability in Example 1. Hence, in a given sample, $t = 1, \dots, n$, the process generates a finite number of transitions between states. However, it differs from said example by having a time-varying shift probability and by not requiring $\eta_{t,0}$ and $\eta_{t,1}$ to have zero means.

Lemma 3. *Suppose that v_t follows the data generating process of Example 6, then*

$$\mathbb{E}[(j^2/n)I_{vv}(\lambda_j)] \leq O(1), \quad \forall [(j^2/n)I_{vv}(\lambda_j)] \leq O(1).$$

Proof. See Appendix B.7. □

Markov switching makes up one way of modeling time-varying shift probabilities as well as stochastic levels. A different way of modeling stochastic level changes, or jumps, in the observable process is to assume they are generated according to a more general compound Lévy process:

Example 7 (Compound Lévy Process). $v_t = \sum_{j=1}^{\pi_t} \eta_j$ where $\pi_t \stackrel{\mathbb{D}}{\sim} \text{Poisson}(p_t/n)$ where the time-varying Poisson intensity is Lipschitz continuous and satisfies for all $t = 1, \dots, n$, $0 < p_t < n$. Moreover, p_t and η_t are independent, and the sequence of jump sizes is bounded with $\min_t |\eta_t| \in (0, \infty)$.

Compared with, e.g., Perron & Qu (2010), the compound Lévy Process allows for general forms of time-variation in the jump intensity and avoids distributional assumptions on the jump magnitudes. In fact, the process in Example 7 does not restrict the dependence between the jump intensity and the underlying signal of interest. Hence, this covers cases where the jump intensity is excited, or positively dependent, on the latent process. This is potentially important for our empirical application to asset return volatility factors, as it accommodates linear dependence between the volatility factors and the jump intensity, similarly to the continuous time option pricing models in, e.g., Andersen et al. (2015). Finally, note that the process in Example 7 is often used to model jumps in the high-frequency volatility estimation literature, see, e.g., the recent work Varneskov (2015) and references therein.

Lemma 4. *Suppose that v_t follows the data generating process of Example 7, then*

$$\mathbb{E}[(j^2/n)I_{vv}(\lambda_j)] \leq O(1), \quad \forall [(j^2/n)I_{vv}(\lambda_j)] \leq O(1).$$

Proof. See Appendix B.8. □

Remark 7. *The bivariate model for low-frequency contamination (A.1) accommodates the case where several sources of contamination affect one series. The results for such a process are easily found by the respective univariate results and the Cauchy-Schwarz inequality to bound their potential covariance.*

B Proofs of Asymptotic Results

This section contains the proofs of the main asymptotic results in the paper. Before proceeding, however, let us introduce some notation. For a generic vector $\mathbf{V}$ let $\mathbf{V}(i)$ index the i -th element, and, similarly, let $\mathbf{M}(i, q)$ denote the (i, q) -th element of a matrix $\mathbf{M}$. $\mathbf{F}_{zz}^*(\ell, m)$ and $\mathbf{I}_{zz}^*(\lambda_j)$ are the trimmed discretely averaged co-periodogram and the co-periodogram, respectively, of $\mathbf{z}^*$. Similarly, we use $\hat{\mathbf{F}}_{x\epsilon}^*(\ell, m)$ to denote the vector $\hat{\mathbf{F}}_{x^*\epsilon^*}(\ell, m)$. Moreover, denote by $K \in (0, \infty)$ a generic constant and $K_s \in (0, \infty)$ a sequence of constants indexed by positive integers $s \in \mathbb{N}^+$. These may take different values in different places. Finally, we remark that sometimes the (stochastic) orders refer to scalars, sometimes to vectors and matrices. We refrain from making distinctions.

B.1 Proof of Theorem 1

Theorem 1 (a) is a key intermediate result for establishing consistency of MBLs in (b). Hence, to verify our remarks in the main text stating that consistency holds *without* assuming equality of memory, that is, $d_i = d \in (0, 1/2)$, $i = 1, \dots, p$, we prove (a) under more general conditions. Specifically, we replace Assumption 2 with a corresponding assumption where the spectral representation not only accommodates different integration orders, but also a complex component in the spectral density matrix at the origin, similar to the representations in Shimotsu (2007) and Robinson (2008). The equivalent consistency result will, then, follow by applying (a), the arguments below for (b), and slightly different bandwidth and trimming conditions than those stated for Theorem 1 (b).

Assumption 2-Complex. *The vector process $\mathbf{z}_t^*$, $t = 0, \pm 1, \dots$, is covariance stationary with spectral density matrix satisfying $\mathbf{f}_{zz}^*(\lambda) \sim \mathbf{\Lambda} \mathbf{G} \bar{\mathbf{\Lambda}}$ as $\lambda \rightarrow 0^+$ where $\mathbf{G}$ is a $p \times p$ real symmetric matrix whose lower right $(p-1) \times (p-1)$ submatrix has full rank and $\mathbf{\Lambda} = \text{diag}[\lambda^{-d_1} e^{i\pi d_1/2}, \dots, \lambda^{-d_p} e^{i\pi d_p/2}]$ for $d_i \in (0, 1/2)$, $i = 1, \dots, p$. However, there exists a $(p-1) \times 1$ vector $\boldsymbol{\beta} \neq \mathbf{0}$ and a constant $g \in (0, \infty)$ such that*

$$(1, -\boldsymbol{\beta}') \mathbf{f}_{zz}^*(\lambda) (1, -\boldsymbol{\beta}')' = f_\epsilon(\lambda) \sim g \lambda^{-2d_\epsilon}, \quad \text{as } \lambda \rightarrow 0^+,$$

where $0 \leq d_\epsilon < d_\perp$, $d_\perp = \min_{i=1, \dots, p} (d_i)$. Define also $d_\top = \max_{i=1, \dots, p} (d_i)$.

Remark 8. *Whereas Theorem 1 (a) will be established under Assumptions 1, 2-Complex and 3, Theorem 1 (b) is shown under the stated trimming conditions and with Assumption 2 replacing 2-Complex. Assumption 2 is nested as a special case of the latter when all integration orders are equal.*

First, for (a), to establish a bound for $|\hat{\mathbf{F}}_{zz}(\ell, m) - \mathbf{G}(\lambda_m)|$, where $\mathbf{G}(\lambda_m)$, here in (a), has (i, k) -th element $\mathbf{G}(\lambda_m, i, k) = G_{i,k} \lambda_m^{1-d_i-d_k} \cos(\pi(d_i - d_k)/2) / (1 - d_i - d_k)$, make the decomposition

$$\hat{\mathbf{F}}_{zz}(\ell, m) - \hat{\mathbf{F}}_{zz}^*(1, m) = -\frac{2\pi}{n} \left(\sum_{j=1}^m \Re(\mathbf{I}_{zz}^*(\lambda_j)) - \sum_{j=\ell}^m \Re(\mathbf{I}_{zz}^*(\lambda_j)) \right) + \hat{\mathbf{F}}_{uu}(\ell, m) + 2\hat{\mathbf{F}}_{zu}^*(\ell, m)$$

$$= -\hat{\mathbf{F}}_{zz}^*(1, \ell - 1) + \frac{2\pi}{n} \sum_{j=\ell}^m O_p\left(\frac{n}{j^2}\right) + \frac{4\pi}{n} \sum_{j=\ell}^m O_p\left(\frac{n^{1/2+d_\top}}{j^{1+d_\top}}\right), \quad (\text{B.2})$$

using Assumptions 1, 2-Complex, and the Cauchy-Schwarz inequality to obtain the stochastic orders for the last two terms. Define $\mathcal{E}(\ell) = 2\pi \sum_{j=\ell}^m O_p(j^{-2})$ and $\mathcal{C}(\ell) = 4\pi n^{d_\top-1/2} \sum_{j=\ell}^m O_p(j^{-(1+d_\top)})$ for the low-frequency contamination and cross-product terms. Then, since $2 > 1 + d_\top > 1$, we may invoke Varneskov (2015, Lemma C.4) to show $|\mathcal{E}(\ell)| \leq O_p(\ell^{-2})$ and $|\mathcal{C}(\ell)| \leq O_p(\ell^{-(1+d_\top)} n^{(d_\top-1/2)})$. Next, we seek to connect the two quantities $\hat{\mathbf{F}}_{zz}^*(1, m)$ and $\mathbf{G}(\lambda_m)$ using the continuously averaged spectral density $\mathbf{F}_{zz}^*(\lambda_m)$. Similarly to the steps used by Robinson & Marinucci (2003, p. 361), see also Lobato (1997), this involves establishing stochastic orders for the remaining two terms in (B.2), that is, for

$$\mathbf{\Lambda}_m \mathbf{\Lambda}_m^{-1} \left(\hat{\mathbf{F}}_{zz}^*(1, m) - \mathbf{F}_{zz}^*(\lambda_m) \right) \bar{\mathbf{\Lambda}}_m^{-1} \bar{\mathbf{\Lambda}}_m - \mathbf{\Lambda}_\ell \mathbf{\Lambda}_\ell^{-1} \hat{\mathbf{F}}_{zz}^*(1, \ell - 1) \bar{\mathbf{\Lambda}}_\ell^{-1} \bar{\mathbf{\Lambda}}_\ell \quad (\text{B.3})$$

since Assumption 2-Complex immediately provides

$$\begin{aligned} \mathbf{F}_{zz}^*(\lambda_m, i, k) &\sim G_{i,k} \lambda_m^{1-d_i-d_k} / (1 - d_i - d_k) \Re(e^{i\pi(d_i-d_k)/2}) \\ &\sim G_{i,j} \lambda_m^{1-d_i-d_k} / (1 - d_i - d_k) \cos(\pi(d_i - d_k)/2) = \mathbf{G}(\lambda_m, i, k) \end{aligned} \quad (\text{B.4})$$

as $\lambda_m \rightarrow 0$ when $m/n \rightarrow 0$, $\forall (i, k) \in (1, \dots, p)^2$ elements. For the *first* of the two terms in (B.3), it follows by Assumption 2-Complex, and using $\Re(e^{i\lambda z}) = 1 + O(\lambda^2)$, $\Im(e^{i\lambda z}) = O(\lambda)$ as $\lambda \rightarrow 0^+$ for any $z \in \mathbb{R}$ in conjunction with Robinson (1994, Theorem 1) along with Lobato (1997, Theorem 1), that

$$\mathbf{\Lambda}_m^{-1} (\hat{\mathbf{F}}_{zz}^*(1, m) - \mathbf{F}_{zz}^*(\lambda_m)) \bar{\mathbf{\Lambda}}_m^{-1} = o_p(\lambda_m(1 + \lambda_m + \lambda_m^2)) = o_p(\lambda_m), \quad (\text{B.5})$$

uniformly, as $\lambda_m \rightarrow 0$ when $m/n \rightarrow 0$ since $\Re(e^{i(\pi-\lambda_m)(d_i-d_j)/2}) = \Re(e^{i\pi(d_i-d_j)/2})(1 + \lambda_m + \lambda_m^2)$ induces only an additional *lower* order approximation error for $\hat{\mathbf{F}}_{zz}^*(1, m)$. See also the corresponding result in Robinson & Marinucci (2003, (A.3)). Hence,

$$|\hat{\mathbf{F}}_{zz}^*(1, m) - \mathbf{G}(\lambda_m)| \leq o_p((m/n)^{1-2d_\top}) \quad (\text{B.6})$$

gives the upper bound for the stochastic order of the approximation error. By the same arguments, it follows for the *second* term in (B.3) that

$$\hat{\mathbf{F}}_{zz}^*(1, \ell - 1) = \mathbf{G}(\lambda_\ell) + o_p((\ell/n)^{1-2d_\top}) \leq O_p((\ell/n)^{1-2d_\top}), \quad (\text{B.7})$$

uniformly. When collecting terms, this gives a general uniform asymptotic approximation,

$$|\hat{\mathbf{F}}_{zz}(\ell, m) - \mathbf{G}(\lambda_m)| \leq O_p\left((\ell/n)^{1-2d_\top}\right) + O_p(\ell^{-2}) + O_p\left(\ell^{-(1+d_\top)} n^{d_\top-1/2}\right) + o_p\left((m/n)^{1-2d_\top}\right),$$

providing the final result in (a) with $d_i = d = d_\top$.

Second, for the consistency result of MBLS estimation in (b), we may invoke the asymptotic approximation in (a) to show that if $1/\ell \rightarrow 0$, then $|\hat{\mathbf{F}}_{xx}(\ell, m) - \mathbf{G}_{xx}(\lambda_m)| = o_p(1)$ where $\mathbf{G}_{xx}(\lambda_m)$ is the lower right $(p-1) \times (p-1)$ submatrix of $\mathbf{G}(\lambda_m)$. Next, decompose $\hat{\mathbf{F}}_{xy}(\ell, m)$ as

$$\begin{aligned}\hat{\mathbf{F}}_{xy}(\ell, m) &= \hat{\mathbf{F}}_{xy^*}(\ell, m) + \hat{\mathbf{F}}_{xu_1}(\ell, m) = \hat{\mathbf{F}}_{xx^*}(\ell, m)\boldsymbol{\beta} + \hat{\mathbf{F}}_{x\epsilon^*}(\ell, m) + \hat{\mathbf{F}}_{xu_1}(\ell, m) \\ &= \hat{\mathbf{F}}_{xx}^*(\ell, m)\boldsymbol{\beta} + \hat{\mathbf{F}}_{x\epsilon}^*(\ell, m) + \hat{\mathbf{F}}_{u_2u_1}(\ell, m) + \mathbf{U}(\ell, m)\end{aligned}\quad (\text{B.8})$$

where u_1 indexes the first element of the low-frequency contamination vector $\mathbf{u}$, $\mathbf{u}_2$ the vector with the remaining $p-1$ elements, and $\mathbf{U}(\ell, m) = \hat{\mathbf{F}}_{u_2x^*}(\ell, m)\boldsymbol{\beta} + \hat{\mathbf{F}}_{u_2\epsilon^*}(\ell, m) + \hat{\mathbf{F}}_{x^*u_1}(\ell, m)$ collects various cross-product terms. Hence, we proceed by analyzing each of the discretely averaged (cross)-periodograms in the decomposition (B.8). For the terms $\hat{\mathbf{F}}_{xx}^*(\ell, m)$ and $\hat{\mathbf{F}}_{u_2u_1}(\ell, m)$, we establish the following uniform bounds using the same arguments as in the proof of (a),

$$|\hat{\mathbf{F}}_{xx}^*(\ell, m) - \mathbf{G}_{xx}(\lambda_m)| \leq O_p((\ell/n)^{1-2d}) + o_p((m/n)^{1-2d}), \quad |\hat{\mathbf{F}}_{u_2u_1}(\ell, m)| \leq O_p(\ell^{-2}). \quad (\text{B.9})$$

Next, since $|\hat{\mathbf{F}}_{x\epsilon}^*(\ell, m, i)| \leq (\hat{\mathbf{F}}_{xx}^*(\ell, m, i)\hat{\mathbf{F}}_{\epsilon\epsilon}^*(\ell, m))^{1/2}$ for all $i = 1, \dots, p-1$ elements of the cross-product vector $\hat{\mathbf{F}}_{x\epsilon}^*(\ell, m, i)$ by the Cauchy-Schwarz inequality and, uniformly,

$$\hat{\mathbf{F}}_{\epsilon\epsilon}^*(\ell, m) \leq (1, -\boldsymbol{\beta}')\hat{\mathbf{F}}_{zz}^*(1, m)(1, -\boldsymbol{\beta}')' = (1, -\boldsymbol{\beta}')(\mathbf{F}_{zz}^*(\lambda_m) + o_p(\mathbf{F}_{zz}^*(\lambda_m)))(1, -\boldsymbol{\beta}')' = O_p(\lambda_m^{1-2d_\epsilon}),$$

using (B.5) and Assumption 2, it follows that $|\hat{\mathbf{F}}_{x\epsilon}^*(\ell, m, i)| \leq O_p(\lambda_m^{1-d-d_\epsilon})$. For the cross-product term, $\mathbf{U}(\ell, m)$, use the triangle inequality, the bound for $\mathcal{C}(\ell)$, and the Cauchy-Schwarz inequality to deduce

$$|\mathbf{U}(\ell, m)| \leq O_p\left(\ell^{-(1+d)}n^{(d-1/2)}\right) + O_p\left(\left(\hat{\mathbf{F}}_{\epsilon\epsilon}^*(\ell, m)\hat{\mathbf{F}}_{u_2u_2}(\ell, m)\right)^{1/2}\right), \quad (\text{B.10})$$

with the last term being $O_p((m/n)^{1/2-d_\epsilon}\ell^{-1})$. Finally, consistency of the MBLS estimator follows by collecting asymptotic bounds for the various components in (B.8) and writing

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\ell, m) &= \boldsymbol{\beta} + \mathbf{G}_{xx}^{-1}(\lambda_m) \left[O_p\left((\ell/n)^{1-2d}\right) + O_p(\ell^{-2}) + O_p\left((m/n)^{1-(d+d_\epsilon)}\right) \right] \\ &\quad + \mathbf{G}_{xx}^{-1}(\lambda_m) \left[O_p\left(\ell^{-(1+d)}n^{(d-1/2)}\right) + O_p((m/n)^{1/2-d_\epsilon}\ell^{-1}) \right] \\ &= \boldsymbol{\beta} + O_p\left((\ell/m)^{1-2d}\right) + O_p\left((n/m)^{1-2d}\ell^{-2}\right) + O_p\left((m/n)^{d-d_\epsilon}\right) \\ &\quad + O_p\left(\left((n/m)^{1-2d}\ell^{-2}\right)^{1/2}m^{-(1/2-d)}\ell^{-d}\right) + O_p\left(\left((n/m)^{1-2d}\ell^{-2}\right)^{1/2}(m/n)^{d-d_\epsilon}\right),\end{aligned}$$

uniformly, showing that $\hat{\boldsymbol{\beta}}(\ell, m) \xrightarrow{\mathbb{P}} \boldsymbol{\beta}$ when $1/m + m/n + 1/\ell + \ell/m \rightarrow 0$ and $(n/m)^{1-2d}\ell^{-2} \rightarrow 0$. $\square$

Remark 9. We use the maximal inequality from Varneskov (2015, Lemma C.4) to bound the low-frequency contamination. The result is originally derived to quantify the impact of market microstructure noise in the context of kernel-based estimation of the quadratic variation of Brownian semimartin-

gales. However, close inspection shows that it applies directly in the present context as well.

B.2 Proof of Theorem 2

First, for (a), make a decomposition similar to (B.2),

$$\hat{\mathbf{F}}_{zz}(\ell_c, m) = \hat{\mathbf{F}}_{zz}^*(\ell_c, m) + \hat{\mathbf{F}}_{uu}(\ell_c, m) + 2\hat{\mathbf{F}}_{zu}^*(\ell_c, m).$$

Then, by combining results from equations (B.4) and (B.7) above, $\hat{\mathbf{F}}_{zz}^*(\ell_c, m)$ may be uniformly approximated as

$$\hat{\mathbf{F}}_{zz}^*(\ell_c, m) = \mathbf{G}(\lambda_m) + o_p\left((m/n)^{1-2d}\right) \leq O_p\left((m/n)^{1-2d}\right),$$

since the addition and subtraction of $\hat{\mathbf{F}}_{zz}^*(1, \ell_c - 1)$ to get $\hat{\mathbf{F}}_{zz}^*(1, m)$ leads to an additional, asymptotically negligible error of order $O_p((\ell_c/n)^{1-2d})$ when ℓ_c is fixed, see, again, (B.7). In fact, if $\ell_c = 1$, then $\hat{\mathbf{F}}_{zz}^*(1, \ell_c - 1) = \mathbf{0}$. Next, for the discretely averaged periodogram of the low-frequency contamination vector, $\mathbf{u}$, we see that

$$\hat{\mathbf{F}}_{uu}(\ell_c, m) = \frac{2\pi}{n} \sum_{j=\ell_c}^m \Re(\mathbf{I}_{uu}(\lambda_j)) \xrightarrow{\mathbb{P}} \mathbf{K} \quad (\text{B.11})$$

where $\mathbf{K}$ is positive semi-definite and $\|\mathbf{K}\| < \infty$, since each $\Re(\mathbf{I}_{uu}(\lambda_j))$ in the sum is positive semi-definite and Assumption 1 implies that $\Re(\mathbf{I}_{uu}(\lambda_j)) = O_p(n/j^2)$, making $2\pi \sum_{j=\ell_c}^m O_p(j^{-2})$ a p -series with exponent $2 > 1$, trivially, that, as a result, converges as $m \rightarrow \infty$. Finally, similarly to (B.2),

$$\hat{\mathbf{F}}_{zu}^*(\ell_c, m) = \frac{2\pi}{n} \sum_{j=\ell_c}^m \Re(\mathbf{I}_{zu}^*(\lambda_j)) \leq \frac{2\pi}{n} \sum_{j=\ell_c}^m O_p\left((n/j)^{1+d} n^{-1/2}\right) = \frac{2\pi}{n^{1/2-d}} \sum_{j=\ell_c}^m O_p\left(j^{-(1+d)}\right),$$

using the Cauchy-Schwarz inequality. Hence, $|\hat{\mathbf{F}}_{zu}^*(\ell_c, m)| \leq O_p(n^{-1/2+d})$ since $2\pi \sum_{j=\ell_c}^m O_p(j^{-1-d})$ is a p -series with exponent $1 + d > 1$ and, as a result, is $O_p(1)$ when $m \rightarrow \infty$. The uniform asymptotic approximation in (a) follows by combining these three results.

Second, for (b), we invoke (a) to show $\hat{\mathbf{F}}_{xx}(\ell_c, m) \xrightarrow{\mathbb{P}} \mathbf{K}_{xx}$, uniformly, as $1/m + m/n \rightarrow 0$. Next, similarly to the decomposition in (B.8), write

$$\hat{\mathbf{F}}_{xy}(\ell_c, m) = \hat{\mathbf{F}}_{xx}(\ell_c, m)\boldsymbol{\beta} + \hat{\mathbf{F}}_{x\epsilon^*}(\ell_c, m) + \hat{\mathbf{F}}_{xu_1}(\ell_c, m) - \hat{\mathbf{F}}_{xu_2}(\ell_c, m)\boldsymbol{\beta},$$

where, as in Appendix B.1, u_1 indexes the first element of the low-frequency contamination vector, $\mathbf{u}$, and $\mathbf{u}_2$ the vector with the remaining $p - 1$ elements. Here, using the triangle inequality, uniform asymptotic approximations from (B.8), and the Cauchy-Schwarz inequality,

$$|\hat{\mathbf{F}}_{x\epsilon^*}(\ell_c, m)| \leq |\hat{\mathbf{F}}_{x\epsilon}^*(\ell_c, m)| + |\hat{\mathbf{F}}_{u_2\epsilon^*}(\ell_c, m)| \leq O_p\left((m/n)^{1-(d+d_\epsilon)}\right) + O_p\left((m/n)^{1/2-d_\epsilon}\right).$$

Moreover, since $\hat{\mathbf{F}}_{xu_1}(\ell_c, m) \xrightarrow{\mathbb{P}} \mathbf{K}_{xy}$ and $\hat{\mathbf{F}}_{xu_2}(\ell_c, m) \xrightarrow{\mathbb{P}} \mathbf{K}_{xx}$ uniformly by (a), this implies

$$\hat{\beta}(\ell_c, m) = \beta + O_p\left((m/n)^{1/2-d_\epsilon}\right) + \mathbf{K}_{xx}^{-1}(\mathbf{K}_{xy} - \mathbf{K}_{xx}\beta)(1 + o_p(1)) \xrightarrow{\mathbb{P}} \mathbf{K}_{xx}^{-1}\mathbf{K}_{xy},$$

uniformly, as $1/m + m/n \rightarrow 0$, using the continuous mapping theorem. This provides (b), from which the result in (c) follows trivially, concluding the proof. $\square$

B.3 Proof of Theorem 3

The following lemma is used to establish Theorem 3 below.

Lemma 5. *Under the conditions of Theorem 3, $\lambda_m^{2d-1}\hat{\mathbf{F}}_{xx}(\ell, m) \xrightarrow{\mathbb{P}} \mathbf{H}_{xx}$.*

Proof. The stated assumptions allow us to invoke Theorem 1 (a), providing, uniformly,

$$\hat{\mathbf{F}}_{xx}(\ell, m) = \mathbf{H}_{xx}(\lambda_m) + O_p\left((\ell/n)^{1-2d}\right) + O_p(\ell^{-2}) + O_p\left(\ell^{-(1+d)}n^{d-1/2}\right) + o_p\left((m/n)^{1-2d}\right).$$

Here, $\lambda_m^{2d-1}\mathbf{H}_{xx}(\lambda_m) = \mathbf{H}_{xx}$ by the definition in (10), $O_p(\lambda_m^{2d-1}(\ell/n)^{1-2d}) = O_p((\ell/m)^{1-2d}) \xrightarrow{\mathbb{P}} \mathbf{0}$ as $\ell/m \rightarrow 0$ and $d \in (0, 1/2)$, and $O_p(\lambda_m^{2d-1}\ell^{-2}) = O_p((n/m)^{1-2d}\ell^{-2}) \xrightarrow{\mathbb{P}} \mathbf{0}$, where the last convergence result is guaranteed by the (fourth) regularity condition in Assumption 5 since

$$(n/m)^{1-2d}\ell^{-2} = (n/m)^{1-d_\epsilon}\ell^{-2}(m/n)^{d-d_\epsilon} < m^{1/2}(n/m)^{1-d_\epsilon}\ell^{-2} \rightarrow 0.$$

The final convergence result, then, follows by manipulating the powers of $O_p(\lambda_m^{2d-1}\ell^{-(1+d)}n^{d-1/2})$ to show

$$(n/m)^{1-2d}\ell^{-(1+d)}n^{d-1/2} = \left((n/m)^{1-d_\epsilon}\ell^{-2}\right)^{1/2} \ell^{-d}m^{-(1/2-d)}(m/n)^{(d-d_\epsilon)/2} \rightarrow 0,$$

by the fourth regularity condition in Assumption 5, similarly to the result above. $\square$

Next, let us continue by using the decomposition in (B.8) to write

$$\begin{aligned} \sqrt{m}\lambda_m^{d_\epsilon-d}\beta(\ell, m) &= \sqrt{m}\lambda_m^{d_\epsilon-d}\hat{\mathbf{F}}_{xx}^{-1}(\ell, m)\hat{\mathbf{F}}_{xx}^*(\ell, m)\beta + \left(\lambda_m^{1-2d}\hat{\mathbf{F}}_{xx}^{-1}(\ell, m)\right)\left(\sqrt{m}\lambda_m^{d_\epsilon-1}\mathbf{F}_{u_2u_1}(\ell, m)\right) \\ &\quad + \left(\lambda_m^{1-2d}\hat{\mathbf{F}}_{xx}^{-1}(\ell, m)\right)\left(\sqrt{m}\lambda_m^{d_\epsilon-1}\mathbf{U}(\ell, m)\right) + \left(\lambda_m^{1-2d}\hat{\mathbf{F}}_{xx}^{-1}(\ell, m)\right)\left(\sqrt{m}\lambda_m^{d_\epsilon-1}\mathbf{F}_{x\epsilon}^*(\ell, m)\right) \\ &\equiv (E.1) + (E.2) + (E.3) + (E.4). \end{aligned}$$

The proof, then, readily follows by sequentially providing individual results for (E.1), (E.2), (E.3), and (E.4). First, for (E.1), we need to show that the last two terms of

$$\hat{\mathbf{F}}_{xx}(\ell, m) = \hat{\mathbf{F}}_{xx}^*(\ell, m) + O_p(\ell^{-2}) + O_p\left(\ell^{-(1+d)}n^{d-1/2}\right)$$

vanish sufficiently fast, their uniform stochastic orders following from (B.2). Here,

$$\sqrt{m}\lambda_m^{d_\epsilon-d}O_p(\ell^{-2}) = O_p\left(m^{1/2}(n/m)^{1-d_c}\ell^{-2}(m/n)^{1-2d}\right) \xrightarrow{\mathbb{P}} \mathbf{0}$$

by the fourth regularity condition in Assumption 5, $m/n \rightarrow 0$, and $d \in (0, 1/2)$. Moreover,

$$\begin{aligned} \sqrt{m}\lambda_m^{d_\epsilon-d}O_p\left(\ell^{-(1+d)}n^{d-1/2}\right) &= O_p\left((n/m)^{1-d_c}(m/n)^{1-2d}(m/n)^{1/2-d}\ell^{-(1+d)}m^d\right) \\ &= O_p\left(\left(m^{1/2}(n/m)^{1-d_c}\ell^{-2}\right)^\varpi (m/n)^{1-2d}\right) \\ &\quad \times m^{d-\varpi/2}(n/m)^{(1-d_c)(1-\varpi)+d-1/2}\ell^{2\varpi-1-d} \end{aligned} \quad (\text{B.12})$$

for some $\varpi > 0$, using algebraic manipulation of the powers. The first term inside the parenthesis goes to 0 by the fourth regularity condition in Assumption 5 and $\varpi > 0$, and since $(m/n)^{1-2d} \rightarrow 0$, we need to show existence of a $\varpi > 0$ such that, under the stated trimming and bandwidth assumptions, the last term outside the parenthesis also vanishes. By algebraic manipulations,

$$m^{d-\varpi/2}(n/m)^{(1-d_c)(1-\varpi)+d-1/2}\ell^{2\varpi-1-d} = m^{(1-\varpi)(d-1/2)}n^{1/2-\varpi(1-d)}\ell^{2\varpi-(1+d)}(m/n)^{d_\epsilon(1-\varpi)}$$

such that, after invoking polynomial forms $m \propto n^\kappa$ and $\ell \propto n^\nu$, we need to show that a $\varpi > 0$ solves

$$\kappa(1-\varpi)(d-1/2) + 1/2 - \varpi(1-d) + \nu(2\varpi - (1+d)) + d_\epsilon(\kappa-1)(1-\varpi) < 0 \quad (\text{B.13})$$

or, by re-arranging and collecting terms, that

$$\varpi((1-\kappa)(d_c-1/2) + 2(\nu-1/4)) < \kappa(1/2-d) + (1-\kappa)d_\epsilon - 1/2 + \nu(1+d).$$

Hence, since we can pick any $\varpi > 0$, it suffices to verify that $(1-\kappa)(d_c-1/2) + 2(\nu-1/4) > 0$ and $\kappa(1/2-d) + (1-\kappa)d_\epsilon - 1/2 + \nu(1+d) > 0$. The latter is trivial by the additional cross-product restriction we impose on ν in (9). For the first term, on the other hand, we may rewrite the positivity condition as

$$\nu > \frac{(1-\kappa)(1/2-d_c)}{2} + \frac{1}{4} = \frac{(1-\kappa)(1-d_c)}{2} + \frac{\kappa}{4}$$

by adding and subtracting $(1-\kappa)/4$. This condition is identical to the fourth regularity condition in Assumption 5 under polynomial forms for m and ℓ . Hence, (B.13) is satisfied by choosing

$$0 < \varpi < \frac{\kappa(1/2-d) + (1-\kappa)d_\epsilon - 1/2 + \nu(1+d)}{(1-\kappa)(d_c-1/2) + 2(\nu-1/4)},$$

thus proving that (B.12) is uniformly $o_p(1)$. As a result, we may write

$$\sqrt{m}\lambda_m^{d_\epsilon-d}\hat{\mathbf{F}}_{xx}^{-1}(\ell, m)\hat{\mathbf{F}}_{xx}^*(\ell, m)\boldsymbol{\beta} = \sqrt{m}\lambda_m^{d_\epsilon-d}\boldsymbol{\beta} \left(1 + o_p\left(m^{-1/2}\lambda_m^{d-d_\epsilon}\right)\right),$$

uniformly, by the continuous mapping theorem, such that

$$\sqrt{m}\lambda_m^{d_\epsilon-d}(\beta(\ell, m) - \beta) = (E.2) + (E.3) + (E.4) + o_p(1).$$

Next, we may use Lemma 5, in conjunction with the continuous mapping theorem, and (B.9), to show

$$|(E.2)| \leq \mathbf{H}_{xx}^{-1} \sqrt{m}\lambda_m^{d_c-1} O_p(\ell^{-2}) = O_p\left(m^{1/2}(n/m)^{1-d_c}\ell^{-2}\right) \xrightarrow{\mathbb{P}} \mathbf{0},$$

where the final convergence result follows by the fourth condition in Assumption 5.

For the cross-product-induced term, (E.3), we may use Lemma 5 and (B.10) to get the uniform stochastic bounds $|(E.3)| \leq O_p(m^{1/2}\lambda_m^{2d-1}\lambda_m^{d_\epsilon-d}\ell^{-1-d}n^{d-1/2}) + O_p(m^{1/2}\lambda_m^{d_c-1}\hat{\mathbf{F}}_{u_2\epsilon^*}(\ell, m))$. The first of the two asymptotic bounds is $o_p(1)$ by the same arguments used for (B.12) above. To show that the second term is asymptotically negligible, as well, requires different arguments since we have to rely on an alternative asymptotic approximation for the stochastic order than the one used in Section B.1. First, use the triangle and Cauchy-Schwarz inequalities to write

$$\left|\hat{\mathbf{F}}_{u_2\epsilon^*}(\ell, m)\right| \leq \frac{2\pi}{n} \sum_{j=\ell}^m |\Re(\mathbf{I}_{u_2\epsilon^*}(\lambda_j))| \leq \frac{2\pi}{n} \sum_{j=\ell}^m O_p\left(\frac{n^{1/2+d_\epsilon}}{j^{1+d_\epsilon}}\right) \leq \frac{2\pi}{n^{1/2-d_\epsilon}} \sum_{j=\ell}^m \frac{m^\varpi}{j^\varpi} O_p\left(j^{-(1+d_\epsilon)}\right)$$

for a different constant $\varpi > 0$. Hence, we may invoke Varneskov (2015, Lemma C.4) to show that $|\hat{\mathbf{F}}_{u_2\epsilon^*}(\ell, m)| \leq O_p(n^{d_\epsilon-1/2}m^\varpi\ell^{-1-d_\epsilon-\varpi})$ since $0 \leq d_\epsilon < d_c < 1/2$ and $1 + d_\epsilon + \varpi > 1$. The need to introduce the scale factor $(m/j)^\varpi$ follows directly from the fact that $d_\epsilon > 0$ is not guaranteed, and an exponent greater than 1 is required to bound the sum. Next, algebraic manipulations provide

$$\begin{aligned} m^{1/2}\lambda_m^{d_c-1}|\hat{\mathbf{F}}_{u_2\epsilon^*}(\ell, m)| &\leq O_p\left(\left(m^{1/2}(n/m)^{1-d_c}\ell^{-2}\right)^{1/2} m^{1/4}(n/m)^{(1-d_c)/2} n^{d_\epsilon-1/2} m^\varpi \ell^{-(d_\epsilon+\varpi)}\right) \\ &= O_p\left(\left(m^{1/2}(n/m)^{1-d_c}\ell^{-2}\right)^{1/2} m^{1/4-(1-d_c)/2+\varpi} n^{(d_\epsilon-d)/2} \ell^{-(d_\epsilon+\varpi)}\right) \xrightarrow{\mathbb{P}} \mathbf{0} \end{aligned}$$

since $m^{1/2}(n/m)^{1-d_c}\ell^{-2} \rightarrow 0$ by the fourth condition in Assumption 5, $0 \leq d_\epsilon < d \leq d_c < 1/2$, and, finally, by choosing $0 < \varpi < (1 - d_c)/2 - 1/4$. By combining results, $|(E.3)| \leq o_p(1)$.

For the last term, (E.4), we may write

$$\sqrt{m}\lambda_m^{d_c-1}\mathbf{F}_{x\epsilon}^*(\ell, m) = \sqrt{m}\lambda_m^{d_c-1}\mathbf{F}_{x\epsilon}^*(1, m) - \sqrt{m}\lambda_m^{d_c-1}\mathbf{F}_{x\epsilon}^*(1, \ell - 1).$$

Hence, it suffices to show $\sqrt{m}\lambda_m^{d_c-1}\mathbf{F}_{x\epsilon}^*(1, \ell - 1) = o_p(1)$ uniformly since, in this case, we may invoke Christensen & Nielsen (2006, Theorem 2) for $\sqrt{m}\lambda_m^{d_c-1}\mathbf{F}_{x\epsilon}^*(1, m)$ and use this in conjunction with Lemma 5, the continuous mapping theorem and Slutsky's theorem to prove the final central limit theorem. To show the former, we use the Cramér-Wold Theorem, cf. Davidson (2002, Theorem 25.5),

for an arbitrary $(p-1) \times 1$ vector, $\boldsymbol{\psi}$,

$$\begin{aligned}
\boldsymbol{\psi}' \sqrt{m} \lambda_m^{d_c-1} \mathbf{F}_{x\epsilon}^*(1, \ell-1) &= \sum_{i=1}^{p-1} \psi_i \sqrt{m} \lambda_m^{d_c-1} \mathbf{F}_{x\epsilon}^*(1, \ell-1, i) \\
&= \sum_{i=1}^{p-1} \psi_i \sqrt{m} \lambda_m^{d_c-1} \frac{2\pi}{n} \sum_{j=1}^{\ell-1} \Re(\mathbf{I}_{x\epsilon}^*(\lambda_j, i)) \\
&= \sum_{i=1}^{p-1} \psi_i \sqrt{m} \lambda_m^{d_c-1} \frac{2\pi}{n} \sum_{j=1}^{\ell-1} \Re(\mathbf{I}_{x\epsilon}^*(\lambda_j, i) - \mathbf{A}^*(\lambda_j, i) \mathbf{J}(\lambda_j) \bar{\mathbf{A}}^*(\lambda_j, p)) \\
&\quad + \sum_{i=1}^{p-1} \psi_i \sqrt{m} \lambda_m^{d_c-1} \frac{2\pi}{n} \sum_{j=1}^{\ell-1} \Re(\mathbf{A}^*(\lambda_j, i) \mathbf{J}(\lambda_j) \bar{\mathbf{A}}^*(\lambda_j, p)) \\
&\equiv (E.4.1) + (E.4.2)
\end{aligned}$$

where, for (E.4.1), we may use summation by parts and Lobato (1999, C.2) to show

$$\begin{aligned}
(E.4.1) &\leq O_p \left(\sum_{i=1}^{p-1} \eta_i \sqrt{m} \lambda_m^{d_c-1} \lambda_\ell^{-d_c} \frac{1}{n} \left[\ell^{1/3} (\log \ell)^{2/3} + \log \ell + \ell^{1/2} / n^{1/4} \right] \right) \\
&= O_p \left(\left(\frac{m}{\ell} \right)^{d_c} \left[\frac{\ell^{1/3}}{m^{1/2}} (\log \ell)^{2/3} + \frac{\log \ell}{m^{1/2}} + \left(\frac{\ell}{m} \right)^{1/2} \frac{1}{n^{1/4}} \right] \right) \\
&= O_p \left(\left(\frac{\ell}{m} \right)^{1/2-d_c} \left[\frac{(\log \ell)^{2/3}}{\ell^{1/6}} + \frac{\log \ell}{\ell^{1/2}} + \frac{1}{n^{1/4}} \right] \right) \xrightarrow{\mathbb{P}} 0
\end{aligned}$$

since $d_c \in (0, 1/2)$ and $\ell/m \rightarrow 0$ as $n \rightarrow \infty$. Lastly, for (E.4.2), since $\mathbf{v}_t$ is a martingale difference sequence with finite fourth moment, $\|\mathbf{J}(\lambda_j)\| \leq K < \infty$, $j = 1, \dots, \ell-1$,

$$\begin{aligned}
(E.4.2) &\leq K \sum_{i=1}^{p-1} \psi_i \sqrt{m} \lambda_m^{d_c-1} \frac{2\pi}{n} \sum_{j=1}^{\ell-1} \Re(\mathbf{A}^*(\lambda_j, i) \bar{\mathbf{A}}^*(\lambda_j, p)) \\
&\leq K \sqrt{m} \lambda_m^{d_c-1} \frac{2\pi}{n} \sup_{i=1, \dots, p-1} \sum_{j=1}^{\ell-1} |\mathbf{f}_{wv}^*(\lambda_j, i, p)| = O \left(\sqrt{m} \lambda_m^{d_c-1} \frac{\ell}{n} \lambda_\ell^{\varphi-d_c} \right) = O \left(\frac{\ell^{1+d_c+\varphi}}{m^{1/2+d_c} n^\varphi} \right),
\end{aligned}$$

which is $o_p(1)$, using Assumption 2' for the first equality, since $G_{i,p} = 0 \ \forall i = 1, \dots, p-1$, and the fifth condition of Assumption 5 for the final convergence result. Note, however, that we need to provide additional conditions on $\varphi \in (0, 2]$ such that the fifth condition of Assumption 5 is mutually consistent with always allowing $d_c \in (0, 1/2)$. Using the polynomial forms $m \propto n^\kappa$ and $\ell \propto n^\nu$, the fifth condition of Assumption 5 is satisfied if $\nu(1 + d_c + \varphi) < \kappa(1/2 + d_c) + \varphi$, implying that

$$d_c > \frac{(\nu-1)\varphi + \nu - 1/2\kappa}{\kappa - \nu}.$$

Since we wish to allow all $d_c \in (0, 1/2)$, and the left-hand-side is strictly decreasing in φ , it suffices to impose $\max(0, (\nu - \kappa/2)/(1 - \nu)) < \varphi \leq 2$. To verify that this restriction always offers a solution, note that conditions two and three of Assumption 5 ensure that $(\nu - \kappa/2)/(1 - \nu) < 2$.

Finally, the use of all the uniform bounds for (E.1), (E.2), (E.3), and (E.4), in conjunction with Lemma 5 and Christensen & Nielsen (2006, Theorem 2), completes the proof. $\square$

B.4 Proof of Lemma 2

Apart from allowing for dependence between z^* and u , the stated assumptions satisfy the corresponding conditions for Iacone (2010, Theorems 2 and 3), see also McCloskey & Perron (2013, Corollary 1, Theorems 2 and 3). Now, careful inspection of the respective proofs provided in Iacone (2010) readily shows that these nowhere rely on independence between z^* and u to eliminate the impact of the low-frequency contamination on $\hat{G}(\delta, z, \ell_1, m_1)$, that is, in showing that the last two terms of

$$\hat{G}(\delta, z, \ell_1, m_1) = \hat{G}(\delta, z^*, \ell_1, m_1) + \hat{G}(\delta, u, \ell_1, m_1) + \frac{2}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \lambda_j^{2\delta} \Re(I_{z^*, u}(\lambda_j))$$

are uniformly asymptotically negligible. Instead, they rely on an asymptotic order assumption for u , which is satisfied by Assumption 1 (with $\phi = 1/2$ in their notation), and the Cauchy-Schwarz inequality for cross-products. Specifically, such arguments are provided in Iacone (2010, (A.7)-(A.9)) for consistency and Iacone (2010, (A.16)-(A.25)) for asymptotic normality. Hence, we may directly invoke Iacone (2010, Theorems 2 and 3), as well as their proofs, to show **(a)** and **(b)**, respectively. $\square$

B.5 Proof of Theorem 4

First, rewrite the observable residuals from a first-stage MBLS regression as

$$\hat{e}_t = y_t - \beta(\ell, m)'x_t = \alpha + \epsilon_t^* + (\beta - \beta(\ell, m))'x_t^* + \tilde{u}_t$$

where $\tilde{u}_t = u_{1,t} - \beta(\ell, m)'u_{2,t}$. Without loss of generality we have set $\mathbf{c} = \mathbf{0}$ in (1), since a constant vector is a special case of the low-frequency contamination in Assumption 1. Hence, we may write

$$\begin{aligned} I_{\hat{e}\hat{e}}(\lambda_j) &= I_{\epsilon\epsilon}^*(\lambda_j) + I_{\tilde{u}\tilde{u}}(\lambda_j) + 2I_{\epsilon\tilde{u}}^*(\lambda_j) + (\beta - \beta(\ell, m))'\Re(\mathbf{I}_{xx}^*(\lambda_j))(\beta - \beta(\ell, m)) \\ &\quad + 2(\beta - \beta(\ell, m))'\Re(\mathbf{I}_{x\epsilon}^*(\lambda_j)) + 2(\beta - \beta(\ell, m))'\Re(\mathbf{I}_{x\tilde{u}}^*(\lambda_j)) + o_p(1) \end{aligned}$$

where $I_{\epsilon\epsilon}^*(\lambda_j) + I_{\tilde{u}\tilde{u}}(\lambda_j) + 2I_{\epsilon\tilde{u}}^*(\lambda_j)$ corresponds to the periodogram analyzed in the objective function in Lemma 2, see also Iacone (2010) and McCloskey & Perron (2013), since $\tilde{u}_t = (1, -\beta')\mathbf{u}_t(1 + o_p(1))$ is a linear combination of low-frequency contamination, which, similarly to (A.1), satisfies the high-level condition in Assumption 1, and $d_\epsilon \in [0, d_c]$. The last three terms are errors from having estimated β

in the first stage using MBLS, one of which depend on low-frequency contamination. Denote

$$\begin{aligned} c_1(\lambda_j) &= 2(\beta - \beta(\ell, m))' \Re(\mathbf{I}_{x\bar{u}}^*(\lambda_j)), \quad c_2(\lambda_j) = (\beta - \beta(\ell, m))' \Re(\mathbf{I}_{xx}^*(\lambda_j))(\beta - \beta(\ell, m)), \\ c_3(\lambda_j) &= 2(\beta - \beta(\ell, m))' \Re(\mathbf{I}_{x\epsilon}^*(\lambda_j)), \end{aligned}$$

then if we can show that $|c_1(\lambda_j)| \leq o_p(I_{\epsilon\bar{u}}^*(\lambda_j))$, the conditions on $m_1 \propto n^{\kappa_1}$ and $\ell_1 \propto n^{\nu_1}$ provided in Assumptions LW' (a) and (b) for consistency and asymptotic normality, respectively, suffice to eliminate this low-frequency contamination term. We readily have $|\Re(\mathbf{I}_{x\bar{u}}^*(\lambda_j))| \leq O_p(n^{1/2+d}j^{-1-d})$ as well as $|I_{\epsilon\bar{u}}^*(\lambda_j)| \leq O_p(n^{1/2+d_\epsilon}j^{-1-d_\epsilon})$ by the Cauchy-Schwarz inequality. Hence, using Theorem 3,

$$|c_1(\lambda_j)| \leq O_p\left(m^{-1/2}\lambda_m^{d-d_\epsilon}(n/j)^{1+d}n^{-1/2}\right) = O_p\left(m^{d-(1/2+d_\epsilon)}j^{-(d-d_\epsilon)}(n/j)^{1+d_\epsilon}n^{-1/2}\right),$$

which is $o_p(I_{\epsilon\bar{u}}^*(\lambda_j))$ since $0 \leq d_\epsilon < d < 1/2$. As a result, the additional error due to first-stage estimation in conjunction with low-frequency contamination is of smaller stochastic order than the remaining low-frequency contamination terms. Next, $c_2(\lambda_j)$ and $c_3(\lambda_j)$ correspond to the additional errors introduced by first-stage estimation for the NBLs estimator in the absence of low-frequency contamination, which are considered in Nielsen & Frederiksen (2011, Theorem 2.2), but their asymptotic impact on the LW estimator differs slightly here since the latter is trimmed with $\ell_1 \propto n^{\nu_1}$. However, since Assumptions 2', 3', 4, and 5-LW' (a), respectively, (b) satisfy the conditions for consistency, respectively, asymptotic normality in Nielsen & Frederiksen (2011, Theorem 2.2), and since Theorem 3 shows that the MBLS estimator enjoys the same asymptotic distribution in the presence of low-frequency contamination as the NBLs estimator does in its absence, this means that if

$$(E.5) = \frac{G_{p,p}^{-1}}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \lambda_j^{2\delta+2d_\epsilon} c_1(\lambda_j) \leq O_p\left(\frac{G_{p,p}^{-1}}{m_1 - \ell_1 + 1} \sum_{j=1}^{m_1} \lambda_j^{2\delta+2d_\epsilon} c_1(\lambda_j)\right), \quad (B.14)$$

$$(E.6) = \frac{G_{p,p}^{-1}}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \lambda_j^{2\delta+2d_\epsilon} c_2(\lambda_j) \leq O_p\left(\frac{G_{p,p}^{-1}}{m_1 - \ell_1 + 1} \sum_{j=1}^{m_1} \lambda_j^{2\delta+2d_\epsilon} c_2(\lambda_j)\right), \quad (B.15)$$

for $\delta = 0$, that is, for the highest asymptotic order of the bias, then trimming has no impact on their analysis, and we may invoke Nielsen & Frederiksen (2011, Theorem 2.2) to show that the errors from first-stage estimation of β in (E.5) and (E.6) are asymptotically negligible such that the final results in (a) and (b) readily follow from Lemmas 2 (a) and (b), respectively. Note that the scaling of (E.5) and (E.6) with $G_{p,p}^{-1}\lambda_j^{2d_\epsilon}$ follows by Nielsen & Frederiksen (2011, (A.15)). First, for (B.14), the bound is immediate as $c_1(\lambda_j) \geq 0 \forall j = 1, \dots, \lfloor n/2 \rfloor$. Next, to establish (B.15), we may write,

$$(E.6) = \frac{2G_{p,p}^{-1}}{m_1 - \ell_1 + 1} (\beta - \beta(\ell, m))' \left(\sum_{j=1}^{m_1} \lambda_j^{2d_\epsilon} \Re(\mathbf{I}_{x\epsilon}^*(\lambda_j)) - \sum_{j=1}^{\ell_1-1} \lambda_j^{2d_\epsilon} \Re(\mathbf{I}_{x\epsilon}^*(\lambda_j)) \right).$$

Hence, as in the proof of Theorem 3, we may use the Cramér-Wold Theorem for an arbitrary $(p-1) \times 1$ vector, ψ , on this decomposition such that verifying condition (B.15) amounts to showing

$$O \left(\left| \sum_{i=1}^{p-1} \psi_i \sum_{j=1}^{\ell_1-1} \lambda_j^{2d_\epsilon} \Re(\mathbf{I}_{x\epsilon}^*(\lambda_j, i)) \right| \right) \leq O \left(\left| \sum_{i=1}^{p-1} \psi_i \sum_{j=1}^{m_1} \lambda_j^{2d_\epsilon} \Re(\mathbf{I}_{x\epsilon}^*(\lambda_j, i)) \right| \right). \quad (\text{B.16})$$

Let us introduce the generic notation $s_1 = \{\ell_1 - 1, m_1\}$. Then, under the stated assumptions, we may invoke Lobato (1999, (C.4)) and Nielsen & Frederiksen (2011, Lemma B1(c)), which, in conjunction with $G_{i,p} = G_{p,i} = 0$, $i = 1, \dots, p-1$, provides the bound,

$$\left| \sum_{i=1}^{p-1} \psi_i \sum_{j=1}^{s_1} \lambda_j^{2d_\epsilon} \Re(\mathbf{I}_{x\epsilon}^*(\lambda_j, i)) \right| = O_p \left(\lambda_{s_1}^{d_\epsilon-d} \left[s_1^{1+\varphi} n^{-\varphi} + s_1^{1/2} (\log s_1) \right] \right).$$

This result, since $s_1^{1+\varphi-d+d_\epsilon}$ and $s_1^{1/2-d+d_\epsilon}$ are both strictly increasing in s_1 , shows that the sum with $s_1 = m_1$ is of a strictly higher stochastic order than the corresponding sum with $s_1 = \ell_1 - 1$ as $\ell_1/m_1 \propto n^{\nu_1-\kappa_1} \rightarrow 0$ when $n \rightarrow \infty$, thus verifying condition (B.16) and thereby (B.15). Hence, we invoke Nielsen & Frederiksen (2011, Theorem 2.2) to asymptotically eliminate the right-hand-side errors in conditions (B.14) and (B.15) and use Lemma 2 to complete the proof. $\square$

B.6 Proof of Theorem 5

Let $I_{\epsilon\epsilon}(\lambda_j)$ be the periodogram of $\epsilon_t = y_t - \beta' \mathbf{x}_t$ at frequency ordinate λ_j . Note that this is not to be confused with $I_{\epsilon\epsilon}^*(\lambda_j)$, the periodogram of $\epsilon_t^* = y_t^* - \beta' \mathbf{x}_t^*$, but merely used as a convenient intermediate step to establish the desired consistency result for $\hat{\epsilon}_t = y_t - \hat{\beta}(\ell, m)' \mathbf{x}_t$ in Theorem 5 (b).

Lemma 6. *Under Assumptions 1, 2', 3', and 4, and for any sequence of positive integers $j = j(n)$ such that $j/n \rightarrow 0$ as $n \rightarrow \infty$, then uniformly*

$$(a) \quad \mathbb{E} \left[\lambda_j^{2d} I_{xx}(\lambda_j) \right] = G_{xx} \left[1 + O \left(\log j/j + (j/n)^\varphi + j^{2d-2} n^{1-2d} + (n/j)^{1-d} n^{-1/2} \right) \right],$$

$$(b) \quad \mathbb{E} \left[\lambda_j^{2d_\epsilon} I_{\epsilon\epsilon}(\lambda_j) \right] = G_{p,p} \left[1 + O \left(\log j/j + (j/n)^\varphi + j^{2d_\epsilon-2} n^{1-2d_\epsilon} + (n/j)^{1-d_\epsilon} n^{-1/2} \right) \right].$$

Proof. Let us write $\tilde{u}_t = (1, -\beta') \mathbf{u}_t$ for the low-frequency contamination of ϵ_t and decompose

$$\mathbf{I}_{xx}(\lambda_j) = \mathbf{I}_{xx}^*(\lambda_j) + \mathbf{I}_{u_2 u_2}(\lambda_j) + 2\mathbf{I}_{x u_2}^*(\lambda_j), \quad I_{\epsilon\epsilon}(\lambda_j) = I_{\epsilon\epsilon}^*(\lambda_j) + I_{\tilde{u}\tilde{u}}(\lambda_j) + 2I_{\epsilon\tilde{u}}^*(\lambda_j).$$

Under Assumptions 2', 3', and 4, $\mathbf{z}_t^*$ satisfies the conditions of Robinson (1995b, Theorem 2). Moreover, $\tilde{u}_t$ forms a linear combination of low-frequency contamination for the process ϵ_t , which, similarly to (A.1), satisfies the high-level Assumption 1. Hence, we may invoke Robinson (1995b, Theorem 2 (a)) to provide the first two asymptotic orders for both of the latent fractional signals, and use this in conjunction with $\mathbb{E}[\mathbf{I}_{u_2 u_2}(\lambda_j)] = O(nj^{-2})$ and $\mathbb{E}[I_{\tilde{u}\tilde{u}}(\lambda_j)] = O(nj^{-2})$, uniformly, by Assumption 1, as well as the Cauchy-Schwarz inequality to provide the asymptotic orders for $\mathbf{I}_{x u_2}^*(\lambda_j)$ and $I_{\epsilon\tilde{u}}^*(\lambda_j)$. $\square$

First, for **(a)**, under Assumptions 1, 2', 3', and 4, the decomposition in Lemma 6 (a) provides us with four approximation errors of stochastic orders

$$(E.7.1) = O_p\left(m_2^{-1} \sum_{j=\ell_2}^{m_2} (\log j) j^{-1}\right), \quad (E.7.2) = O_p\left(m_2^{-1} n^{-\varphi} \sum_{j=\ell_2}^{m_2} j^\varphi\right),$$

$$(E.7.3) = O_p\left(m_2^{-1} n^{1-2d} \sum_{j=\ell_2}^{m_2} j^{-(2-2d)}\right), \quad (E.7.4) = O_p\left(m_2^{-1} n^{1/2-d} \sum_{j=\ell_2}^{m_2} j^{-(1-d)}\right),$$

respectively. Next, to establish stochastic bounds for (E.7.1) and (E.7.2), we use the result in Nielsen & Frederiksen (2011, (A.17)), which states that

$$\sup_{-1 \leq \tau \leq K} \left| m_2^{-\tau-1} (\log m_2)^{-1} \sum_{j=1}^{m_2} j^\tau \right| = O(1), \quad \text{for } K \in (1, \infty). \quad (\text{B.17})$$

This immediately provides

$$|(E.7.1)| \leq O_p\left(m_2^{-1} (\log m_2) \sum_{j=1}^{m_2} j^{-1}\right) \leq O_p\left((\log m_2)^2 m_2^{-1}\right), \quad |(E.7.2)| \leq O_p\left((m_2/n)^\varphi (\log m_2)\right),$$

by multiplying and dividing through with the relevant quantities. For (E.7.3), and similarly to the proof of Theorem 1 (a), we may apply Varneskov (2015, Lemma C.4) since $2(1-d) > 1$, thus providing the uniform bound $|(E.7.3)| \leq O_p(m_2^{-1} n^{1-2d} \ell_2^{2d-2})$. For the last term, (E.7.4), the direct use of (B.17) to bound the sum leads to strict requirements on the rate at which the bandwidth, m_2 , tends to infinity. Hence, and similarly to the proof of Theorem 3, define some $\varpi > 0$ and write for (E.7.4),

$$\frac{n^{1/2-d}}{m_2} \sum_{j=\ell_2}^{m_2} j^{-(1-d)} \leq \frac{n^{1/2-d}}{m_2^{1-(d+\varpi)}} \sum_{j=\ell_2}^{m_2} j^{-(1+\varpi)} \leq \left(\frac{m_2}{n}\right)^d \frac{n^{1/2}}{m_2 \ell_2} \left(\frac{m_2}{\ell_2}\right)^\varpi. \quad (\text{B.18})$$

Since (B.18) is valid for any $\varpi > 0$, we choose ϖ arbitrarily small such that the fourth condition Assumption 5-G (a) provides $|(E.7.4)| \leq o_p(1)$. Similarly, condition one, two and four of Assumption 5-G (a) provide $|(E.7.1)| \leq o_p(1)$, $|(E.7.2)| \leq o_p(1)$, and $|(E.7.3)| \leq o_p(1)$. This proves **(a)**.

For **(b)**, we first use exactly the same arguments as for **(a)** to show

$$\hat{G}(d_\epsilon, \epsilon, \ell_2, m_2) \xrightarrow{\mathbb{P}} G_{p,p}. \quad (\text{B.19})$$

Next, write $\hat{e}_t = y_t - \hat{\beta}(\ell, m)' \mathbf{x}_t = \epsilon_t + (\beta - \hat{\beta}(\ell, m))' \mathbf{x}_t \equiv \epsilon_t + \hat{x}_t$ such that we may decompose the periodogram of $\hat{e}_t$ as $I_{\hat{e}\hat{e}}(\lambda_j) = I_{\epsilon\epsilon}(\lambda_j) + I_{\hat{x}\hat{x}}(\lambda_j) + 2I_{\epsilon\hat{x}}(\lambda_j)$. Then, due to (B.19), it immediately follows that if we can show $\hat{G}(d_\epsilon, \hat{x}, \ell_2, m_2) \xrightarrow{\mathbb{P}} 0$, this will establish **(b)** since the contribution from cross-products between ϵ_t and $\hat{x}_t$ to $\hat{G}(d_\epsilon, \hat{e}, \ell_2, m_2)$ will vanish asymptotically, using the Cauchy-Schwarz

inequality bound: $|\hat{G}(d_\epsilon, \epsilon, \ell_2, m_2)|^{1/2} |\hat{G}(d_\epsilon, \hat{x}, \ell_2, m_2)|^{1/2}$. Hence, write

$$\begin{aligned}\hat{G}(d_\epsilon, \hat{x}, \ell_2, m_2) &= (\beta - \hat{\beta}(\ell, m))' \left(\frac{1}{m_2 - \ell_2 + 1} \sum_{j=\ell_2}^{m_2} \lambda_j^{d_\epsilon - d} \lambda_j^d \mathbf{I}_{xx}(\lambda_j) \right) (\beta - \hat{\beta}(\ell, m)) \\ &\leq K(\beta - \hat{\beta}(\ell, m))' \left(n^{d-d_\epsilon} m_2^{-1} \sum_{j=\ell_2}^{m_2} \lambda_j^d \mathbf{I}_{xx}(\lambda_j) \right) (\beta - \hat{\beta}(\ell, m)) \\ &= O_p \left(m^{-1/2} (m/n)^{d-d_\epsilon} n^{d-d_\epsilon} m_2^{-1} m^{-1/2} (m/n)^{d-d_\epsilon} \right) \xrightarrow{\mathbb{P}} 0\end{aligned}$$

using $m_2/(m_2 - \ell_2 + 1) \leq K$ and $n/j \leq n$ for the first inequality, Theorem 3 and **(a)** for the final equality, and $m/n \rightarrow 0$ and $0 \leq d_\epsilon < d < 1/2$ for the final convergence result. This proves **(b)**. $\square$

B.7 Proof of Lemma 3

First, define some auxiliary processes. Let $p_t = p_{01} \mathbf{1}\{s_t = 0\} + p_{10} \mathbf{1}\{s_t = 1\}$ and

$$\mu_t = (\eta_{t,1} - \eta_{t,0}) \mathbf{1}\{s_{t-1} = 0\} + (\eta_{t,0} - \eta_{t,1}) \mathbf{1}\{s_{t-1} = 1\} \equiv \eta_{t,01} \mathbf{1}\{s_{t-1} = 0\} + \eta_{t,10} \mathbf{1}\{s_{t-1} = 1\}$$

where $\eta_{t,01} = \eta_{t,1} - \eta_{t,0} \stackrel{\mathbb{D}}{\sim} \text{i.i.d. N}(\mu_{10}, \sigma_\eta^2)$ and $\eta_{t,10} = \eta_{t,0} - \eta_{t,1} \stackrel{\mathbb{D}}{\sim} \text{i.i.d. N}(\mu_{01}, \sigma_\eta^2)$ with $\mu_{10} = \mu_1 - \mu_0$, $\mu_{01} = -\mu_{10}$, and $\sigma_\eta^2 = \sigma_{\eta,0}^2 + \sigma_{\eta,1}^2$. Then, the Markov switching process, v_t , may be rewritten as

$$v_t = v_1 + \tilde{v}_t, \quad \tilde{v}_t = \sum_{i=2}^t \pi_i \mu_i, \quad \pi_i \stackrel{\mathbb{D}}{\sim} \text{Bernoulli}(p_{i-1}/n), \quad t \geq 2, \quad (\text{B.20})$$

where π_i is independent, but not identically distributed. Note that v_1 and $\tilde{v}_t$, $t \geq 2$, are independent due to $\eta_{t,0}$, $\eta_{t,1}$ and π_t all being independent over time. Next, add and subtract v_1 in $I_{vv}(\lambda_j)$ and make the decomposition,

$$\begin{aligned}I_{vv}(\lambda_j) &= \frac{1}{2\pi n} \sum_{s=1}^n \sum_{t=1}^n (v_s - v_1)(v_t - v_1) \cos\left(\frac{2\pi j(t-s)}{n}\right) + \frac{v_1}{\pi n} \sum_{s=1}^n \sum_{t=1}^n v_t \cos\left(\frac{2\pi j(t-s)}{n}\right) \\ &\quad - \frac{v_1^2}{2\pi n} \sum_{s=1}^n \sum_{t=1}^n \cos\left(\frac{2\pi j(t-s)}{n}\right) \equiv I_{vv}^1(\lambda_j) + I_{vv}^2(\lambda_j) - I_{vv}^3(\lambda_j),\end{aligned}$$

and note that $I_{vv}^1(\lambda_j)$ corresponds to the periodogram of $\tilde{v}_t$, $t \geq 2$. Next, define $\bar{v}_t = \sum_{i=1}^t \pi_i$ with π_i defined as in (B.20), and $\bar{v}_{t,h} = \sum_{i=1}^t \pi_{i,h}$, $h = \{0, 1\}$, where $\pi_{i,0} \stackrel{\mathbb{D}}{\sim} \text{i.i.d. Bernoulli}(p_{01}/n)$ and, similarly, $\pi_{i,1} \stackrel{\mathbb{D}}{\sim} \text{i.i.d. Bernoulli}(p_{10}/n)$, then we may write (up to an $o_p(1)$ error)

$$\begin{aligned}I_{vv}^1(\lambda_j) &= \frac{1}{2\pi n} \sum_{s=2}^n \sum_{t=2}^n \tilde{v}_s \tilde{v}_t \cos\left(\frac{2\pi j(t-s)}{n}\right) \\ &\leq \frac{K}{2\pi n} \sum_{s=2}^n \sum_{t=2}^n \bar{v}_s \bar{v}_t \cos\left(\frac{2\pi j(t-s)}{n}\right)\end{aligned}$$

$$\leq \frac{K_1}{2\pi n} \sum_{s=2}^n \sum_{t=2}^n \bar{v}_{s,0} \bar{v}_{t,0} \cos\left(\frac{2\pi j(t-s)}{n}\right) + \frac{K_2}{2\pi n} \sum_{s=2}^n \sum_{t=2}^n \bar{v}_{s,1} \bar{v}_{t,1} \cos\left(\frac{2\pi j(t-s)}{n}\right),$$

for which the two final terms on the right-hand-side are scaled periodograms for special cases of the Lévy processes covered by Example 7. Hence, we refer to Lemma 4 above and Section B.8 below for a proof that

$$\mathbb{E}[(j^2/n)I_{vv}^1(\lambda_j)] \leq O(1), \quad \mathbb{V}[(j^2/n)I_{vv}^1(\lambda_j)] \leq O(1).$$

Next, for $I_{vv}^2(\lambda_j)$, use independence of v_1 and $\tilde{v}_t$, $t \geq 2$, to write

$$\begin{aligned} \mathbb{E}\left[\frac{j^2}{n}I_{vv}^2(\lambda_j)\right] &= \frac{j^2\mathbb{E}[v_1^2]}{\pi n^2} \sum_{s=1}^n \sum_{t=1}^n \cos\left(\frac{2\pi j(t-s)}{n}\right) \\ &\rightarrow \frac{j^2\mathbb{E}[v_1^2]}{\pi} \int_0^1 \int_0^1 \cos(2\pi j(t-s)) dt ds = \frac{j^2\mathbb{E}[v_1^2]}{\pi} \times \frac{\sin^2(\pi j)}{\pi^2 j^2} = O(1), \end{aligned}$$

using Riemann integration. The asymptotic order for the variance follows similarly as

$$\begin{aligned} \mathbb{V}\left[\frac{j^2}{n}I_{vv}^2(\lambda_j)\right] &\leq \frac{j^4}{\pi^2 n^4} \mathbb{E}\left[\left(v_1 \sum_{s=1}^n \sum_{t=1}^n v_t \cos\left(\frac{2\pi j(t-s)}{n}\right)\right) \left(v_1 \sum_{s=1}^n \sum_{t=1}^n v_s \cos\left(\frac{2\pi j(t-s)}{n}\right)\right)\right] \\ &= \frac{j^4\mathbb{E}[v_1^4]}{\pi^2 n^4} \left(\sum_{s=1}^n \sum_{t=1}^n \cos\left(\frac{2\pi j(t-s)}{n}\right)\right) \left(\sum_{s=1}^n \sum_{t=1}^n \cos\left(\frac{2\pi j(t-s)}{n}\right)\right), \end{aligned}$$

which is $O(1)$ by the same argument as for the expected value. Finally, since the corresponding bounds for $I_{vv}^3(\lambda_j)$ follow by identical arguments, we can collect results for the respective unconditional means and variances of the three terms $I_{vv}^1(\lambda_j)$, $I_{vv}^2(\lambda_j)$, and $I_{vv}^3(\lambda_j)$, and use them, in conjunction with the Cauchy-Schwarz inequality, to bound their covariance, thus providing the final result. $\square$

B.8 Proof of Lemma 4

First, using boundedness of the jump sizes, the periodogram of v_t may be bounded as

$$\begin{aligned} \frac{j^2}{n}I_{vv}(\lambda_j) &\leq \frac{K}{2\pi} \frac{j^2}{n^2} \sum_{s=1}^n \sum_{t=1}^n \pi_s \pi_t \cos\left(\frac{2\pi j(s-t)}{n}\right) \\ &= (1 + o_p(n^{-1})) \frac{Kj^2}{2\pi} \int_0^1 \int_0^1 \pi(s)\pi(t) \cos(2\pi j(s-t)) ds dt \equiv I_{\pi\pi}(\lambda_j)(1 + o_p(n^{-1})) \end{aligned}$$

where $\pi(t) \stackrel{\mathbb{D}}{\sim} \text{Poisson}(p_t \cdot t)$, which follows by Riemann integration in conjunction with boundedness and Lipschitz continuity of the time-varying intensity p_t . Hence, to establish Lemma 4, it suffices to show that the unconditional mean and variance of $I_{\pi\pi}(\lambda_j)$ are bounded. Before proceeding, however, let us define $\xi_t \equiv p_t \cdot t$ and recall that by Dobinski's formula,

$$\mathbb{E}[\pi(t)^2|p_t] = \xi_t(1 + \xi_t), \quad \mathbb{E}[\pi(t)^3|p_t] = \xi_t(1 + 3\xi_t + \xi_t^2), \quad \mathbb{E}[\pi(t)^4|p_t] = \xi_t(1 + 7\xi_t + 6\xi_t^2 + \xi_t^3).$$

Then, the following lemma will be used in the remainder of the proof.

Lemma 7. *Let $\pi(t) \stackrel{\mathbb{D}}{\sim} \text{Poisson}(p_t \cdot t)$ with the time-varying jump intensity p_t satisfying the conditions of Example 7, then for $0 \leq t \leq s \leq g \leq h \leq 1$,*

$$(a) \quad \mathbb{E} [\pi(t)\pi(s)] \leq Kt(1 + O(t) + O(s - t))$$

$$(b) \quad \mathbb{E} [\pi(t)\pi(s)\pi(g)\pi(h)] \leq Kt(1 + O(t) + O(s - t) + O(g - s) + O(h - g))$$

Proof. First, for (a), write

$$\mathbb{E} [\pi(t)\pi(s)] = \mathbb{E} [\pi(t)^2] + \mathbb{E} [\pi(t)] \mathbb{E} [(\pi(s) - \pi(t))]$$

by splitting terms and using independence of the Poisson increments. Then, by Dobinski's formula, the law of iterated expectations, and Lipschitz continuity and boundedness of p_t ,

$$\mathbb{E} [\pi(t)^2] = \mathbb{E} [\mathbb{E} [\pi(t)^2 | p_t]] = \mathbb{E} [\xi_t(1 + \xi_t)] \leq K_1 t(1 + K_1 t), \quad \mathbb{E} [\pi(t)] \leq K_2 t,$$

and $\mathbb{E} [(\pi(s) - \pi(t))] = \mathbb{E} [p_t(s - t)(1 + O(s - t))] \leq K_3(s - t)$. The result follows immediately by combining these bounds. Next, for (b), introduce the notation $\pi(s, t) = \pi(s) - \pi(t)$ for two arbitrary indices $0 \leq t \leq s \leq 1$, and write

$$\begin{aligned} \mathbb{E} [\pi(t)\pi(s)\pi(g)\pi(h)] = \\ \mathbb{E} [\pi(t) \times (\pi(t) + \pi(s, t)) \times (\pi(t) + \pi(s, t) + \pi(g, s)) \times (\pi(t) + \pi(s, t) + \pi(g, s) + \pi(h, g))] . \end{aligned}$$

Then, the final bound follows by writing out the parentheses, collecting powers and independent cross products, and repeatedly using the same arguments as for (a). $\square$

Let us now continue and establish bounds for the first two moments of $I_{\pi\pi}(\lambda_j)$. First, for the expected value, write

$$\begin{aligned} \mathbb{E} [I_{\pi\pi}(\lambda_j)] &= \frac{K}{2\pi} j^2 \int_0^1 \int_0^t \mathbb{E} [\pi(s)\pi(t)] \cos(2\pi j(s - t)) ds dt \\ &\quad + \frac{K}{2\pi} j^2 \int_0^1 \int_t^1 \mathbb{E} [\pi(s)\pi(t)] \cos(2\pi j(s - t)) ds dt \\ &\leq \frac{K}{2\pi} j^2 \int_0^1 \int_0^t K_1 s(1 + O(s) + O(t - s)) \cos(2\pi j(s - t)) ds dt \\ &\quad + \frac{K}{2\pi} j^2 \int_0^1 \int_t^1 K_2 t(1 + O(t) + O(s - t)) \cos(2\pi j(s - t)) ds dt \\ &\leq \frac{K_1}{2\pi} j^2 \int_0^1 \int_0^t s \cos(2\pi j(s - t)) ds dt + \frac{K_2}{2\pi} j^2 \int_0^1 \int_t^1 t \cos(2\pi j(s - t)) ds dt \leq \frac{K}{4\pi^3} \end{aligned}$$

using Lemma 7 (a) for the first inequality and integration of the cosine functions for the last inequality, similarly to Section B.7. The arguments to bound the variance are tedious. First, write

$$\begin{aligned}\mathbb{V}[I_{\pi\pi}(\lambda_j)] &\leq \left(\frac{Kj^4}{2\pi}\right) \mathbb{E} \left[\left(\int_0^1 \int_0^1 \pi(s)\pi(t) \cos(2\pi j(s-t)) ds dt \right)^2 \right] \\ &= \left(\frac{Kj^4}{2\pi}\right) \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbb{E}[\pi(s)\pi(t)\pi(g)\pi(h)] \cos(2\pi j(s-t)) \cos(2\pi j(g-h)) ds dt dg dh \\ &\leq \left(\frac{Kj^4}{2\pi}\right) \int_0^1 \int_0^1 \int_0^1 \int_0^1 K_1 \min(s, t, g, h) \cos(2\pi j(s-t)) \cos(2\pi j(g-h)) ds dt dg dh\end{aligned}$$

where the last inequality follows by applying Lemma 7 (b). Next, split the integrals up into intervals with $\min(s, t) \leq \min(g, h)$ and $\min(s, t) \geq \min(g, h)$, and let us consider the first interval. The analogous results for the second interval follow immediately by symmetry. Similarly, let us reduce the problem further and consider $s \leq t \leq (g, h)$ since the case $t \leq s \leq (g, h)$ is, again, symmetric. In this case, $\min(s, t, g, h) = s$ such that showing a bound for $\mathbb{V}[I_{\pi\pi}(\lambda_j)]$ amounts to showing that

$$j^4 \int_0^1 \int_0^t s \cos(2\pi j(s-t)) ds \times f(t, j) dt \tag{B.21}$$

is bounded, where $f(u, j)$ is a function that decomposes $f(t, j) = f_g(t, j) + f_h(t, j)$ with

$$f_g(t, j) = \int_t^1 \int_g^1 \cos(2\pi j(g-h)) dh dg = \frac{\sin^2(\pi j(t-1))}{2\pi^2 j^2} > 0, \quad \forall (t, j) \in [0, 1] \times \{1, \dots, \lfloor n/2 \rfloor\}$$

and

$$f_h(t, j) = \int_t^1 \int_h^1 \cos(2\pi j(g-h)) dg dh = \frac{\sin^2(\pi j(t-1))}{2\pi^2 j^2} > 0, \quad \forall (t, j) \in [0, 1] \times \{1, \dots, \lfloor n/2 \rfloor\}$$

corresponding to the cases $g \leq h$ and $h \leq g$, respectively. Computing the inner integral of the expression (B.21) yields

$$\int_0^t s \cos(2\pi j(s-t)) ds = \frac{\sin^2(\pi jt)}{2\pi^2 j^2},$$

implying that we may insert all three expressions into (B.21) to obtain

$$j^4 \int_0^1 \frac{\sin^2(\pi jt)}{2\pi^2 j^2} \times \frac{\sin^2(\pi j(t-1))}{\pi^2 j^2} dt = \frac{1}{2\pi^4} \int_0^1 \sin^4(\pi jt) dt$$

using $\sin(\pi jt) = -\sin(\pi j(t-1))$. Since

$$\frac{1}{2\pi^4} \int_0^1 \sin^4(\pi jt) dt = \frac{1}{2\pi^4} \left(\frac{12\pi j - 8\sin(2\pi j) + \sin(4\pi j)}{32\pi j} \right) = \frac{1}{2\pi^4} (3/8 + O(j^{-1})),$$

and the remaining terms follow similarly by symmetry, this provides the final result. $\square$.