AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

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

Bent Jesper Christensen and Rasmus T. Varneskov

CREATES Research Paper 2015-25

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

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

May 27, 2015

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 level shifts, i.e., structural changes in the means of the series, and deterministic trends. In particular, the proposed medium band least squares (MBLS) estimator uses sample 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, and participants in the Conference on Recent Developments in Financial Econometrics and Empirical Finance at the University of Essex, June 2014, for useful comments and suggestions. Financial support from the Department of Economics and Business and Aarhus Institute of Advanced Studies (AIAS), 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 make 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

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

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 formalized 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 (Duecker & 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)

[1]Strictly 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.

[2]Other 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. Moreover, they develop the necessary asymptotic central limit theory for maximum likelihood (ML) based inference such that the model may be formally analyzed.

Meanwhile, for the univariate case, a parallel literature has documented the presence 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 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 deterministic trends, possibly in conjunction with short memory dynamics, for a U.S. inflation rate series. Moreover, for example, Granger & Hyung (2004), Mikosch & Stărică (2004), Stărică & Granger (2005), Lu & Perron (2010), Perron & Qu (2010), and Varneskov & Perron (2015) 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), Hillebrand (2005), McCloskey & Perron (2013), and McCloskey & Hill (2014), this literature further 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, respectively, of the integration order upward, toward indicating fractional integration, and unit root tests toward non-rejection, thereby 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, 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 uses sample dependent trimming of the frequencies in the vicinity of the origin to asymptotically eliminate the impact of low-frequency contamination, utilizing 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 fully-modified NBLS estimator in the spirit of Phillips & Hansen (1990).

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

[4] Among 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 present framework with low-frequency contamination (see Corollary 1 below), and a similar result will hold for the FCVAR approach. Hence, to consistently analyze 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-consistent estimation of the cointegrating relations in the model.

Our solution is related to those in McCloskey & Perron (2013) and McCloskey & Hill (2014), 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, and achieve robustness to low-frequency contamination by 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. Hence, we are able to robustly assess whether the basic assumption of cointegration among the latent fractional signals is satisfied, and test whether low-frequency contamination does, in fact, exert an impact on existing estimators of the cointegrating vector.

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 by 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 the 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 empirical analysis. Section 6 concludes. The appendix contains proofs of the main asymptotic results in the paper. The following notation is used throughout: $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$, $o_p(\cdot)$ denote the usual (stochastic) orders of magnitude, which sometimes refer to matrices, vectors, and scalars; "$\sim$" means that the ratio of the left- and right-hand side tends to one in the limit; "$\overset{\mathbb{D}}{\sim}$" means distributed as; $\|\cdot\|$ is the Euclidean matrix norm; "$\to$ ", "$\xrightarrow{\mathbb{P}}$", and "$\xrightarrow{\mathbb{D}}$" indicate the limit, convergence in probability, and convergence in distribution, respectively.

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 $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t')'$, $t = 1, \ldots, n$, which is characterized by the decomposition

$$\boldsymbol{z}_t = \boldsymbol{c} + \boldsymbol{z}_t^* + \boldsymbol{u}_t, \tag{1}$$

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

$$y_t^* = \alpha + \boldsymbol{\beta}' \boldsymbol{x}_t^* + \epsilon_t^*, \tag{2}$$

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

$$y_t = a + \boldsymbol{b}' \boldsymbol{x}_t + e_t. \tag{3}$$

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 $\boldsymbol{x}_t^*$. 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 specific problem of estimating cointegrating relations of 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.*

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 (2014). Before proceeding, however, let

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

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

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

defines the cross-periodogram, with $\bar{\boldsymbol{w}}_k(\cdot)$ indicating complex conjugation of $\boldsymbol{w}_k(\cdot)$, and $\Re(\cdot)$ and $\Im(\cdot)$ the real and imaginary parts, respectively. We then characterize the low-frequency contaminating component $\boldsymbol{u}_t$ in (1) using the following high-level assumption:

Assumption 1. $\boldsymbol{u}_t \perp\!\!\!\perp \boldsymbol{z}_s^* \ \forall t, s$, $\mathbb{E}\left[\boldsymbol{I}_{uu}(\lambda_j)\right] = O(n/j^2)$, *and* $\lim_{n\to\infty} \mathbb{V}\left[(j^2/n)\|\boldsymbol{I}_{uu}(\lambda_j)\|\right] < \infty \ \forall j \neq 0$.

Example 1 (Bivariate Random Level Shifts)**.** *Consider a bivariate random level shift process* $(u_{1,t}, u_{2,t})$, $t = 1, \dots, n$, *defined as*

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

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

Lemma 1. *Suppose* $u_{1,t}$ *and* $u_{2,t}$ *follow the data generating process of Example 1. Then, for* $j = 1, \dots, \lfloor n/2 \rfloor$, *it follows that*

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

(2) $\lim_{n\to\infty} \mathbb{E}[(j^2/n)I_{u_2}(\lambda_j)] = (p_2\sigma^2_{\eta,2} + p_3\sigma^2_{\eta,3})/(4\pi^3)$ *and* $\lim_{n\to\infty} \mathbb{V}[(j^2/n)I_{u_2}(\lambda_j)] < \infty$.

(3) $\lim_{n\to\infty} \mathbb{E}[(j^2/n)I_{u_1u_2}(\lambda_j)] \leq (p_1\sigma^2_{\eta,1} + p_3\sigma^2_{\eta,3})^{1/2}(p_2\sigma^2_{\eta,2} + p_3\sigma^2_{\eta,3})^{1/2}/(4\pi^3)$ *and the limiting variance satisfies* $\lim_{n\to\infty} \mathbb{V}[(j^2/n)I_{u_1u_2}(\lambda_j)] < \infty$.

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

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. Along the same lines, a Cauchy-Schwarz inequality argument in conjunction with results from the univariate literature shows that Assumption 1 is satisfied when the low-frequency contamination process contains deterministic level shifts (Mikosch & Stărică (2004) and McCloskey & Perron (2013)), deterministic trends (Qu 2011), fractional trends, and outliers (Iacone 2010).[5] In addition, we note that Assumption 1 is sufficient, but not always necessary, for all these forms of contamination, and it may be relaxed for special cases. However, its general form 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.

3 Consistency

In this section, we motivate and introduce our medium band least squares estimator of the cointegrating vector $\boldsymbol{\beta}$ in the latent regression relation of interest (2), and we establish its consistency. To this end, we introduce the following assumptions in order to derive the theoretical results.

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

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

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

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

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

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

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

Assumptions 2 and 3 are similar to those provided 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, $\boldsymbol{z}_t^*$, rather than the observable $\boldsymbol{z}_t$ in (1), consistent with the nature of the estimation problem at hand. Furthermore, Assumption 2 generalizes the corresponding assumption in the two earlier articles by using the spectral representation analyzed by, e.g., Shimotsu (2007) and Robinson (2008) in the context of local Whittle estimation for multivariate stationary fractionally integrated processes. For example, this allows the latent process of interest $\boldsymbol{z}_t^*$ to exhibit vector fractional ARIMA dynamics with non-zero coherence between its individual elements at the origin. The potential presence of low-frequency contamination in our case, however, induces a bias of the (co-)peridogram in the *close* vicinity of the origin, which is highlighted by the decomposition

$$\boldsymbol{I}_{zz}(\lambda_j) = \boldsymbol{I}_{zz}^*(\lambda_j) + \boldsymbol{I}_{uu}(\lambda_j) + 2\boldsymbol{I}_{zu}^*(\lambda_j) = O_p((n/j)^{2d_\top}) + O_p(n/j^2) + O_p((n/j)^{1+d_\top}n^{-1/2}) \quad (4)$$

as $\lambda_j \to 0^+$, where the $(^*)$ superscript is used as generic notation for a (co-)periodogram or spectral density for the latent $\boldsymbol{z}_t^*$. The component $\boldsymbol{u}_t$, thus, dominates the peridogram for frequencies λ_j such that $j = o(n^{(1-2d_\top)/(2-2d_\top)})$ and vice versa for $jn^{(2d_\perp 1)/(2-2d_\perp)} \to \infty$, see McCloskey & Perron (2013) for a detailed discussion. We note, however, 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 $\boldsymbol{\beta}$ in (2) by carefully selecting which frequency ordinates to include in the analysis. Finally, we should, strictly speaking, make distinctions between the asymptotic order of the low-frequency contamination bias for each entry of $\boldsymbol{I}_{zz}(\lambda_j)$, depending on the size of $d_i \in (0, 1/2)$, $i = 1, \ldots, p$. However, as the bias is strictly decreasing in d_i, we will use the generic notation $d_\perp$ and $d_\top$ to indicate the largest bias since we seek to eliminate the impact of low-frequency contamination for *all* elements of $\boldsymbol{I}_{zz}(\lambda_j)$.

From (4), as the bias due to low-frequency contamination is particularly pronounced for frequencies in the close vicinity of the origin, this motivates the introduction of the trimmed discretely averaged co-periodogram,

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

which we use to define our trimmed narrow-band least squares (NBLS) estimator, labelled the medium band least squares (MBLS) estimator, as

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

The trimming function, ℓ, is crucial for selecting the frequency range where $\boldsymbol{I}_{zz}(\lambda_j)$ is asymptotically free from low-frequency contamination bias, such that we may consistently estimate $\boldsymbol{\beta}$, and form asymptotically normal inference. While NBLS is recovered as the special case $\ell = 1$, we note that, in general, more sophisticated conditions on the trimming function are required to obtain consistency and asymptotic normality. These will be stated formally below.[6]

Finally, let us introduce the continuously averaged spectral density matrix $\boldsymbol{F}^*_{zz}(\lambda) = \int_0^\lambda \Re(\boldsymbol{f}^*_{zz}(s))ds$, which is important for establishing our asymptotic results below, and appears in a scaled version in the asymptotic variance of the MBLS estimator in our central limit theorem in Section 4. Under Assumptions 2 and 3, and from the results, e.g., in Robinson & Marinucci (2003) and Shimotsu (2007), the local asymptotic behavior of this integral satisfies

$$\boldsymbol{F}^*_{zz}(\lambda) \sim \boldsymbol{G}(\lambda) \qquad \text{as} \quad \lambda \to 0^+,$$

where $\boldsymbol{G}(\lambda)$ has (i,k)-th element $\boldsymbol{G}(\lambda,i,k) = G_{i,k}\lambda^{1-d_i-d_k}/(1-d_i-d_k)C_{i,k}$, with $G_{i,k}$ the (i,k)-th element of $\boldsymbol{G}$ and $C_{i,k} = \cos(\pi(d_i - d_k)/2)$ an adjustment factor for the case $d_i \neq d_k$, reflecting the possibility of a complex component in the spectral density matrix at the origin. We are now ready to establish consistency of the MBLS estimator:

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

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

(b) *If additionally $1/\ell \to 0$, then $\hat{\boldsymbol{\beta}}(\ell,m) \xrightarrow{\mathbb{P}} \boldsymbol{\beta}$.*

Proof. See Appendix A.1. □

Theorem 1 shows that if the trimming function $\ell \to \infty$, but more slowly than the bandwidth, m, then the MBLS estimator is consistent for $\boldsymbol{\beta}$. Indeed, trimming is necessary, as highlighted by the following corollary:

Corollary 1. *Under the conditions of Theorem 1 (a), if $\ell = O(1)$ and $\boldsymbol{\beta} \neq \boldsymbol{b}$, then $\hat{\boldsymbol{\beta}}(\ell,m)$ is inconsistent for $\boldsymbol{\beta}$.*

Corollary 1 highlights the general inconsistency of NBLS in the presence of low-frequency contamination, and similarly holds for the band spectrum regression estimators in Engle (1974) and Yamamoto & Perron (2013), even though they discard a fixed, respectively, increasing number of frequencies in the vicinity of the origin. For the former, this is immediately evident, as the estimator dictates $\ell = O(1)$ and $m/n = O(1)$. The trimmed band spectrum regression estimator in the latter, on the other hand, allows $1/\ell \to 0$ and $\ell/m \to 0$, 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, $\boldsymbol{z}^*_t$, nor cointegration among its elements.

The condition $\boldsymbol{\beta} \neq \boldsymbol{b}$ rules out the case where the low-frequency contamination is "cointegrated" in exactly the same way as the latent fractional signals. The decomposition of errors in Theorem 1 (a) reflects the tradeoff that arises from trimming frequencies. The first term measures the loss of

[6]Formally, $\hat{\boldsymbol{\beta}}(1,m) \equiv \hat{\boldsymbol{\beta}}_{NBLS}$, and OLS is algebraically equivalent to $\hat{\boldsymbol{\beta}}(1,n-1) \equiv \hat{\boldsymbol{\beta}}_{OLS}$.

information, and the second term the bias from low-frequency contamination. The third term mirrors the error in the uncontaminated case considered by Robinson & Marinucci (2003). The second term in the decomposition, thus, corroborates the heuristic arguments made above by highlighting the importance of trimming frequencies to avoid the bias from low-frequency contamination, manifesting itself in the discretely averaged co-periodogram and causing inconsistency of NBLS.

Remark 2. *The fact that we need the trimming parameter ℓ to diverge is consistent with the findings of Iacone (2010), McCloskey & Perron (2013), McCloskey & Hill (2014), and McCloskey (2013), 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.*

4 Asymptotic Normality

We proceed with our analysis and establish asymptotic normality of our MBLS estimator of the cointegrating vector, $\boldsymbol{\beta}$, in the presence of low-frequency contamination. For simplicity, and in the remainder of the paper, we will impose equality of the integration orders for the variables in $\boldsymbol{z}_t^*$, that is, $d_i = d \in (0, 1/2)$, $i = 1, \ldots, p$, and subsequently discuss our reasoning behind this. As in the frameworks without contamination, e.g., Lobato (1999), Robinson & Marinucci (2003), and Christensen & Nielsen (2006), we will strengthen our assumptions on the latent variables and restate them in terms of $\boldsymbol{w}_t^* = ((\boldsymbol{x}_t^*)', \epsilon_t^*)'$ instead of $\boldsymbol{z}_t^*$. Finally, 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 $\boldsymbol{w}_t^*$, $t = 0, \pm 1, \ldots,$ is covariance stationary with spectral density matrix satisfying $\boldsymbol{f}_{ww}^*(\lambda) \sim \boldsymbol{\Lambda G \Lambda}$ as $\lambda \to 0^+$. In particular, $\boldsymbol{\Lambda} = \operatorname{diag}[\lambda^{-d}, \ldots, \lambda^{-d}, \lambda^{-d_\epsilon}]$ is $p \times p$ with $0 \leq d_\epsilon < d$, there exists a $\varphi \in (0, 2]$ such that $|\boldsymbol{f}_{ww}^*(\lambda, i, k) - G_{i,k}\lambda^{-2d}| = O(\lambda^{\varphi - 2d})$ as $\lambda \to 0^+$ for $i, k = 1, \ldots, p-1$, where $\boldsymbol{f}_{ww}^*(\lambda, i, k)$ is the (i, k)-th element of $\boldsymbol{f}_{ww}^*(\lambda)$, $|\boldsymbol{f}_{ww}^*(\lambda, p, p) - G_{p,p}\lambda^{-2d_\epsilon}| = O(\lambda^{\varphi - 2d_\epsilon})$ as $\lambda \to 0^+$, and $|\boldsymbol{f}_{ww}^*(\lambda, i, p) - G_{i,p}\lambda^{-d_c}| = O(\lambda^{\varphi - d_c})$ as $\lambda \to 0^+$. Finally, $G_{i,p} = G_{p,i} = 0$, and the remaining terms $G_{i,k}$, $i, k = 1, \ldots, p-1$, and $G_{p,p}$ correspond to the (i, k)-elements of $\boldsymbol{G}$, which has the same properties as in Assumption 2.*

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

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

Assumption 5. *The bandwidth m and trimming ℓ satisfy*

$$\frac{1}{m} + \frac{m^{1+2\varphi}}{n^{2\varphi}} + \frac{\ell}{m} + \frac{n^{1-d_c}}{m^{1/2-d_c}} \frac{1}{\ell^2} + \frac{\ell^{1+d_c+\varphi}}{m^{1/2+d_c} n^{\varphi}} \to 0 \qquad \text{as} \quad n \to \infty.$$

Assumptions 2', 3', and 4 resemble the assumptions in Christensen & Nielsen (2006), and they share similarities with those in Robinson (1995*a*) and Lobato (1999), who analyze uni- and multivariate local Whittle estimation of the fractional integration order. In particular, we note that for a special case of the smoothness condition in Assumption 2', $\varphi = 2$, $\boldsymbol{w}_t^*$ follows a vector fractional ARIMA process.

The zero-coherence condition $G_{i,p} = G_{p,i} = 0$ in Assumption 2' is important for our central limit theorem below, and to maintain this, we need equality of integration orders, as emphasized by, e.g., Shimotsu (2007) and Nielsen & Frederiksen (2011).[7] Intuitively, it 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., common breaks (see Example 1), common deterministic trends, and similar behavior of the observed series, which by Corollary 1 leads to inconsistency of standard estimators of the fractional cointegration vector. Finally, we emphasize that the assumption of equality of memory is common to the literature testing for fractional cointegration rank, see, e.g., Robinson & Yajima (2002) and Nielsen & Shimotsu (2007).

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 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 $1/\ell \to 0$, and the fifth term the loss of information from trimming frequencies. We defer specific recommendations for the selection of ℓ and m to Section 5.1.

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

$$\boldsymbol{H}_{xx}(\lambda) = \frac{\lambda^{1-2d}}{1-2d}\boldsymbol{G}_{xx}, \quad \boldsymbol{H}_{xx} = \boldsymbol{H}_{xx}(\lambda)\lambda^{2d-1}, \quad \boldsymbol{J}_{xx} = \frac{1}{2(1-2d_c)}\boldsymbol{G}_{xx}. \tag{7}$$

Theorem 2. *Under Assumptions 1, 2', 3', 4, and 5 along with* $0 \le d_\epsilon < d_c < 1/2$,

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

Proof. See Appendix A.2. □

Theorem 2 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

[7] Too see this, note that when comparing the cross-spectral densities at the origin in Assumptions 2 and 2', it follows by setting $d_i = d \in (0, 1/2)$ and $G_{i,p} = G_{p,i} = 0$ for all i elements that the (complex-valued) representation in the former becomes real-valued and equivalent to the corresponding representation in the latter. Hence, in this case, the implicit requirement in Assumption 2' that cross-autocorrelations are symmetric with respect to time is consistent with the one-sided moving average representations of the latent series in both Assumptions 3 and 3'.

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. Moreover, rewriting the scaled asymptotic variance of the MBLS estimator from Theorem 2 as

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

illustrates how the accuracy of the MBLS inference procedure depends on the noise-to-signal ratio of the latent long-run variance of $(1-L)^{d_\epsilon}\epsilon_t^*$, L being the lag operator, to the latent long-run covariance matrix of $(1-L)^d\boldsymbol{x}_t^*$. 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. However, the result in the latter is established under much stricter assumptions than in the present paper, for example, without consideration of low-frequency contamination in the observed series, and with neither fractional integration nor cointegration.

Remark 3. *Despite the fact that we need the trimming parameter ℓ to diverge for consistency and asymptotic normality, similarly to, e.g., McCloskey (2013) and McCloskey & Hill (2014), the rate requirement for Theorem 2 differs quite suggestively from corresponding rates in the literature. To see this, suppose $\ell = n^\nu$ and $m = n^\kappa$. Then Assumption 5 dictates*

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

thus with lower- and upper bounds that depend on the size of the bandwidth, κ, and the collective memory of the system, d_c. Notice that the trimming requirement (i.e., the lower bound) is generally decreasing in both κ and d_c. The corresponding bounds in McCloskey (2013) and McCloskey & Hill (2014) 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 former and a semiparametric estimate 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 not only depends 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. Similar comparisons can be made with the univariate estimators in Iacone (2010) and McCloskey & Perron (2013). This is carried out implicitly, however, as part of our discussion of feasible inference procedures next.

4.1 Robust Feasible Inference

So far, we have assumed that d and d_ϵ are known. However, knowledge of the lower and upper bounds on the rate of required trimming, along with feasible inference on the MBLS estimator in Theorem 2, depend on consistent estimates of these fractional integration orders. For this purpose, we apply the trimmed local Whittle (LW) estimator introduced by Iacone (2010), and refined in McCloskey

& Perron (2013), to each element of the multivariate process $\boldsymbol{z}_t$ separately, to obtain estimates $\hat{d}_i$, $i = 1, \dots, p$, and then compute the average $\hat{d} = p^{-1} \sum_{i=1}^{p} \hat{d}_i$ as a global estimate, following Robinson & Yajima (2002) and Nielsen & Shimotsu (2007). In particular, for a generic univariate series z_t in the vector $\boldsymbol{z}_t$, where $z_t^* \in I(d)$, the trimmed LW estimator is defined by

$$d_{\text{LW}}(z, \ell_1, m_1) = \underset{\delta \in [0,1/2)}{\operatorname{argmin}} \log \hat{G}(\delta, z, \ell_1, m_1) - \frac{2\delta}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \log \lambda_j, \tag{8}$$

where

$$\hat{G}(\delta, z, \ell_1, m_1) = \frac{1}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \lambda_j^{2\delta} \Re \left(I_{zz}(\lambda_j) \right), \tag{9}$$

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 and the previous section, 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 $\boldsymbol{z}_t$ 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. 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, then if additionally*

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

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

Proof. Under the stated assumptions, the results follow by Iacone (2010, Theorems 2 and 3). □

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 since the signal becomes more detectable the higher its fractional integration order. 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 which is robust to low-frequency contamination, and of a 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),

$$\hat{e}_t = y_t - \hat{\boldsymbol{\beta}}(\ell, m)'\boldsymbol{x}_t, \tag{10}$$

differ from the true errors, ϵ_t^*, not only due to low-frequency contamination in y_t and $\boldsymbol{x}_t$, but also since the MBLS estimate, $\hat{\boldsymbol{\beta}}(\ell, m)$, may differ from the true $\boldsymbol{\beta}$ 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, 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* (10) *satisfy either*

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

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

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 $n \to \infty$. 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 3. *Under Assumptions 1, 2', 3', and 4, then if additionally*

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

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

Proof. See Appendix A.3. □

Theorem 3 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 3 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 $\boldsymbol{G}_{xx}$ and $G_{p,p}$

As a final step in developing feasible inference procedures for Theorem 2, we also need consistent estimates of $\boldsymbol{G}_{xx}$ and $G_{p,p}$. Hence, we generalize the representation in (9) to accommodate a generic vector time series, $\boldsymbol{h}_t$, $t = 1, \dots, n$, with fractional integration order $d_h \in [0, 1/2)$, i.e.,

$$\hat{\boldsymbol{G}}(d_h, \boldsymbol{h}, \ell_2, m_2) = \frac{1}{m_2 - \ell_2 + 1} \sum_{j=\ell_2}^{m_2} \lambda_j^{2d_h} \Re \left(\boldsymbol{I}_{hh}(\lambda_j) \right), \tag{11}$$

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 their cross-restrictions.

Assumption 5-G. *The bandwidth m_2 and trimming ℓ_2 satisfy either*

(a) $(\log m_2)^2/m_2 + (m_2/n)^{\varphi}(\log m_2) + 1/\ell_2 + \ell_2/m_2 + (n/\ell_2)^{1-2d}(\ell_2 m_2)^{-1} \to 0$, *or*

(b) $(\log m_2)^2/m_2 + (m_2/n)^{\varphi}(\log m_2) + 1/\ell_2 + \ell_2/m_2 + (n/\ell_2)^{1-2d_\epsilon}(\ell_2 m_2)^{-1} \to 0$, *or*

(c) *Assumption 5-G (b) and* $(\log n)^4(\log m_2)(m/m_2)^{d-d_\epsilon} \to 0$, *as* $n \to \infty$.

Assumption 5-G imposes mild restrictions on the required trimming, ℓ_2, which depends on the memory of the underlying process. The extra condition in Assumption 5-G (c) relative to 5-G (b) corresponds to the one imposed in Assumption 5-LW' (a), that is, we need an additional restriction on the bandwidth parameter m_2 to account for the pre-estimated cointegrating vector. The split into Assumptions 5-G (b) and 5-G (c) reflects our use of an intermediate step in establishing consistency of $\hat{\boldsymbol{G}}(d_\epsilon, \hat{e}, \ell_2, m_2)$ for $G_{p,p}$. First, we establish consistency of $\hat{\boldsymbol{G}}(d_\epsilon, \epsilon, \ell_2, m_2)$ for $G_{p,p}$ using the latent variable $\epsilon_t = y_t - \boldsymbol{\beta}'\boldsymbol{x}_t$, which differs from the observable $\hat{e}_t$ by not having the cointegrating vector, $\boldsymbol{\beta}$, consistently pre-estimated. The final step is, then, provided using arguments similar to those for Theorem 3 to account for the additional sampling errors in $\hat{e}_t$.

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

(a) *Assumption 5-G (a) holds, then* $\hat{\boldsymbol{G}}(d, \boldsymbol{x}, \ell_2, m_2) \xrightarrow{\mathbb{P}} \boldsymbol{G}_{xx}$.

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

(c) *Assumption 5-G (c) holds, then* $\hat{G}(d_\epsilon, \hat{e}, \ell_2, m_2) \xrightarrow{\mathbb{P}} G_{p,p}$.

Proof. See Appendix A.4. □

Hence, we can estimate d and d_ϵ consistently by Lemma 2 and Theorem 3, along with $\boldsymbol{G}_{xx}$ and $G_{p,p}$ using Theorem 4, and thereby obtain consistent estimates of $\boldsymbol{H}_{xx}$ and $\boldsymbol{J}_{xx}$ in (7) by the continuous mapping theorem. This enables us to perform feasible inference for the MBLS estimator in the presence of low-frequency contamination using Theorem 2.

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

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

where, again, $\boldsymbol{z}_t^* \in I(d)$ 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 $b > 0$. Second, rejection in favor of $d - d_\epsilon > b > 0$ indicates a stronger fractional cointegrated relation. Ideally, we would like to test $\mathcal{H}_0$ for the case $b = 0$, 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 2, and orthogonality of $\boldsymbol{x}_t^*$ and ϵ_t^*, which hold only under fractional cointegration, and thus requires $b > 0$ (Assumption 2' would be violated for $b = 0$). Informally, failure to reject for a very small value of b could provide indicative evidence against cointegration.

For ease of exposition, we introduce the shorthand notation $\hat{d}_M = d_{\mathrm{LW}}(z, \ell_1, m_1)$ and $\hat{d}_{\epsilon,M} = d_{\mathrm{LW}}(\hat{e}, \ell_1, m_1)$ for the trimmed LW estimators of the fractional integration orders of $\boldsymbol{z}_t^*$ and ϵ_t^*, respectively, using the MBLS residuals. By Lemma 2 and Theorem 3, both are robust to low-frequency contamination. Hence, we may introduce the following test.

Corollary 2. *Under Assumptions 1, 2', 3', 4, and 5-LW',*

$$t_b = m_1^{1/2} 2(\hat{d}_M - \hat{d}_{\epsilon,M} - b) \xrightarrow{\mathbb{D}} N(0,1). \tag{12}$$

We assess the power properties of this test for different values of b, cointegration strength $d - d_\epsilon$, bandwidth, and sample size in Section 5.1. Furthermore, we illustrate how the test may be applied as a diagnostic tool in our empirical analysis in Section 5.3 below.

4.2.2 Testing Consistency of NBLS

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

$$(1, -\boldsymbol{\beta}')\boldsymbol{f}_{zz}^*(\lambda)(1, -\boldsymbol{\beta}')' = f_\epsilon(\lambda) \sim g\lambda^{-2d_\epsilon}, \qquad \text{as} \quad \lambda \to 0^+, \tag{13}$$

for $0 \leq d_\epsilon < d$. Since $\boldsymbol{f}^*_{zz}$ is positive definite and the lower $(p-1) \times (p-1)$ submatrix of $\boldsymbol{G}$ 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 $\boldsymbol{\beta}$. An inconsistent estimator of $\boldsymbol{\beta}$ 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 $\boldsymbol{\beta}$ 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 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 $c \in (0,1)$, when estimating the fractional integration order of the NBLS residuals, while keeping the bandwidth m_1 in the trimmed LW estimator applied to the MBLS residuals unchanged. Formally, we write $\hat{d}_{\epsilon,N}(1,c)$ for the resulting estimator.

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

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

Corollory 3 highlights the importance of slowing down the rate of convergence of $\hat{d}_{\epsilon,N}(c)$ 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 (13), one source of inconsistency arises from NBLS being inconsistent for $\boldsymbol{\beta}$. Another source is from 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 4. *Under Assumptions 2', 3', 4, 5-LW' and* $\boldsymbol{u}_t = \boldsymbol{0}\ \forall t = 1, \ldots, n$,

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

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. A detailed assessment of this issue, however, is deferred to Section 5.1.

5 Implementation and Application

In this section, we first discuss implementation details for the MBLS framework, including specific choices of tuning parameters and applications of the testing procedures introduced in Section 4.2. This is followed by an empirical analysis where we provide new perspectives on volatility factors using long-span realized variance series for S&P 500 equities.

5.1 Implementation

So far, 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 $\hat{\boldsymbol{G}}(d_h, \boldsymbol{h}, \ell_2, m_2)$ from (11), we need to provide some more specific guidelines for their selection.

5.1.1 Discussion of Tuning Parameters

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 $n \to \infty$. For simplicity of exposition, we let $\kappa_1 = \kappa_2$. The asymptotic theory, cf. Assumption 5-LW' for Theorem 3, then dictates the bound $\kappa < \kappa_1 < 2\varphi/(1+2\varphi)$. The latter simplifies by assuming that the latent process of interest $\boldsymbol{w}_t^*$, 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 either explicitly on d and d_ϵ, or implicitly through d_c. We select the trimming rate for the local Whittle estimator according to the simple, and conservative, rule-of-thumb $\nu_1 = 1/2 + \zeta$, $\zeta = 0.05$, proposed and extensively studied by McCloskey & Perron (2013), which is valid irrespectively of the value of $d \in (0, 1/2)$. This implies $\kappa_1 \in (\kappa \wedge 0.55, 4/5)$.

Finally, we note that for $\varphi = 2$, the upper bound on ν is always determined by κ.[8] 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 $\nu > 0.35$.

5.1.2 Size and Power of Empirical Diagnostic Tools

We investigate the size and power characteristics of the two testing procedures introduced in Section 4.2 by simulating the empirical rejection rates of the tests statistics in Corollary 2 and 3, respectively, according to the theoretical properties of $\hat{d}_M$, $\hat{d}_{\epsilon,M}$, and $\hat{d}_{\epsilon,N}(1, c)$.

[8] This 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 $\varphi = 2$.

First, for the test of cointegration strength, t_b from Corollary 2, we simulate the empirical memory gap as

$$\hat{d}_M - \hat{d}_{\epsilon,M} = d - d_\epsilon + m_1^{1/2}/2\eta_1 - m_1^{1/2}/2\eta_2,$$

for sample sizes $n \in \{500, 2000\}$, a bandwidth $m_1 = n^{\kappa_1}$ for which $\kappa_1 \in \{0.6, 0.7, 0.8\}$, and cointegration strength $d - d_\epsilon \in \{0.05, 0.1, 0.2, 0.4\}$, where η_1 and η_2 denote two independent standard Gaussian random variables. The choices of sample size correspond well with the monthly and weekly series, respectively, used in the empirical analysis below, and κ_1 with the bounds on the bandwidth parameter discussed above, which suggest $\kappa_1 \in (0.6, 0.8)$ when $\kappa = 0.6$. Finally, the memory gaps cover both the case of a fairly strong cointegration, $d - d_\epsilon = 0.4$, and almost no cointegration, $d - d_\epsilon = 0.05$. The latter is representative of the empirically relevant case where failure to reject t_b for a small value of b may provide indicative evidence against cointegration. We have divided and displayed the simulated empirical rejection rates for two different intervals, $b \leq d - d_\epsilon$ and $b \geq d - d_\epsilon$, in Figures 1 and 2, respectively. Rejection of the former is of particular interest, as it provides evidence of a stronger fractionally cointegrated relation, assuming $b > 0$. Finally, note that the x-axis in each figure has been normalized by $d - d_\epsilon$, and that rejection is for a two-sided test at a 5% significance level.

From Figure 1, we observe, not surprisingly, that the empirical rejection rates are uniformly increasing as the relative distance $b/(d - d_\epsilon)$, measured on the x-axis, decreases, since b moves further below the true cointegration strength. Moreover, it illustrates that for a fixed ratio, $b/(d - d_\epsilon)$, the power of the test is uniformly increasing in $d - d_\epsilon$. Similarly, in Figure 2, we observe that power is increasing in $(b - (d - d_\epsilon))/d_U$, with $d_U = 0.49$ fixed as the upper bound on d, since b increases relative to the true cointegration strength $d - d_\epsilon$. In contrast to Figure 1, however, for a fixed $(b - (d - d_\epsilon))/d_U$, the power of the test is uniformly decreasing in $d - d_\epsilon$. Collectively, this demonstrates that the test performs better the larger the distance between b and $d - d_\epsilon$, and it suggests lower power for rejecting the boundary hypotheses $d - d_\epsilon - b > 0$ when $d - d_\epsilon$ is small and $d - d_\epsilon - b < 0$ when $d - d_\epsilon$ is large. Finally, we observe from both figures that the empirical rejection rates are increasing in both n and κ_1. This is not surprising, as both increase the effective sample size $m_1 = n^{\kappa_1}$.

We perform a similar size and power analysis for the consistency test of NBLS inference, $t_d(c)$ from Corollary 3, by simulating the empirical memory gap as

$$\hat{d}_{\epsilon,N}(1, c) - \hat{d}_{\epsilon,M} = d_{\epsilon,N} - d_{\epsilon,M} + (cm_1)^{1/2}/2\eta_1 - m_1^{1/2}/2\eta_2,$$

where m_1 is chosen as above, $d_{\epsilon,N} - d_{\epsilon,M} \in \{0, 0.05, 0.1, 0.2\}$, and $c \in (0, 1)$. In particular, we consider the empirical rejection rate from implementing a one-sided test at a 5% significance level as a function of c in Figure 3. Thus, the results for $d_{\epsilon,N} - d_{\epsilon,M} = 0$ show the size of the test statistic $t_d(c)$, and $d_{\epsilon,N} - d_{\epsilon,M} > 0$ the power.

Figure 3 clearly illustrates the size-power tradeoff induced by c; the power curves increase uniformly as c increases, but so do size distortions. Hence, the researcher must choose c to strike a balance between the two different rejection rates. Intuitively, this suggests maximizing the distance between the size and power curves in Figure 3. Hence, from an informal gauge, we select $c \in \{0.5, 0.7\}$ for the empirical analysis, but emphasize that conclusions must be interpreted with care, since this procedure is slightly liberal on size. Finally, once more, the properties of the test improve as κ_1 and n are increased.

5.2 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), Andersen, Bollerslev, Diebold & Ebens (2001), and Andersen, Bollerslev, Diebold & Labys (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.[9] A similar factor decomposition arises from a different literature, which considers estimation of continuous time diffusion models based on discrete time data.[10] 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 in the mean of the process, that is, by a type of low-frequency contamination.

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

$$V_{i,t} = c_i + V_{i,t,1} + V_{i,t,2} + \boldsymbol{\iota}'\boldsymbol{u}_{i,t} \tag{14}$$

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 $\boldsymbol{u}_{i,t}$ as a $\tau \times 1$ vector containing 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 in, for example, McCloskey & Perron (2013) and McCloskey & Hill (2014), we see that $\boldsymbol{u}_{i,t}$ not only accommodates features such as random level shifts and deterministic trends in the log-variance process, but also many types of finite activity Lévy innovations, which have been shown to improve the volatility model fit in, e.g., Eraker, Johannes & Polson (2003) and Todorov (2011).

Next, let us combine the insights from (14) 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 $r_{i,t} = \alpha_i + \lambda_i r_{M,t} + \boldsymbol{\Lambda}_{\boldsymbol{i}}'\boldsymbol{f}_t + \eta_{i,t}$, with $r_{M,t}$ the excess return on the market portfolio, $\boldsymbol{f}_t$ a $k \times 1$ vector of additional pricing factors *orthogonal* to $r_{M,t}$, $\eta_{i,t}$ 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 an affine

[9] This is, for example, the case when fitting fractional ARIMA models to the volatility series, e.g., Andersen, Bollerslev, Diebold & Labys (2003), Koopman, Jungbacker & Hol (2005), Christensen & Nielsen (2007), Chiriac & Voev (2011), 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 discrete time stochastic volatility models with long memory, e.g., Breidt, Crato & de Lima (1998), Hurvich, Moulines & Soulier (2005), and Deo, Hurvich & Lu (2006).

[10] See, among others, Gallant, Hsu & Tauchen (1999), Bates (2000), Duffie, Pan & Singleton (2000), Chernov, Gallant, Ghysels & Tauchen (2003), Christoffersen, Heston & Jacobs (2009), Todorov, Tauchen & Grynkiv (2011), Andersen, Fusari & Todorov (2014), and many references therein.

function of the factor- and idiosyncratic variances. This decomposition in terms of factor variances motivates the alternative volatility model

$$V_{i,t} = c_i + \beta_i V_{M,t,1} + V_{i,t,2} + \boldsymbol{\iota}'\boldsymbol{u}_{i,t}, \tag{15}$$

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. Besides the affine dependence on $V_{M,t,1}$, the model (15) shares all the generic features of (14). 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.

5.3 Empirical Analysis of Volatility Factors

We analyze MBLS and NBLS estimation of β_i in (15) 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. We will, thus, 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 (ACF's) in Figure 4. 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, the latter may be caused by either a persistent long memory component, a combination of structural breaks and short memory dynamics, or both. Hence, it is not surprising that all series display slowly decaying ACF's. We proceed by considering standard descriptive statistics of the series, provided in Table 2. 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 2 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 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 (15) for each of the three stocks, using both NBLS with a bandwidth $\kappa = 0.6$ and MBLS with an additional trimming parameter, ν. Specifically, we consider three different selections, $\nu \in \{1/3 + \zeta, 2/5 + \zeta, 1/2 + \zeta\}$, $\zeta = 0.05$, which all satisfy the conservative restriction $\nu > 0.35$. 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 b when implementing the test of cointegration strength, t_b, we utilize the empirically estimated cointegration strength, $\hat{d} - \hat{d}_\epsilon$, and the asymptotic distribution from Corollary 2 to back out the upper (right-tail) critical values of b 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 3 and 4 for the monthly and weekly series, respectively.

Table 3 contains several interesting results. First, the MBLS estimates of β_i are uniformly above the corresponding NBLS estimates for all three equities and often by more than two standard errors, judging by the MBLS inference.[11] Hence, low-frequency contamination introduces a downward bias in the NBLS estimator, understating the magnitude of market volatility spillover into the individual equities. This may have dramatic consequences, e.g., 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, he or she will generally be under-hedged. 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 on MBLS, indicating the presence of cointegration, it is small and close to zero when using NBLS, suggesting the opposite conclusion. These estimates also show that the condition $\hat{d}+\hat{d}_\epsilon < 1/2$ is violated for NBLS, which is the reason for not reporting standard errors with the NBLS estimates. Third, the critical values of b, backed out from the corresponding empirical estimates, substantiate the claim of cointegration based on the MBLS results, as they are positive at the 1% significance level for IBM and KO, and at the 10% level for WMT. Fourth, we strongly reject consistency of the NBLS inference procedure at all conventional significance levels, the smallest value of $t_d(c)$ being larger than 7. Hence, even if the size of the test is slightly liberal, this provides compelling evidence in favor of applying the proposed MBLS inference procedure to analyze volatility coherence. Finally, when decomposing the source of NBLS inconsistency using the

[11] We have also run the corresponding OLS regressions. The results are highly similar to the NBLS results and are therefore omitted for brevity.

trimmed LW estimator to estimate the respective fractional integration orders for $\tilde{t}_d(c)$, as described in Corollary 4, the evidence against consistency of the NBLS estimates of β_i remains strong.

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 NBLS. Hence, our analysis of the proposed volatility model (15) 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 4, 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 condition for inference, $\hat{d} + \hat{d}_\epsilon < 1/2$, is violated in all cases. Hence, no standard errors are reported. However, Theorem 1 demonstrates that the MBLS estimates remain consistent in this case. Furthermore, based on the point estimates, all qualitative conclusions pertain to the weekly series, as well.

6 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 deterministic trends. It uses sample dependent trimming of frequencies 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 develop a rigorous feasible inference procedure and two empirical diagnostic tools; a test of cointegration strength and a test of inconsistency of NBLS inference.

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	$\ell = n^{\nu}$	$0 < \nu$	$\nu < \kappa$	-
LW' bandwidth	$m = n^{\kappa_1}$	$0 < \nu_1 < \kappa_1$	$\kappa_1 < 1$	$\kappa < \kappa_1$
LW' trimming	$\ell = n^{\nu_1}$	$\max\left\{0, 1 - \frac{\kappa_1}{1-2d}\right\} < \nu_1$	$\nu_1 < \kappa_1$	-
$\boldsymbol{G}_{xx}$ bandwidth	$m = n^{\kappa_2}$	$0 < \nu_2 < \kappa_2$	$\kappa_2 < 1$	-
$\boldsymbol{G}_{xx}$ trimming	$\ell = n^{\nu_2}$	$\max\left\{0, \frac{1-2d-\kappa_2}{2(1-d)}\right\} < \nu_2$	$\nu_2 < \kappa_2$	-
$G_{p,p}$ bandwidth	$m = n^{\kappa_2}$	$0 < \nu_2 < \kappa_2$	$\kappa_2 < 1$	$\kappa < \kappa_2$
$G_{p,p}$ trimming	$\ell = n^{\nu_2}$	$\max\left\{0, \frac{1-2d_{\epsilon}-\kappa_2}{2(1-d_{\epsilon})}\right\} < \nu_2$	$\nu_2 < \kappa_2$	-
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	$\ell = n^{\nu}$	$\frac{(1-d_c)(1-\kappa)}{2} + \frac{\kappa}{4} < \nu$	$\nu < \min\left\{\frac{\varphi+\kappa(1/2+d_c)}{1+d_c+\varphi}, \kappa\right\}$	-
LW' bandwidth	$m = n^{\kappa_1}$	$0 < \nu_1 < \kappa_1$	$\kappa_1 < 2\varphi/(1+2\varphi)$	$\kappa < \kappa_1$
LW' trimming	$\ell = n^{\nu_1}$	$\max\left\{0, 1 - \frac{\kappa_1}{2(1-2d)}\right\} < \nu_1$	$\nu_1 < \kappa_1$	-

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 for feasible inference on the cointegrating vector, $\boldsymbol{\beta}$, 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, recall that φ measures the smoothness of the spectral density matrix of the latent signal, $\boldsymbol{w}_t^*$, in the vicinity of the origin, see Assumption 2', and $d_c = d + d_{\epsilon}$ is the collective fractional integration order of the system. Finally, 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 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.

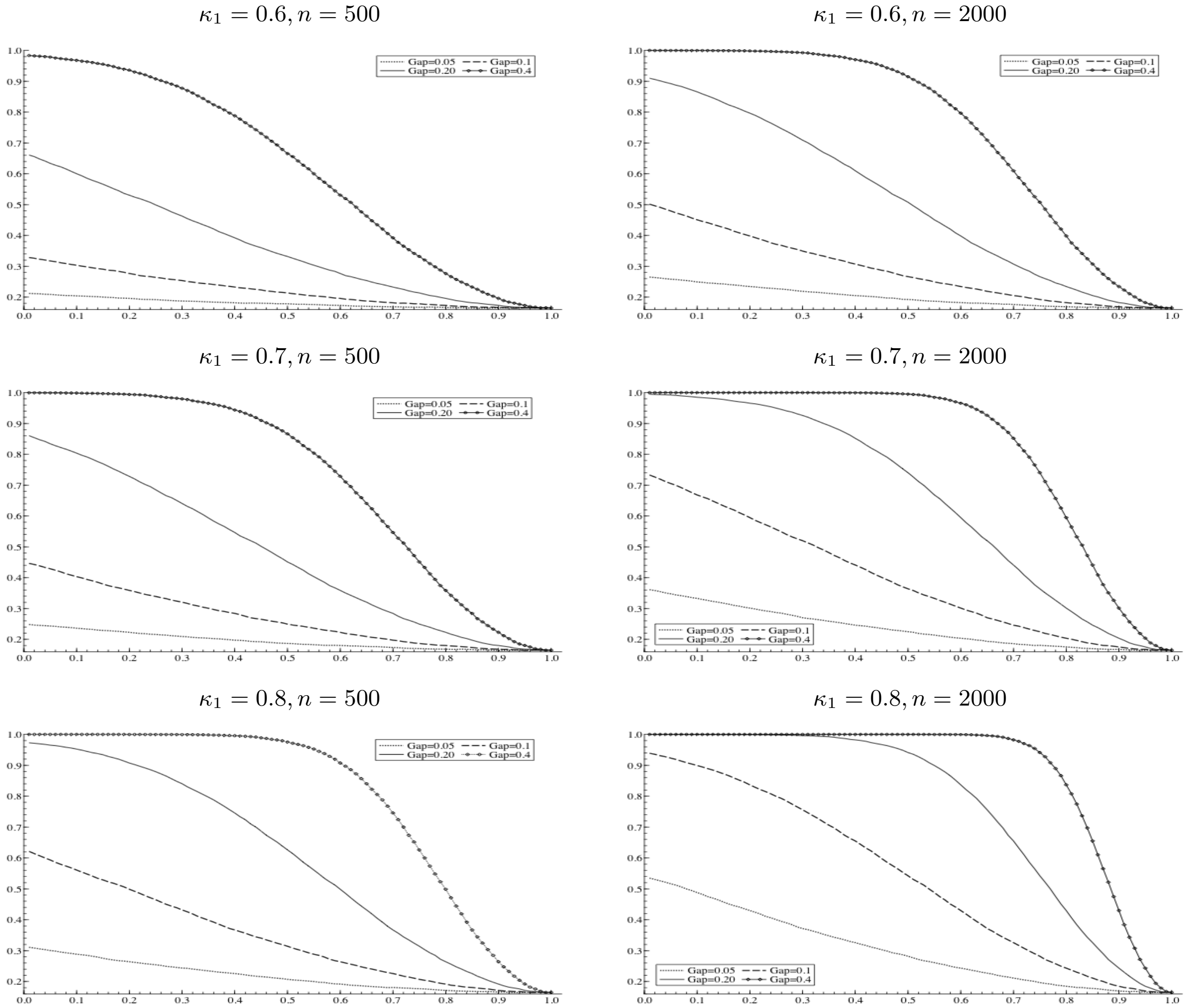

Figure 1: Size and power for the cointegration strength test when $b \leq d - d_\epsilon$. This figure illustrates the empirical rejection rates for the test statistic suggested in Corollary 2 when $b \leq d - d_\epsilon$ and the estimated memory gap is simulated as $\hat{d}_M - \hat{d}_{\epsilon,M} = d - d_\epsilon + m_1^{-1/2}/2\eta_1 - m_1^{-1/2}/2\eta_2$ for sample size $n \in \{500, 2000\}$, $m_1 = n^{\kappa_1}$, and $d - d_\epsilon \in \{0.05, 0.1, 0.2, 0.4\}$ where η_1 and η_2 are independent standard Gaussian random variables. In particular, the y-axis displays rejection rates, whereas the x-axis has $b/(d - d_\epsilon)$ for $b \in (0, d - d_\epsilon]$, i.e., the strength normalized by the memory gap. The test is implemented using a two-sided 5% significance level and 10000 replications.

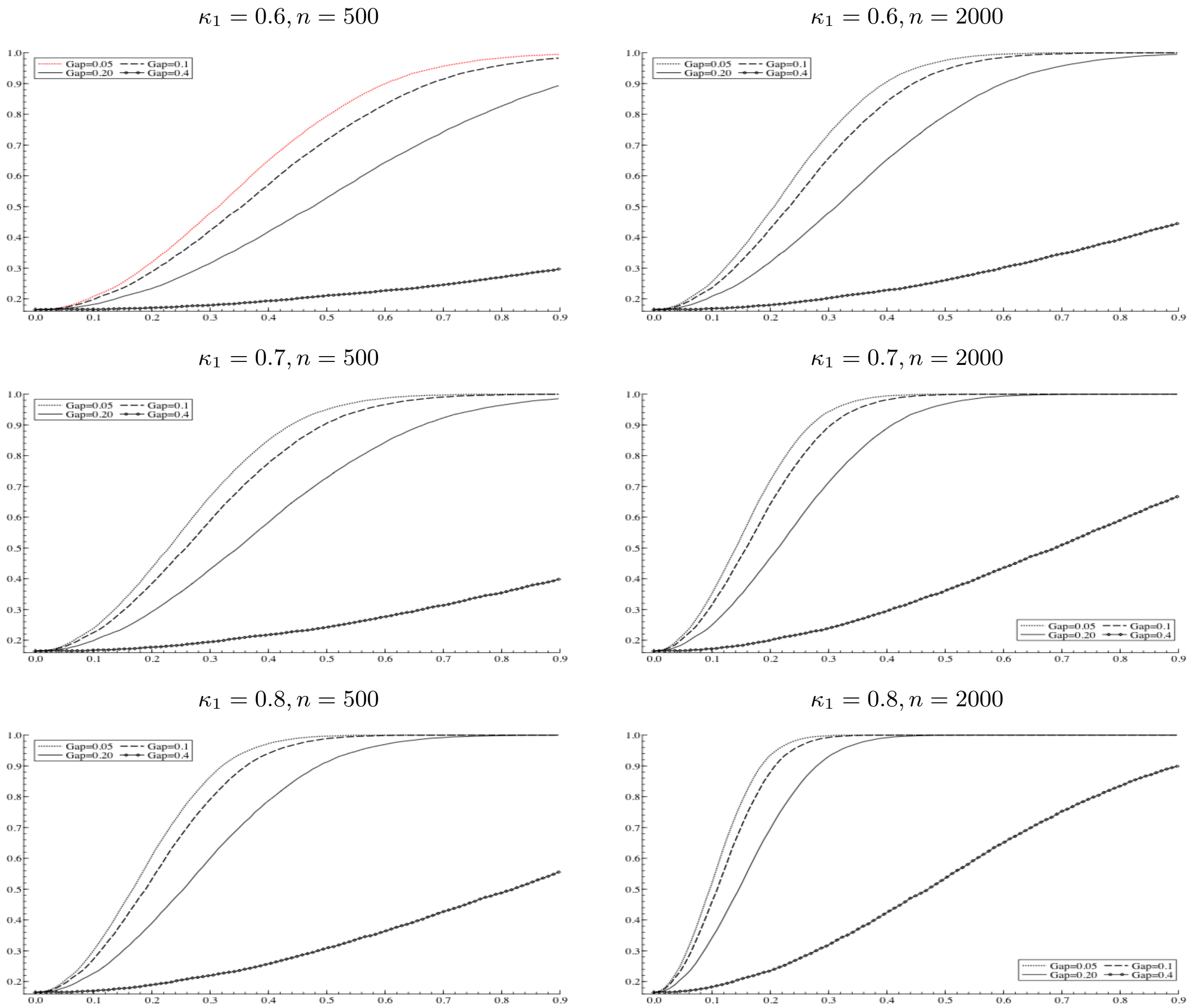

Figure 2: Size and power for the cointegration strength test when $b \geq d - d_\epsilon$. This figure illustrates empirical rejection rates for the test statistic suggested in Corollary 2 when $b \geq d - d_\epsilon$ and the estimated memory gap is simulated as $\hat{d}_M - \hat{d}_{\epsilon,M} = d - d_\epsilon + m_1^{-1/2}/2\eta_1 - m_1^{-1/2}/2\eta_2$ for sample size $n \in \{500, 2000\}$, $m_1 = n^{\kappa_1}$, and $d - d_\epsilon \in \{0.05, 0.1, 0.2, 0.4\}$ where η_1 and η_2 are independent standard Gaussian random variables. In particular, the y-axis displays rejection rates, whereas the x-axis has $(b - (d - d_\epsilon))/d_U$ for $b \in [d - d_\epsilon, d_U)$ where $d_U = 0.49$, i.e., the strength normalized by the memory gap. The test is implemented using a two-sided 5% significance level and 10000 replications.

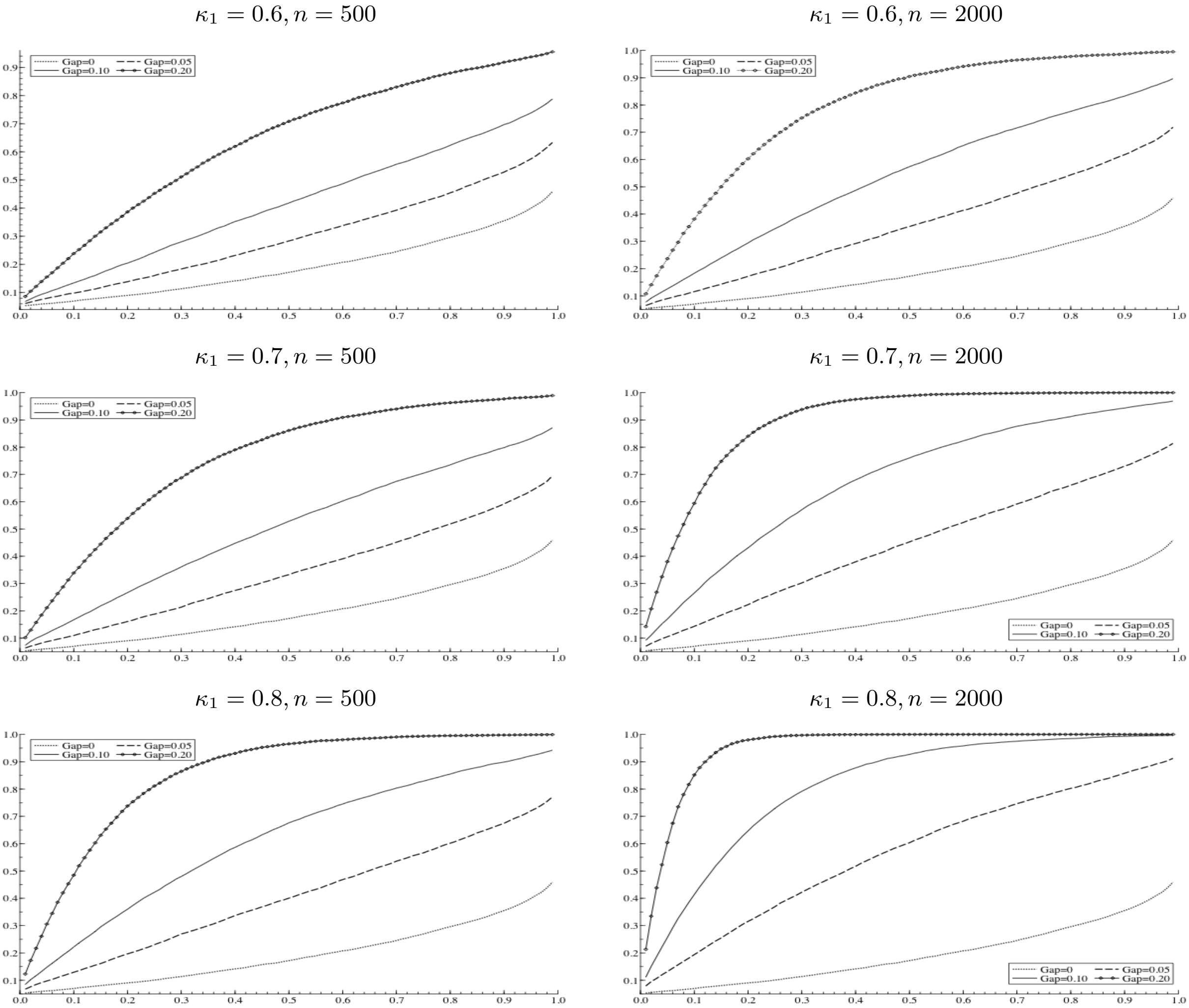

Figure 3: Size and power for consistency test of NBLS inference. This figure illustrates the empirical rejection rates for the Hausman-type test statistic suggested in Corollary 3 when the "estimated" memory gap is simulated as $\hat{d}_{\epsilon,N}(1,c) - \hat{d}_{\epsilon,M} = d_{\epsilon,N} - d_{\epsilon,M} + (cm_1)^{-1/2}/2\eta_1 - m_1^{-1/2}/2\eta_2$ for $n \in \{500, 2000\}$, $m_1 = n^{\kappa_1}$, and $d_{\epsilon,N} - d_{\epsilon,M} \in \{0, 0.05, 0.1, 0.2\}$ where η_1 and η_2 are independent standard Gaussian random variables. The y-axis displays rejection rates, whereas the x-axis has the tuning parameter $c \in (0, 1)$, i.e., the scale of the bandwidth to achieve a non-degenerate distribution. The test is implemented using a one-sided 5% significance level and 10000 replications.

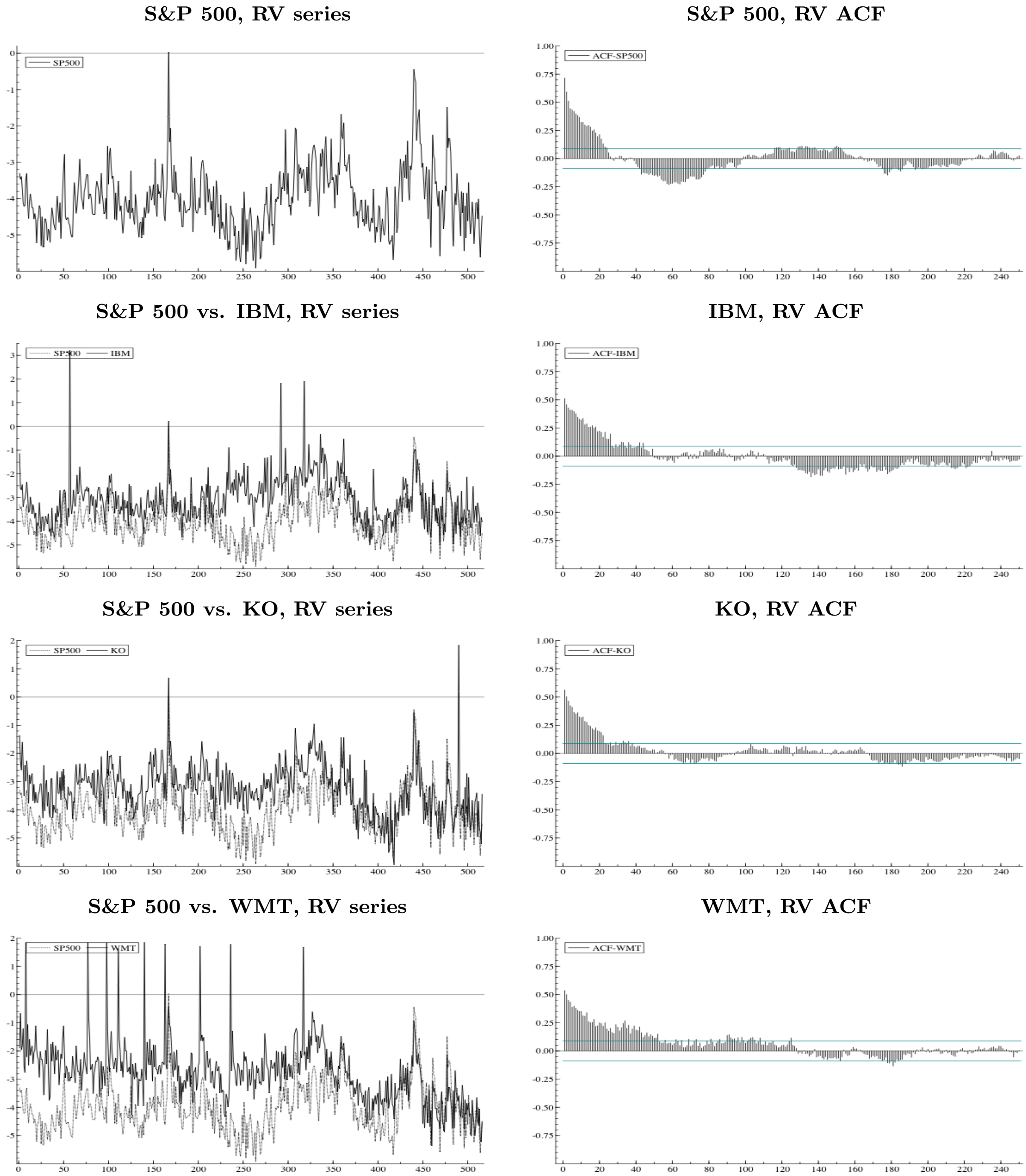

Figure 4: Realized variance series and ACF's. We depict the monthly log-realized variance series for S&P 500 alone and together with the corresponding series for the three equities IBM, KO, and WMT, using data from 1975 through August 2014, amounting to $n = 516$ observations. Furthermore, we illustrate the respective empirical ACF's 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 2: 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 the trimmed LW estimator, respectively, which are implemented with a bandwidth $m = n^{\kappa_1}$, for $\kappa_1 = 0.7$, and a trimming function $\ell = n^{\nu_1}$ for $\nu_1 = 1/2 + \zeta$, $\zeta = 0.05$.

Parameter Estimates for MBLS and NBLS: Monthly RV

	IBM, $\nu =$				**KO**, $\nu =$				**WMT**, $\nu =$			
	$1/3+\zeta$	$2/5+\zeta$	$1/2+\zeta$	NLBS	$1/3+\zeta$	$2/5+\zeta$	$1/2+\zeta$	NBLS	$1/3+\zeta$	$2/5+\zeta$	$1/2+\zeta$	NBLS
$\hat{\boldsymbol{\beta}}(\ell, m)$	0.6880 (0.0498)	0.6905 (0.0497)	0.6243 (0.0551)	0.5711 (−)	0.7640 (0.0277)	0.7028 (0.0294)	0.6885 (0.0301)	0.5970 (−)	0.7741 (0.1180)	0.7279 (0.1158)	0.5839 (0.1221)	0.4342 (−)
$\hat{d}_\epsilon(\ell_1, m_1)$	0.0452 (0.0662)	0.0445 (0.0662)	0.0649 (0.0662)	0.4377 (0.0563)	−0.0465 (0.0662)	−0.0250 (0.0662)	−0.0184 (0.0662)	0.4280 (0.0563)	0.1375 (0.0662)	0.1366 (0.0662)	0.1424 (0.0662)	0.4640 (0.0563)
$\hat{b}(\ell_1, m_1)$	0.3076	0.3083	0.2879	0.0271	0.4336	0.4121	0.4056	0.0422	0.1781	0.1790	0.1732	-0.0029
$\mathcal{CV}^{-}_{90}(t_b)$	0.1361	0.1368	0.1165	-	0.2621	0.2406	0.2341	-	0.0066	0.0075	0.0017	-
$\mathcal{CV}^{-}_{95}(t_b)$	0.1033	0.1039	0.0836	-	0.2293	0.2078	0.2012	-	-0.0263	-0.0254	-0.0312	-
$\mathcal{CV}^{-}_{99}(t_b)$	0.0391	0.0397	0.0194	-	0.1651	0.1436	0.1370	-	-0.0905	-0.0896	-0.0954	-
$t_d(0.5)$	9.159**	9.309**	8.995**	-	9.608**	9.264**	9.159**	-	7.464**	7.478**	7.385**	-
$t_d(0.7)$	12.53**	12.55**	12.05**	-	12.87**	12.34**	12.18**	-	9.254**	9.276**	9.134**	-
$\tilde{t}_d(0.7)$	9.594**	9.610**	9.113**	-	8.781**	8.256**	8.094**	-	7.293**	7.316**	7.173**	

Table 3: 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 $\nu \in \{1/3+\zeta, 2/5+\zeta, 1/2+\zeta\}$ where $\zeta = 0.05$. 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 2-4 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 $\nu_1 = \nu_2 = 1/2+\zeta$. 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, $\hat{d}_\epsilon(\ell_1, m_1)$, is implemented with $\kappa_1 = 0.7$ along with $\nu_1 = 1/2+\zeta$ for MBLS and $\nu_1 = 0$ for NBLS. The estimated cointegration strength is defined as $\hat{b}(\ell_1, m_1) = \hat{d}(\ell_1, m_1) - \hat{d}_\epsilon(\ell_1, m_1)$. The values $\mathcal{CV}^{-}_{1-2\alpha}(t_b)$ for $\alpha \in \{0.05, 0.025, 0.005\}$ are the upper critical values of b using a significance level $1-2\alpha$ implied by the test for cointegration strength in Corollary 2. Furthermore, $t_d(c)$, for $c \in \{0.5, 0.7\}$, denotes the Hausman-type t-test for consistency of NBLS inference in Corollary 3 based on the estimates $\hat{d}_\epsilon(\ell_1, m_1)$. The test $\tilde{t}_d(c)$, for $c = 0.7$, is similar, however with the fractional integration order of the residuals estimated using the trimmed LW with $\nu_1 = 1/2+\zeta$ for *both* MBLS and NBLS. See the main text for detail. (*) and (**) denote rejection at a 5% and 1% significance level, respectively.

Parameter Estimates for MBLS and NBLS: Weekly RV

	IBM				KO				WMT			
	$1/3+\zeta$	$2/5+\zeta$	$1/2+\zeta$	0	$1/3+\zeta$	$2/5+\zeta$	$1/2+\zeta$	0	$1/3+\zeta$	$2/5+\zeta$	$1/2+\zeta$	0
$\hat{\beta}(\ell, m)$	0.7039 (−)	0.6431 (−)	0.6142 (−)	0.5873 (−)	0.6663 (−)	0.6353 (−)	0.5564 (−)	0.5900 (−)	0.6447 (0.0363)	0.5705 (0.0380)	0.5477 (0.0390)	0.4528 (−)
$\hat{d}_\epsilon(\ell_1, m_1)$	0.1142 (0.0392)	0.1160 (0.0392)	0.1186 (0.0392)	0.3639 (0.0346)	0.2159 (0.0392)	0.2200 (0.0392)	0.2347 (0.0392)	0.3467 (0.0346)	0.0260 (0.0392)	0.0358 (0.0392)	0.0401 (0.0392)	0.3852 (0.0346)
$\hat{b}(\ell_1, m_1)$	0.2939	0.2921	0.2895	0.0923	0.2433	0.2392	0.2246	0.1228	0.3540	0.3442	0.3399	0.0749
$\mathcal{CV}_{90}^{-}(t_b)$	0.1740	0.1722	0.1696	-	0.1234	0.1192	0.1046	-	0.2340	0.2242	0.2198	-
$\mathcal{CV}_{95}^{-}(t_b)$	0.1510	0.1492	0.1466	-	0.1004	0.0962	0.0816	-	0.2110	0.2013	0.1969	-
$\mathcal{CV}_{99}^{-}(t_b)$	0.1061	0.1043	0.1017	-	0.0555	0.0513	0.0367	-	0.1661	0.1563	0.1520	-
$t_d(0.5)$	10.03**	9.980**	9.908	-	5.723**	5.607**	5.197**	-	13.64**	13.37**	13.24**	-
$t_d(0.7)$	12.36**	12.28**	12.17**	-	5.854**	5.677**	5.051**	-	19.10**	18.68**	18.49**	-
$\tilde{t}_d(0.7)$	8.271**	8.193**	8.083**	-	-0.1265	-0.3040	-0.9299	-	16.67**	16.25**	16.07**	-

Table 4: 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 $\nu \in \{1/3+\zeta, 2/5+\zeta, 1/2+\zeta\}$ where $\zeta = 0.05$. 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 2-4 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 $\nu_1 = \nu_2 = 1/2+\zeta$. 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, $\hat{d}_\epsilon(\ell_1, m_1)$, is implemented with $\kappa_1 = 0.7$ along with $\nu_1 = 1/2+\zeta$ for MBLS and $\nu_1 = 0$ for NBLS. The estimated cointegration strength is defined as $\hat{b}(\ell_1, m_1) = \hat{d}(\ell_1, m_1) - \hat{d}_\epsilon(\ell_1, m_1)$. The values $\mathcal{CV}_{1-2\alpha}^{-}(t_b)$ for $\alpha = \{0.05, 0.025, 0.005\}$ are the lower critical values of b using a significance level $1-2\alpha$ implied by the test for cointegration strength in Corollary 2. Furthermore, $t_d(c)$, for $c = \{0.5, 0.7\}$, denotes the Hausman-type t-test for consistency of NBLS inference in Corollary 3 based on the estimates $\hat{d}_\epsilon(\ell_1, m_1)$. The test $\tilde{t}_d(c)$, for $c = 0.7$, is similar, however with the fractional integration order of the residuals estimated using the trimmed LW with $\nu_1 = 1/2+\zeta$ for *both* MBLS and NBLS. See the main text for detail. (*) and (**) denote rejection at a 95% and 99% significance level, respectively.

References

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. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

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

Andersen, T. G., Fusari, N. & Todorov, V. (2014), 'Parametric inference and dynamic state recovery from option panels', *Econometrica* . forthcoming.

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.

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.

Chiriac, R. & Voev, V. (2011), 'Modelling and forecasting multivariate realized volatility', *Journal of Applied Econometrics* **28**, 922–947.

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.

Christoffersen, P., Heston, S. & Jacobs, K. (2009), 'The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well', *Management Science* **55**, 1914–1932.

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.

Duecker, 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.

Duffie, D., Pan, J. & Singleton, K. (2000), ‘Transform analysis and asset pricing for affine jump-diffusions’, *Econometrica* **68**, 1343–1376.

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.

Eraker, B., Johannes, M. & Polson, N. (2003), ‘The impact of jumps in volatility and returns’, *The Journal of Finance* **58**, 1269–1300.

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.

Hillebrand, E. T. (2005), ‘Neglecting parameter changes in GARCH models’, *Journal of Econometrics* **129**, 121–138.

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.

Koopman, S. J., Jungbacker, B. & Hol, E. (2005), 'Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements', *Journal of Empirical Finance* **12**, 445–475.

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. (2014), Parameter estimation robust to low-frequency contamination. Unpublished Manuscript, Department of Economics, Brown University.

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.

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.

Phillips, P. C. & Hansen, B. E. (1990), ‘Statistical inference in instrumental variables regression with I(1) variables’, *Review of Economic Studies* **57**, 99–125.

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. (1995*a*), ‘Gaussian semiparametric estimation of long range dependence’, *The Annals of Statistics* **23**, 1630–1661.

Robinson, P. M. (1995*b*), ‘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.

Todorov, V. (2011), 'Econometric analysis of jump-driven stochastic volatility models', *Journal of Econometrics* **160**, 12–21.

Todorov, V., Tauchen, G. & Grynkiv, I. (2011), 'Realized laplace transforms for estimation of jump diffusive volatility models', *Journal of Econometrics* **164**, 367–381.

Varneskov, R. T. (2014), 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 Proofs

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 $\boldsymbol{V}$ let $\boldsymbol{V}(i)$ index the i-th element, and, similarly, let $\boldsymbol{M}(i,q)$ denote the (i,q)-th element of a matrix $\boldsymbol{M}$. $\boldsymbol{F}^*_{zz}(\ell,m)$ and $\boldsymbol{I}^*_{zz}(\lambda_j)$ are the trimmed discretely averaged co-periodogram and the co-periodogram, respectively, of $\boldsymbol{z^*}$. Similarly, we use $\hat{\boldsymbol{F}}^*_{x\epsilon}(\ell,m)$ to denote the vector $\hat{\boldsymbol{F}}_{x^*\epsilon^*}(\ell,m)$. Moreover, $K \in (0,\infty)$ is a generic constant, which may take different values in different places. Finally, we remark that sometimes the (stochastic) orders refer to scalars, sometimes to matrices. We refrain from making distinctions.

A.1 Proof of Theorem 1

First, for **(a)**, to establish a bound for $|\hat{\boldsymbol{F}}_{zz}(\ell,m) - \boldsymbol{G}(\lambda_m)|$, use Assumption 1 to decompose

$$\begin{aligned}\hat{\boldsymbol{F}}_{zz}(\ell,m) - \hat{\boldsymbol{F}}^*_{zz}(1,m) &= -\frac{2\pi}{n}\left(\sum_{j=1}^{m}\Re(\boldsymbol{I}^*_{zz}(\lambda_j)) - \sum_{j=\ell}^{m}\Re(\boldsymbol{I}^*_{zz}(\lambda_j))\right) + \frac{2\pi}{n}\sum_{j=\ell}^{m} O_p\left(\frac{n}{j^2}\right) \\ &= -\hat{\boldsymbol{F}}^*_{zz}(1,\ell-1) + 2\pi\sum_{j=\ell}^{m} O_p\left(j^{-2}\right), \end{aligned} \tag{A.1}$$

and define $\mathcal{E}(\ell) = 2\pi\sum_{j=\ell}^{m} O_p(j^{-2})$. Then, since $2 > 1$, trivially, we may invoke Varneskov (2014, Lemma C.4) to show that $|\mathcal{E}(\ell)| \leq O_p(\ell^{-2})$. Next, we seek to connect the two quantities $\hat{\boldsymbol{F}}^*_{zz}(1,m)$ and $\boldsymbol{G}(\lambda_m)$ using the continuously averaged spectral density $\boldsymbol{F}^*_{zz}(\lambda_m)$. Similar 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 (A.1), that is, for

$$\boldsymbol{\Lambda}_m\boldsymbol{\Lambda}_m^{-1}\left(\hat{\boldsymbol{F}}^*_{zz}(1,m) - \boldsymbol{F}^*_{zz}(\lambda_m)\right)\bar{\boldsymbol{\Lambda}}_m^{-1}\bar{\boldsymbol{\Lambda}}_m - \boldsymbol{\Lambda}_\ell\boldsymbol{\Lambda}_\ell^{-1}\hat{\boldsymbol{F}}^*_{zz}(1,\ell-1)\bar{\boldsymbol{\Lambda}}_\ell^{-1}\bar{\boldsymbol{\Lambda}}_\ell \tag{A.2}$$

since Assumption 2 immediately provides

$$\begin{aligned}\boldsymbol{F}^*_{zz}(\lambda_m,i,k) &\sim G_{i,k}\lambda_m^{1-d_i-d_k}/(1-d_i-d_k)\Re(e^{\mathrm{i}\pi(d_i-d_k)/2}) \\ &\sim G_{i,j}\lambda_m^{1-d_i-d_k}/(1-d_i-d_k)\cos(\pi(d_i-d_k)/2)\end{aligned} \tag{A.3}$$

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

$$\boldsymbol{\Lambda}_m^{-1}(\hat{\boldsymbol{F}}^*_{zz}(1,m) - \boldsymbol{F}^*_{zz}(\lambda_m))\boldsymbol{\Lambda}_m^{-1} = o_p\left(\lambda_m(1+\lambda_m+\lambda_m^2)\right) = o_p(\lambda_m), \tag{A.4}$$

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

$$|\hat{\boldsymbol{F}}^*_{zz}(1,m) - \boldsymbol{G}(\lambda_m)| \le o_p((m/n)^{1-2d_\top}) \tag{A.5}$$

gives the upper bound for the stochastic order of the approximation error. By the same arguments, it follows for the *second* term in (A.2) that

$$\hat{\boldsymbol{F}}^*_{zz}(1,\ell-1) = \boldsymbol{G}(\lambda_\ell) + o_p((\ell/n)^{1-2d_\top}) \le O_p((\ell/n)^{1-2d_\top}), \tag{A.6}$$

uniformly, thus providing the final result **(a)**.

Second, for the consistency result in **(b)**, we know from **(a)** that $|\hat{\boldsymbol{F}}_{xx}(\ell,m) - \boldsymbol{G}_{xx}(\lambda_m)| = o_p(1)$ where $\boldsymbol{G}_{xx}(\lambda_m)$ is the lower $(p-1)\times(p-1)$ submatrix of $\boldsymbol{G}(\lambda_m)$. Next, decompose $\hat{\boldsymbol{F}}_{xy}(\ell,m)$ as

$$\begin{aligned}\hat{\boldsymbol{F}}_{xy}(\ell,m) &= \hat{\boldsymbol{F}}_{xy^*}(\ell,m) + \hat{\boldsymbol{F}}_{xu_1}(\ell,m) = \hat{\boldsymbol{F}}_{xx^*}(\ell,m)\boldsymbol{\beta} + \hat{\boldsymbol{F}}_{x\epsilon^*}(\ell,m) + \hat{\boldsymbol{F}}_{xu_1}(\ell,m)\\ &= \hat{\boldsymbol{F}}^*_{xx}(\ell,m)\boldsymbol{\beta} + \hat{\boldsymbol{F}}^*_{x\epsilon}(\ell,m) + \hat{\boldsymbol{F}}_{u_2u_1}(\ell,m) + o_p(1)\end{aligned} \tag{A.7}$$

where u_1 indexes the first element of the low-frequency contamination vector $\boldsymbol{u}$, and u_2 the vector with the remaining $p-1$ elements. The last equality follows immediately since $\boldsymbol{x}^*_t \perp\!\!\!\perp \boldsymbol{u}_s\ \forall t,s$ by Assumption 1. For the terms $\hat{\boldsymbol{F}}^*_{xx}(\ell,m)$ and $\hat{\boldsymbol{F}}_{u_2u_1}(\ell,m)$, we establish the following uniform bounds using the same arguments as in the proof of **(a)**,

$$|\hat{\boldsymbol{F}}^*_{xx}(\ell,m) - \boldsymbol{G}_{xx}(\lambda_m)| \le O_p((\ell/n)^{1-2d_\top}) + o_p((m/n)^{1-2d_\top}), \quad |\hat{\boldsymbol{F}}_{u_2u_1}(\ell,m)| \le O_p(\ell^{-2}). \tag{A.8}$$

Finally, as $|\hat{\boldsymbol{F}}^*_{x\epsilon}(\ell,m,i)| \le (\hat{\boldsymbol{F}}^*_{xx}(\ell,m,i,i)\hat{F}^*_{\epsilon\epsilon}(\ell,m))^{1/2}$ for all $i=1,\ldots,p-1$ elements of the cross-product vector $\hat{\boldsymbol{F}}^*_{x\epsilon}(\ell,m,i)$ by the Cauchy-Schwarz inequality and

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

using **(a)** and Assumption 2, the final consistency result follows using the same arguments provided by Robinson & Marinucci (2003, pp. 361-362) in conjunction with **(a)**. □

Remark 4. *We use the maximal inequality from Varneskov (2014, 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 semimartingales. However, close inspection shows that it applies directly in the present context as well.*

A.2 Proof of Theorem 2

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

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

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

As $\lambda_m^{2d-1}\boldsymbol{H}_{xx}(\lambda_m)=\boldsymbol{H}_{xx}$ by definition in (7), $O_p(\lambda_m^{2d-1}(\ell/n)^{1-2d})=O_p((\ell/m)^{1-2d})\xrightarrow{\mathbb{P}}\mathbf{0}$ as $\ell/m\to 0$ and $d\in(0,1/2)$, and, finally, $O_p(\lambda_m^{2d-1}\ell^{-2})=O_p((n/m)^{1-2d}\ell^{-2})\xrightarrow{\mathbb{P}}\mathbf{0}$ where the last convergence result is guaranteed by the (fourth) regularity condition $m^{1/2}(n/m)^{1-d_c}\ell^{-2}\to 0$ in Assumption 5. $\square$

First, use the decomposition in (A.7) to write

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

Then, for the term $(E.1)$, we may use (A.1) to show that $\hat{\boldsymbol{F}}_{xx}(\ell,m)=\hat{\boldsymbol{F}}_{xx}^{*}(\ell,m)+O_p(\ell^{-2})$. Hence, as $\ell\to\infty$, we have $\sqrt{m}\lambda_m^{d_\epsilon-d}\hat{\boldsymbol{F}}_{xx}^{-1}(\ell,m)\hat{\boldsymbol{F}}_{xx}^{*}(\ell,m)\boldsymbol{\beta}=\sqrt{m}\lambda_m^{d_\epsilon-d}\boldsymbol{\beta}(1+o_p(m^{-1/2}\lambda_m^{d-d_\epsilon}))$ by the continuous mapping theorem such that

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

Next, we may use Lemma 3, in conjunction with the continuous mapping theorem, and (A.8) to show,

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

where the final convergence result follows by the (fourth) regularity condition in Assumption 5. For the last term, $(E.3)$, we may write

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

Hence, it suffices to show $\sqrt{m}\lambda_m^{d_c-1}\boldsymbol{F}_{x\epsilon}^{*}(1,\ell-1)=o_p(1)$ since, in this case, we may invoke Christensen & Nielsen (2006, Theorem 2) for $\sqrt{m}\lambda_m^{d_c-1}\boldsymbol{F}_{x\epsilon}^{*}(1,m)$ and use this in conjunction with Lemma 3 and the continuous mapping 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, $\boldsymbol{\psi}$,

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

$$
\begin{aligned}
&+\sum_{i=1}^{p-1}\psi_i\sqrt{m}\lambda_m^{d_c-1}\frac{2\pi}{n}\sum_{j=1}^{\ell-1}\Re\left(\boldsymbol{A}^*(\lambda_j,i)\boldsymbol{J}(\lambda_j)\bar{\boldsymbol{A}}^*(\lambda_j,p)\right)\\
&\equiv(E.3.1)+(E.3.2)
\end{aligned}
$$

where, for $(E.3.1)$, we may use summation by parts and Lobato (1999, C.2) to show

$$
\begin{aligned}
(E.3.1)&\leq O_p\left(\sum_{i=1}^{p-1}\eta_i\sqrt{m}\lambda_m^{d_c-1}\lambda_\ell^{-d_c}\frac{1}{n}\left[\ell^{1/3}(\log\ell)^{2/3}+\log\ell+\ell^{1/2}/n^{1/4}\right]\right)\\
&=O_p\left(\left(\frac{m}{\ell}\right)^{d_c}\left[\frac{\ell^{1/3}}{m^{1/2}}(\log\ell)^{2/3}+\frac{\log\ell}{m^{1/2}}+\left(\frac{\ell}{m}\right)^{1/2}\frac{1}{n^{1/4}}\right]\right)\\
&=O_p\left(\left(\frac{\ell}{m}\right)^{1/2-d_c}\left[\frac{(\log\ell)^{2/3}}{\ell^{1/6}}+\frac{\log\ell}{\ell^{1/2}}+\frac{1}{n^{1/4}}\right]\right)\xrightarrow{\mathbb{P}}0
\end{aligned}
$$

since $d_c\in(0,1/2)$ and $\ell/m\to0$ as $n\to\infty$. Last, for $(E.3.2)$, since $\boldsymbol{v}_t$ is a martingale difference sequence with finite fourth moment, $\|\boldsymbol{J}(\lambda_j)\|\leq K<\infty$, $j=1,\ldots,\ell-1$,

$$
\begin{aligned}
(E.3.2)&\leq K\sum_{i=1}^{p-1}\psi_i\sqrt{m}\lambda_m^{d_c-1}\frac{2\pi}{n}\sum_{j=1}^{\ell-1}\Re\left(\boldsymbol{A}^*(\lambda_j,i)\bar{\boldsymbol{A}}^*(\lambda_j,p)\right)\\
&\leq K\sqrt{m}\lambda_m^{d_c-1}\frac{2\pi}{n}\sup_{i=1,\ldots,p-1}\sum_{j=1}^{\ell-1}|\boldsymbol{f}^*_{ww}(\lambda_j,i,p)|=O\left(\sqrt{m}\lambda_m^{d_c-1}\frac{\ell}{n}\lambda_\ell^{\varphi-d_c}\right)=O\left(\frac{\ell^{1+d_c+\varphi}}{m^{1/2+d_c}n^{\varphi}}\right),
\end{aligned}
$$

which is $o_p(1)$ using Assumption 2' for the first equality, since $G_{i,p}=0\ \forall i=1,\ldots,p-1$, and the fifth condition of Assumption 5 for the final convergence result, concluding the proof. □

A.3 Proof of Theorem 3

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

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

where $\tilde{u}_t=u_{1,t}-\boldsymbol{\beta}(\ell,m)'u_{2,t}$. Without loss of generality we have set $\boldsymbol{c}=\boldsymbol{0}$ in (1), since a constant vector is a special case of the low-frequency contamination in Assumption 1. Since $\boldsymbol{x}_t^*\perp\!\!\!\perp\boldsymbol{u}_s\ \forall t,s$, we may write

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

where $I^*_{\epsilon\epsilon}(\lambda_j)+I_{\tilde{u}\tilde{u}}(\lambda_j)$ correspond to the periodogram analyzed in the objective function by Iacone (2010) and McCloskey & Perron (2013) since $\tilde{u}_t=(1,-\boldsymbol{\beta})\boldsymbol{u}_t(1+o_p(1))$ is simply a linear combination of low-frequency contamination and $d_\epsilon \in [0, d_c)$. The last two terms are the additional errors from having estimated $\boldsymbol{\beta}$ in the first stage. Define $c_1(\lambda_j) = (\boldsymbol{\beta} - \boldsymbol{\beta}(\ell, m))'\Re(\boldsymbol{I}^*_{xx}(\lambda_j))(\boldsymbol{\beta} - \boldsymbol{\beta}(\ell, m))$ and $c_2(\lambda_j) = 2(\boldsymbol{\beta} - \boldsymbol{\beta}(\ell, m))'\Re(\boldsymbol{I}^*_{x\epsilon}(\lambda_j))$, then 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), it suffices to show

$$(E.4) = \frac{G^{-1}_{p,p}}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \lambda_j^{2\delta+2d_\epsilon} c_1(\lambda_j) \leq O_p\Big(\frac{G^{-1}_{p,p}}{m_1 - \ell_1 + 1} \sum_{j=1}^{m_1} \lambda_j^{2\delta+2d_\epsilon} c_1(\lambda_j)\Big), \tag{A.9}$$

$$(E.5) = \frac{G^{-1}_{p,p}}{m_1 - \ell_1 + 1} \sum_{j=\ell_1}^{m_1} \lambda_j^{2\delta+2d_\epsilon} c_2(\lambda_j) \leq O_p\Big(\frac{G^{-1}_{p,p}}{m_1 - \ell_1 + 1} \sum_{j=1}^{m_1} \lambda_j^{2\delta+2d_\epsilon} c_2(\lambda_j)\Big), \tag{A.10}$$

for $\delta = 0$, that is, for the highest asymptotic order of the bias, to invoke their theorem and eliminate the additional errors from the first-stage estimation of $\boldsymbol{\beta}$. The scaling with $G^{-1}_{p,p}\lambda_j^{2d_\epsilon}$ follows by Nielsen & Frederiksen (2011, (A.15)). First, for condition (A.9), the bound is immediate as $c_1(\lambda_j) \geq 0$ $\forall j = 1, \ldots, \lfloor n/2 \rfloor$. Next, for condition (A.10), we may write,

$$(E.5) = \frac{2G^{-1}_{p,p}}{m_1 - \ell_1 + 1}(\boldsymbol{\beta} - \boldsymbol{\beta}(\ell, m))'\Big(\sum_{j=1}^{m_1} \lambda_j^{2d_\epsilon}\Re(\boldsymbol{I}^*_{x\epsilon}(\lambda_j)) - \sum_{j=1}^{\ell_1-1} \lambda_j^{2d_\epsilon}\Re(\boldsymbol{I}^*_{x\epsilon}(\lambda_j))\Big).$$

Hence, as in the proof of Theorem 2, we may use the Cramer-Wold Theorem for an arbitrary $(p-1)\times 1$ vector, $\boldsymbol{\psi}$, on this decomposition such that verifying condition (A.10) amounts to showing

$$O\left(\Big|\sum_{i=1}^{p-1} \psi_i \sum_{j=1}^{\ell_1-1} \lambda_j^{2d_\epsilon}\Re(\boldsymbol{I}^*_{x\epsilon}(\lambda_j, i))\Big|\right) \leq O\left(\Big|\sum_{i=1}^{p-1} \psi_i \sum_{j=1}^{m_1} \lambda_j^{2d_\epsilon}\Re(\boldsymbol{I}^*_{x\epsilon}(\lambda_j, i))\Big|\right). \tag{A.11}$$

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 $G_{i,p} = G_{p,i} = 0$, $i = 1, \ldots, p-1$, provides the bound,

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

This result, since $s_1^{1+\varphi-d+d_\epsilon}$ and $s_1^{1/2-d+d_\epsilon}$ are 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 $\ell_1/m_1 \propto n^{\nu_1-\kappa_1} \to 0$ when $n \to \infty$, thus verifying condition (A.11) and thereby (A.10). Hence, we invoke Nielsen & Frederiksen (2011, Theorem 2.2) to asymptotically eliminate the right-hand-side errors in conditions (A.9) and (A.10) and use Lemma 2 to complete the proof. □

A.4 Proof of Theorem 4

Let $I_{\epsilon\epsilon}(\lambda_j)$ be the periodogram of $\epsilon_t = y_t - \boldsymbol{\beta}'\boldsymbol{x}_t$ at frequency ordinate λ_j. Note that this is not to be confused with $I^*_{\epsilon\epsilon}(\lambda_j)$, the periodogram of $\epsilon^*_t = y^*_t - \boldsymbol{\beta}'\boldsymbol{x}^*_t$, but merely used to establish the desired consistency result for $\hat{e}_t = y_t - \boldsymbol{\beta}(\ell, m)'\boldsymbol{x}_t$ in Theorem 4 (c).

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

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

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

Proof. Since $\boldsymbol{z}^*_t$ satisfies the conditions of Robinson (1995*b*, Theorem 2(a)) and since $(1, -\boldsymbol{\beta})\boldsymbol{u}_t$ form a linear combination of low-frequency contamination for ϵ_t, we may invoke the former to provide the first two asymptotic orders for the latent fractional signal, and use it in conjunction with McCloskey & Perron (2013, Theorem 1 (i)) and the Cauchy-Schwarz inequality to provide the asymptotic order for low-frequency contamination, concluding the proof. □

For **(a)**, the decomposition in Lemma 4 (a) provides us with three errors of stochastic orders

$$(E.6.1) = O_p\Big(m_2^{-1}\sum_{j=\ell_2}^{m_2}(\log j)j^{-1}\Big),\ (E.6.2) = O_p\Big(m_2^{-1}\sum_{j=\ell_2}^{m_2}(j/n)^{\varphi}\Big),\ (E.6.3) = O_p\Big(m_2^{-1}\sum_{j=\ell_2}^{m_2}\frac{n^{1-2d}}{j^{2-2d}}\Big),$$

respectively. First, as in the proof of Theorem 1 (a), the bound $(E.6.3) \le O_p(m_2^{-1}n^{1-2d}\ell_2^{2d-2})$ follows using Varneskov (2014, Lemma C.4). Next, to bound the two remaining terms, we may use Nielsen & Frederiksen (2011, (A.17)), which states that

$$\sup_{-1\le\tau\le K}\Big|m_2^{-\tau-1}(\log m_2)^{-1}\sum_{j=1}^{m_2}j^{\tau}\Big| = O(1), \quad \text{for } K \in (1, \infty). \tag{A.12}$$

This immediately gives the bound

$$(E.6.1) \le O_p(m_2^{-1}(\log m_2)\sum_{j=1}^{m_2}j^{-1}) \le O_p\left((\log m_2)^2 m_2^{-1}\right)$$

and, similarly, $(E.6.2) \le O_p(m_2^{-1}n^{-\varphi}\sum_{j=1}^{m_2}j^{\varphi}) \le O_p\left((m_2/n)^{\varphi}(\log m_2)\right)$. Lastly, the final convergence result follows by imposing the rates stated in Assumption 5-G (a). **(b)** follows by the same arguments. For **(c)**, we know from the proof of Theorem 3 above and the proof of Nielsen & Frederiksen (2011, Theorem 2.2) that the condition $(\log n)^4(\log m_2)(m/m_2)^{d-d_\epsilon} \to 0$ as $n \to \infty$ suffices to eliminate the additional errors from having estimated $\boldsymbol{\beta}$ in the first stage when establishing consistency using the scaled periodogram $G_{p,p}^{-1}\lambda_j^{2d_\epsilon}I_{\hat{e}\hat{e}}(\lambda_j)$. Hence, the result follows by **(b)**. □

Research Papers 2013

CREATES
Center for Research in Econometric Analysis of Time Series

2015-08: Alfonso Irarrazabal and Juan Carlos Parra-Alvarez: Time-varying disaster risk models: An empirical assessment of the Rietz-Barro hypothesis

2015-09: Daniela Osterrieder, Daniel Ventosa-Santaulària and Eduardo Vera-Valdés: Unbalanced Regressions and the Predictive Equation

2015-10: Laurent Callot, Mehmet Caner, Anders Bredahl Kock and Juan Andres Riquelme: Sharp Threshold Detection Based on Sup-norm Error rates in High-dimensional Models

2015-11: Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen and Anders Rahbek: Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)

2015-12: Tommaso Proietti, Martyna Marczak and Gianluigi Mazzi: EuroMInd-D: A Density Estimate of Monthly Gross Domestic Product for the Euro Area

2015-13: Michel van der Wel, Sait R. Ozturk and Dick van Dijk: Dynamic Factor Models for the Volatility Surface

2015-14: Tim Bollerslev, Andrew J. Patton and Rogier Quaedvlieg: Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting

2015-15: Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Effects of Macroeconomic Uncertainty upon the Stock and Bond Markets

2015-16: Markku Lanne, Mika Meitz and Pentti Saikkonen: Identification and estimation of non-Gaussian structural vector autoregressions

2015-17: Nicholas M. Kiefer and C. Erik Larson: Counting Processes for Retail Default Modeling

2015-18: Peter Reinhard Hansen: A Martingale Decomposition of Discrete Markov Chains

2015-19: Peter Reinhard Hansen, Guillaume Horel, Asger Lunde and Ilya Archakov: A Markov Chain Estimator of Multivariate Volatility from High Frequency Data

2015-20: Henri Nyberg and Harri Pönkä: International Sign Predictability of Stock Returns: The Role of the United States

2015-21: Ulrich Hounyo and Bezirgen Veliyev: Validity of Edgeworth expansions for realized volatility estimators

2015-22: Girum D. Abate and Niels Haldrup: Space-time modeling of electricity spot prices

2015-23: Eric Hillebrand, Søren Johansen and Torben Schmith: Data revisions and the statistical relation of global mean sea-level and temperature

2015-24: Tommaso Proietti and Alessandra Luati: Generalised partial autocorrelations and the mutual information between past and future

2015-25: Bent Jesper Christensen and Rasmus T. Varneskov: Medium Band Least Squares Estimation of Fractional Cointegration in the Presence of Low-Frequency Contamination