Robust Multivariate Local Whittle Estimation and Spurious Fractional Cointegration*

Janis Becker, Christian Leschinski, Philipp Sibbertsen[1]

Institute of Statistics, Faculty of Economics and Management,
Leibniz University Hannover, Germany

Abstract

This paper derives a multivariate local Whittle estimator for the memory parameter of a possibly long memory process and the fractional cointegration vector robust to low frequency contaminations. This estimator as many other local Whittle based procedures requires a priori knowledge of the cointegration rank. Since low frequency contaminations bias inference on the cointegration rank, we also provide a robust estimator of the cointegration rank. As both estimators are based on the trimmed periodogram we further derive some insights in the behaviour of the periodogram of a process under very general types of low frequency contaminations. An extensive Monte Carlo exercise shows the applicability of our estimators in finite samples.
Our procedures are applied to realized betas of two American energy companies discovering that the series are fractionally cointegrated. As the series exhibit low frequency contaminations, standard procedures are unable to detect this relation.

JEL-Numbers: C13, C32

Keywords: Multivariate Long Memory · Fractional Cointegration · Random Level Shifts · Semiparametric Estimation

*Financial support of the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. The authors are grateful to Fabricio Iacone, Remigijus Leipus, Guillaume Chevillon, the participants of the econometrics seminar at the University of Frankfurt(Oder), the CFE 2018 in Pisa, and the SNDE 2019 in Dallas for helpful comments.

[1]Corresponding author:
Phone: +49-511-762-3783
E-Mail: sibbertsen@statistik.uni-hannover.de

1 Introduction

It has been a well established fact that level shifts among many other so-called low-frequency contaminations can be mistaken as long memory. Künsch (1986), Granger and Ding (1996), Diebold and Inoue (2001) and Granger and Hyung (2004), among others, show that various forms of low frequency contaminations such as deterministic breaks and trends can cause spurious long memory. This leads to a bias of semiparametric estimators for the memory parameter which mainly use these frequencies. This feature has been used by Qu (2011) to test against spurious long memory.
However, the notion of spurious long memory is not restricted to the univariate case but can as well be found in multivariate systems. An extension of the test by Qu (2011) to the multivariate case relying on the same idea can be found in Sibbertsen et al. (2018). This paper shows that working in a multivariate system can result in efficiency gains and is therefore preferable where suitable. Multivariate local Whittle estimation of the memory parameter has been considered in Shimotsu (2007). Robinson (2008b) extends this to the case of possible fractional cointegration and simultaneously estimates the cointegration vector. Two series are called fractionally cointegrated if they have the same memory parameter and their linear combination has a reduced order of integration (see among many others Nielsen (2007)). Neither of these two estimators is robust against low frequency contaminations.
The aim of this paper is to provide such a robust multivariate local Whittle estimator of the memory parameter and the fractional cointegration vector that remains consistent in case of low frequency contaminations. Similar to the estimator of Robinson (2008b) our proposed estimator requires a priori knowledge of the cointegration rank. This is because local Whittle based methods need the inverse of the so-called $\boldsymbol{G}$ matrix of the spectral density which becomes singular in the case of fractional cointegration. Christensen and Santucci de Magistris (2010) and Kellard et al. (2015) discuss that in case of low frequency contaminations inference on the fractional cointegration rank is likely to be biased. For example, simultaneous breaks in the series can cause tests and estimators to falsely indicate the series to be fractionally cointegrated.
We therefore additionally suggest a robust estimator of the cointegration rank. For this purpose we investigate what we call spurious fractional cointegration further by generalizing the definition of cobreaking in Hendry and Massmann (2007) to what we call common low frequency contaminations. We show that low frequency contaminations dominate the $\boldsymbol{G}$ matrix of the periodogram for frequencies close to the origin and therefore empirically effect among others local Whittle based procedures. Due to this dominance of low frequency contaminations in the observed $\boldsymbol{G}$ matrix we find that common low frequency contaminations spuriously indicate the presence of fractional cointegration whereas distinct low frequency contaminations falsely indicate the absence of fractional cointegration.
In order to obtain our estimators we use the idea of Iacone (2010) of trimming away the contaminated frequencies. This idea is applied to provide a consistent estimator of the

cointegrating rank of the system by proposing a trimmed version of the procedure by Robinson and Yajima (2002) as well as to construct a robust multivariate local Whittle estimator for the memory parameter and the cointegrating vector. All of our approaches are spectral based and therefore semiparametric. To keep notation sparse we concentrate on the bivariate case although an extension to higher dimensions is straightforward. As our estimators rely on properties of the periodogram of processes with low frequency contaminations we find it useful to additionally provide some deeper understanding of the behaviour of the periodogram in this situation.

The paper is structured as follows. First, we provide some results for the periodogram of a contaminated process in a rather general framework of low frequency contaminations generalizing previously obtained results in Section 2. Section 3 formally defines common low frequency contaminations and contains our robust procedure to estimate the cointegration rank while Section 4 has the robust multivariate local Whittle estimator. Section 5 contains some Monte Carlo and Section 6 an empirical example. Section 7 concludes. All the proofs are gathered in the appendix.

2 The Periodogram of Spurious Long Memory Processes

In this section we obtain some properties of the periodogram for a very general class of low frequency contaminations which are partwise needed later but are also of an interest on its own. We therefore discuss it in more detail then necessary for our robust estimators and see this as an additional contribution of the paper. Although the focus of this paper is on multivariate estimation, we derive the results in this section in a univariate setup to avoid notational complexity. We will later assume the trend process to be independent from the noise process which means that an extension of the results to a multivariate framework is straightforward. It further implies that effects of additional noise components are irrelevant for the mean process so that we are only concerned with the behaviour of the pseudo-periodogram of a time-varying mean process in this section.

For the mean process we use a very general specification allowing for deterministic mean shifts, smooth deterministic trends and random level shifts with rare shift asymptotics as well as random level shifts with medium rare shifts where the number of shifts tend to infinity with sample size but with a slower rate. This model embeds many of the processes discussed in the literature to generate spurious long memory such as the fractional trend of Bhattacharya et al. (1983) or the STOPBREAK model of Engle and Smith (1999) among many others.

The mean process that can be either deterministic or stochastic is represented by

$$\mu_t = \mu_0 + \sum_{k=1}^{K} \Delta\mu_k 1(t \geq T_k), \quad (1)$$

or

$$\mu_t = \mu_0 + \sum_{k=0}^{K} \mu_k 1(T_{k-1} \le t < T_k). \tag{2}$$

Here, K is either a fixed number or a random variable giving the number of breaks, T is the length of the series, 1 corresponds to the indicator function, T_k denotes the breakpoint, and $\mu_0 = \mu_1 - 1/T \sum_{t=1}^{T} \mu_t$.

The expression in Equation (1) is a suitable representation of the mean for processes which have a nonstationary nature. Examples include deterministic trends or mean shifts as well as nonstationary random level shift models or the STOPBREAK model. The model is accumulative in the sense that the break at time t depends on all shifts that occurred before t. It also nests a random walk. The model in (2) on the other hand has a stationary character and seems appropriate for models such as the Markov-Switching model or stationary random level shift models. It has a non-cumulative structure and nests the White Noise.

In the following we denote by $I_z(\lambda_j) = w_z(\lambda_j) w_z^*(\lambda_j)$ the periodogram of the series z_t at frequency λ_j. Here, $w_z(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} z_t e^{i\lambda_j t}$ is the Fourier transform of the series z_t and the asterix denotes complex conjugation. $I_\mu(\lambda_j)$ is then the pseudo-periodogram of the mean process. We focus on the behaviour of this pseudo-periodogram at the Fourier frequencies $\lambda_j = \frac{2\pi j}{T}$ for $\lambda_j \to 0_+$.

We now derive the properties of the induced periodogram of the process (1) and (2). Let us first consider the case of a smooth trend $h(s,T)$ and assume:

Assumption A1. $|h(s,T)|$, $\left|\frac{\partial h(s,T)}{\partial s}\right| < \infty$, *for* $s \in [0,1]$.

We have the following Lemma:

Lemma 1. *If* $\mu_t = h(t/T,T)$, *under Assumption A1 we have*

$$I_\mu(\lambda_j) \sim \frac{T}{8\pi^3 j^2} \left\{ \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi js) ds \right]^2 + \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} (1 - \cos(2\pi js)) ds \right]^2 \right\}.$$

Since the integrals in Lemma 1 are functions of j (and possibly T), it can be seen that the exact rate of the periodogram $I_\mu(\lambda_j)$ depends on the derivative of the trend function. Therefore, if the trend function is known and the integrals have a closed form solution, it is possible to determine the exact order. If this is not the case, we can still recover the upper bound on the rate of decay for increasing j that was established by Künsch (1986), Qu (2011), and Iacone (2010). To see this, note that $\sin(2\pi js) \le 1$ and $1 - cos(2\pi js) \le 2$ for all j and s. It therefore follows immediately for $\mu_t = h(s,T) = h(s)$ that the periodogram is $I_\mu(\lambda_j) = O(Tj^{-2})$.

We now turn to the behavior of the periodogram of abrupt level-shift processes. To simplify the exposition, let ζ_k denote either $\Delta\mu_k$ or μ_k, depending on whether the accumulative structural-change model (1) or the non-accumulative model (2) is considered.

To characterize the behavior of different groups of processes, we require different groups of assumptions. First, in the case of deterministic structural breaks, we assume:

Assumption A2. *$|\zeta_k| < \infty$ and the $\delta_k = T_k/T$ are deterministic with $0 < \delta_k < 1$, for $k = 1, ..., K < \infty$.*

For stochastic level shifts we require the following assumptions.

Assumption A3. *$E[\zeta_k] = 0$ and $Var[\zeta_k] = \sigma_\Delta^2 T^{-\beta}$, for some $0 \leq \beta \leq 1$ and $0 < \sigma_\Delta^2 < \infty$.*

Assumption A4. *$P(t \in \{T_1, ..., T_K\}) = p_t$, where $0 \leq p_t \leq 1$, and $E[p_t] = \tilde{p}T^{-\alpha}$, for some $0 \leq \alpha \leq 1$. Furthermore, the dependence in p_t is limited such that $E[K] = \tilde{p}T^{1-\alpha}$, $E\left[((T_k - T_{k-1})/T)^2\right] = \frac{2\tilde{D}}{\tilde{p}^2}T^{2(\alpha-1)}$, and $E\left[((T_k - T_{k-1})/T)^4\right] = O(T^{4(\alpha-1)})$, for some $0 < \tilde{p}, \tilde{D} < \infty$.*

Assumption A5. *p_t is independent of ζ_k for all $k = 1, ..., K$ and $t = 1, ..., T$. Additionally, $\sum_{\tau=1}^{\infty} |E[\zeta_k \zeta_{k-\tau}]| = Var[\zeta_k]\tilde{C}$, for $k = 1, 2, ...,$ and $0 \leq \tilde{C} < \infty$.*

The rate $T^{-\beta}$ in Assumption A3 is required to nest a number of mean-change processes from the literature, such as the STOPBREAK process of Engle and Smith (1999). For other processes setting $\beta = 0$ gives the familiar setup with non-degenerate breaks.
Assumption A4 imposes a structure on the nature of the mean change process. The nature of the dependence in p_t is restricted by the additional requirement that the expected squared length of the kth regime expressed as a fraction of the sample is $\frac{2\tilde{D}}{\tilde{p}^2}T^{2(\alpha-1)}$, which means that the second moment of the regime lengths is still of the same order as that of a geometric distribution. In this context, the constant $\tilde{D}$ depends on the dependence in p_t, and it is equal to one, if $p_t = p$ for all $t = 1, ..., T$.
Since there are T observations in the sample, the expected number of mean shifts in the series is $E[K] = \tilde{p}T^{1-\alpha}$. The parameter α controls the asymptotic frequency of level changes. The expected number of shifts remains constant for $\alpha = 1$, whereas it goes to infinity for $\alpha < 1$. The first case ($\alpha = 1$) is referred to as *rare shifts* asymptotics or *low-frequency contaminations*. We refer to the second case ($\alpha < 1$) as *intermediate-frequency contaminations*. Here, we have $K \to \infty$ but $K/T \to 0$, as $T \to \infty$. That means we asymptotically have an infinite number of shifts, but also an infinite number of observations between shifts. Finally, for $\alpha > 1$ shifts are so rare that we will no longer observe any in a sample, asymptotically.
Even though it may seem unusual to tie the properties of the process to the sample size, this is a common approach in the related literature. Guégan (2005) refers to this practice as a thought experiment. The validity of this approach depends on the purpose of the analysis. Obviously, it is unreasonable to assume that structural changes will become less common in the future if the objective is to forecast a time series. On the other hand, if the objective is statistical inference based on a given sample, we argue that assuming that the frequency of structural change is tied to the sample size T can be thought of as an asymptotic framework that is better suited to approximate the statistical properties of the quantities of interest than keeping p fixed. The latter would imply, for example,

that level changes are so frequent that the mean between two shifts cannot be estimated consistently.

Finally, we require some bound on the degree of dependence between the means or mean changes ζ_k in consecutive segments. This is imposed by Assumption A5 according to which the autocovariance function of the ζ_k has to be absolutely summable. We then obtain the following result.

Lemma 2. *Denote by $\kappa > 0$ a finite constant and by $|\kappa_T| \leq 1$ a sequence of constants. Then, for $j/T \to 0$ and level-shift processes characterized by (1),*

i.) $I_\mu(\lambda_j) \sim \frac{T}{4\pi^3 j^2}\kappa$, *under Assumption A2.*

ii.) $E\left[I_\mu(\lambda_j)\right] \sim \frac{\sigma_\Delta^2 \tilde{p} T^{2-\alpha-\beta}}{4\pi^3 j^2}\left(1+\kappa_T \tilde{C}\right)$, *for $\alpha \leq 1$, and under Assumptions A3, A4, and A5.*

Lemma 2 establishes the properties of the periodogram of the accumulative mean-change process in (1). The first case i.) derives the growth rate of the peak near the origin and the rate of decay for frequencies further away from the zero frequency for a deterministic structural break process. This order was previously established by McCloskey and Perron (2013). Interesting is the contrast to case ii.), where rare random level shifts are considered. In contrast to i.), the periodogram becomes stochastic instead of deterministic. Furthermore, the scaling factor $T^{-\beta}$ influences the scaling of the peak local to the origin, which is of order $T^{1-\beta}$ instead of T. In i.) the periodogram is a deterministic function. In ii.) there is a well defined expectation and the process is not ergodic for $\alpha = 1$, the expected number of shifts in the sample is always given by $E[K] = \tilde{p}$. The case of $\alpha < 1$, on the other hand, covers intermediate frequency contaminations, so that the expected number of shifts is $E[K] = \tilde{p}T^{1-\alpha}$ and the process is ergodic. In this situation, the scaling of the peak near the origin is determined by both – α and β. Since $\alpha < 1$, the growth rate is always faster than that in case i.) and for $\alpha = 1$. The rate of decay for increasing j, however, is the same for all three types of processes.

Similar results to these can be obtained for the non-accumulative mean-change process in (2).

Lemma 3. *Denote by $\kappa' > 0$, $|\kappa'_T| \leq 1$, and $\kappa'_{P,T}$, a finite constant, a sequence of constants, and a sequence of positive valued random variables with constant expectation and finite variance, respectively. Then, for $j/T \to 0$ and level-shift processes characterized by (2),*

i.) $I_\mu(\lambda_j) \sim \frac{T}{2\pi^3 j^2}\kappa'$, *under Assumption A2.*

ii.) $I_\mu(\lambda_j) \sim \frac{\sigma_\Delta^2 \tilde{p} T^{1-\beta}}{2\pi^3 j^2}\kappa'_{P,T}$, *for $\alpha = 1$, and under Assumptions A3 and A4.*

iii.) $E\left[I_\mu(\lambda_j)\right] \sim \frac{\sigma_\Delta^2 \tilde{D}}{\pi \tilde{p}} T^{\alpha-\beta}\left(1+\kappa'_T \tilde{C}\right)$, *for $\alpha < 1$, and under Assumptions A3, A4, and A5.*

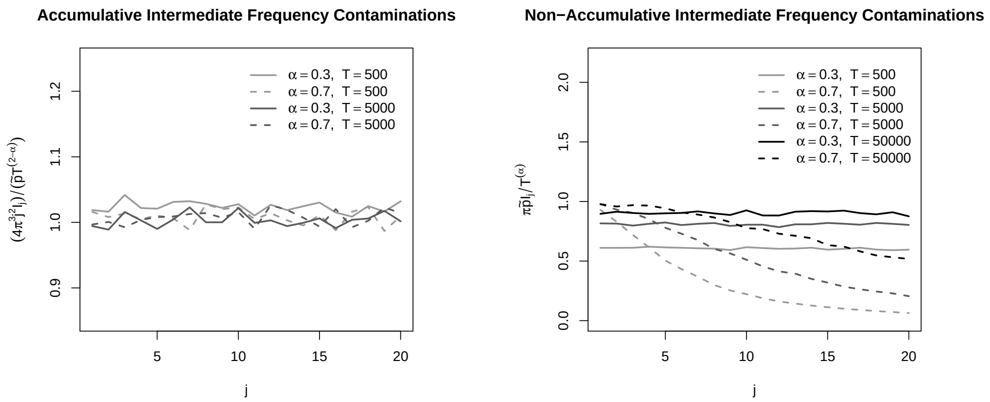

Figure 1: Average rescaled periodogram for accumulative and non-accumulative processes with intermediate frequency contaminations. In the left plot the simulated DGP is $\mu_t = \mu_{t-1} + \pi_t \eta_t$ and in the right plot it is $\mu_t = (1-\pi_t)\mu_{t-1} + \pi_t \eta_t$. In both cases $\pi_t \sim B(p)$, $p = 5/T^{\alpha}$, and $\eta_t \sim N(0,1)$.

As one can see, the orders for cases i.) and ii.) in Lemma 3 are identical to those in Lemma 2. This means that accumulative and non-accumulative structural change have the same impact on the periodogram local to zero, as long as the mean changes are deterministic or rare. In contrast to that, the case $\alpha < 1$ is remarkably different and needs to be treated separately. In presence of intermediate frequency contaminations, when the process becomes ergodic, the order of the peak is reduced to $T^{\alpha-\beta}$, instead of $T^{2-\alpha-\beta}$. Furthermore, the peak local to zero no longer decays for increasing j. This is a behaviour similar to a white noise reflecting the stationary structure of the non-accumulative approach.

Important special cases of both the accumulative and the non-accumulative process are obtained for $\alpha = 0$. In this case the accumulative process boils down to a unit root process and the non-accumulative process becomes a simple stationary short memory process. In this situation, case ii.) in Lemma 2 and iii.) in Lemma 3 reduces to the well known result that the periodogram of the unit root process local to the origin is of order $O_P(T^2/j^2)$ and that of the short memory process is $O_P(1)$.

The precision of the statements in case iii.) of Lemmas 2 and 3 in finite samples is investigated in a small Monte Carlo study. The results are shown in Figure 1. As examples for accumulative and non-accumulative structural-change processes with intermediate frequency contaminations, we simulate the random level-shift process $\mu_t = \mu_{t-1} + \pi_t \eta_t$ and its stationary counterpart $\mu_t = (1-\pi_t)\mu_{t-1} + \pi_t \eta_t$. In both cases $\pi_t \sim B(p)$, $p = 5/T^{\alpha}$, and $\eta_t \overset{iid}{\sim} N(0,1)$.

To analyze the accuracy of the asymptotic approximations in Lemmas 2 and 3, we calculate the average periodogram from 5,000 realizations of the respective process for different sample sizes and standardize it with the rate implied by the theorems. The resulting rescaled averaged periodogram is expected to be flat, if the theorem applies. Since the results are obtained under the assumption that $j/T \to 0$, it can be expected that the accuracy of the approximation is decreasing in j. On the left-hand side of Figure 1, it can be seen that the asymptotic approximation for accumulative structural-change processes in Lemma 2 ii.) with $\alpha < 1$ is precise for small as well as for larger samples. On the other hand, the plot on the right-hand side of Figure 1 shows that the results in Lemma 3 for non-accumulative structural-change processes hold up well in finite samples for $\alpha = 0.3$, but the approximation seems to be much more imprecise for $\alpha = 0.7$.

The explanation for this effect can be seen from (19) in the proof of Lemma 3, where the first term of order $O(T^{\alpha-\beta})$ is the dominating one that drives the asymptotic result, but there is an approximation error of order $O(T^{-2+3\alpha-\beta})$, whose impact vanishes only slowly if α is relatively large.

Nevertheless, it can be seen that the average rescaled periodogram converges to its predicted value as T increases.

3 Robust Fractional Cointegration Rank Estimator

In this section we first provide evidence that existing semiparametric estimators and tests for the fractional cointegration rank are biased in case of low frequency contaminations. We then derive our robust fractional cointegration rank estimator as an extension of the rank estimator by Robinson and Yajima (2002).

Point of departure is a vector valued long memory process y_t with low frequency contaminations. For expositional simplicity we focus on a bivariate system, i.e. $y_t = (y_{at}, y_{bt})'$, extensions to higher dimensions are straightforward. The process under investigation is

$$y_t = x_t + \mu_t, \tag{3}$$

where $\mu_t = (\mu_{at}, \mu_{bt})'$ is a bivariate low frequency contamination process that is independent of x_t and where for μ_{at} and μ_{bt} either of the Assumptions A1, A2, or A3 with $\beta = 0$ holds. Moreover, $x_t = (x_{at}, x_{bt})'$ is a bivariate long memory process whose spectral density matrix $f(\lambda_j)$ at frequency λ_j fulfills

$$f(\lambda_j) \sim \Lambda_j(d) G_x \Lambda_j^*(d), \tag{4}$$

where $\Lambda_j(d) = diag(\lambda_j^{-d_a} e^{i(\pi-\lambda_j)d_a/2}, \lambda_j^{-d_b} e^{i(\pi-\lambda_j)d_b/2})$ with $i = \sqrt{-1}$ and $d = (d_a, d_b)$ are the memory parameters. Further, $A \sim B$ denotes that $A/B \to 1$, as $\lambda \to 0_+$ and A^* denotes the complex conjugate of A.

As shown in Marinucci and Robinson (2001), the matrix G_x is positive definite if and only if x_t is not (fractionally) cointegrated and it becomes singular otherwise. Consequently, by estimating the rank of the G_x matrix we can investigate whether the time series are (fractionally) cointegrated as suggested by Robinson and Yajima (2002). However, if instead of the pure memory process we observe a contaminated process such as (3), then our estimate of the rank will be based on G_y which comprises the influence of G_x and G_μ.

To illustrate this, note that the periodogram as an estimate of G_y is given by $I_y(\lambda_j) = I_\mu(\lambda_j) + I_x(\lambda_j) + I_{\mu x}(\lambda_j) + I_{x\mu}(\lambda_j)$, where $I_{\mu x}(\lambda_j) = w_\mu(\lambda_j) w_x^*(\lambda_j)$ and $I_{x\mu}(\lambda_j) = w_x(\lambda_j) w_\mu^*(\lambda_j)$

are the cross-periodograms of μ_t and x_t, so that $E[I_y(\lambda_j)] = E[I_\mu(\lambda_j)] + E[I_x(\lambda_j)]$ if x_t and μ_t are assumed to be independent.

For the long memory component it holds that $E[I_x(\lambda_j)] = \left(\frac{T}{j}\right)^{2\max\{d_a,d_b\}} G_x$ as $j/T \to 0$. The properties of $E[I_\mu(\lambda_j)]$ can be derived based on our results presented in Section 2. Here we are interested in the empirically relevant situations of a smooth trend, a deterministic break or a random level shift process with rare shifts. These low frequency contaminations can be distinct, i.e. each series faces different contaminations, or they can be common as discussed in Hendry and Massmann (2007). Their definition is limited to contemporaneous mean cobreaking, i.e. common deterministic structural changes. Furthermore, their definition refers to changes relative to some initial parametrization, so that processes with a stable monotonous trend for example are not included. We therefore propose the following slightly modified definition:

Definition 1 (CLFC)**.** *The bivariate process y_t in (3) has common low frequency components if there exists a 2×1 matrix $\Phi \neq 0_{21}$, such that $\Phi'(\mu_t - \mu_1) = 0$ for all $t = 1, ..., T$.*

Now we are able to derive the order of the expected pseudo-periodogram of the trend process.

Theorem 1. *Suppose y_t is generated by (3) and $j/T \to 0$ we have*

$$E[I_\mu(\lambda_j)] = \frac{T}{j^2} G_\mu,$$

and G_μ has rank 1 if and only if μ_t is a common low frequency component according to Definition 1.

It is obvious from the rate in the theorem that the G_μ matrix dominates the G_y matrix for low frequencies while the G_x matrix is the dominating one for higher frequencies. If we now observe time series which are not fractionally cointegrated but exhibit joint breaks, then the estimator by Robinson and Yajima (2002) might spuriously identify a fractional cointegration relation since the G_μ matrix is singular. On the other hand, if we observe fractionally cointegrated time series which exhibit distinct breaks, then the same estimator might wrongly identify no fractional cointegration relation since the G_μ matrix has full rank. These problems do not only arise for the estimator by Robinson and Yajima (2002) but for all existing semiparametric estimators and tests concerning the fractional cointegration relation since all of them use the G_y matrix in some form. We will demonstrate this by simulations in Section 5. It should further be noted that testing the homogeneity of fractional difference parameters as suggested by Robinson and Yajima (2002) is also not possible in the case of low frequency contaminations. This is due to two reasons. First, the estimates of the memory parameter will be biased when using the standard local Whittle estimator. This issue can be overcome by considering a robust estimator such as those by Iacone (2010), McCloskey and Perron (2013) or Hou and Perron (2014). However, the test statistics also includes an estimate of the G_x matrix which is based on the first m_1 frequencies. In the case of joint breaks this matrix might be estimated to be singular letting the statistics converge to zero no matter if the order of integrations are truly equal.

Let us now introduce an approach to estimate the G_x matrix consistently also in the case of low frequency contaminations. We know from Theorem 1 that the G_μ matrix only dominates for the low frequencies. If we trim these away then we can estimate the rank of the G_x matrix without distortions no matter if low frequency contaminations are present or not. We first show that the estimated G_y matrix trimmed by the first frequencies converges to the estimated G_x matrix. For this purpose we need to introduce the following assumptions.

Assumption B1. *It holds that*

$$x_t - E[x_t] = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j},$$

with $\sum_{j=0}^{\infty} \|A_j\|^2 < \infty$ *and* $\|\cdot\|$ *denotes the supremum norm. It is assumed that* $E[\varepsilon_t|\mathfrak{F}_{t-1}] = 0$*,* $E[\varepsilon_t\varepsilon_t^{'}|\mathfrak{F}_{t-1}] = I_q$ *a.s. for* $t = 0, \pm 1, \pm 2, \ldots$ *where* $\mathfrak{F}_t$ *denotes the* σ*-field generated by* ε_s *and* I_q *is an identity matrix,* $s \leq t$*. Furthermore, there exists a scalar random variable* ε *such that* $E[\varepsilon^2] < \infty$ *and for all* $\tau > 0$ *and some* $C > 0$ *it is* $P(\|\varepsilon_t\|^2 > \tau) \leq CP(\varepsilon^2 > \tau)$*.*

Assumption B2. *As* $T \to \infty$*,*

$$\frac{l}{m_1} + \frac{m_1}{T} \to 0,$$

where l *is a trimming parameter with* $l = max(1, [c_l T^{\delta_l}])$ *and* $m_1 = max(l+1, [c_{m_1} T^{\delta_{m_1}}])$ *is the bandwidth parameter with* $0 \leq \delta_l < \delta_{m_1} < 1$ *and* $c_l, c_{m_1} \in (0, \infty)$*.*

Let us further assume for the moment that the order of integration is known and denote

$$\begin{aligned}
\hat{G}_y(d,l,m_1) &= (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_y(\lambda_j) \Lambda_j^*(d) \\
&= (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_x(\lambda_j) \Lambda_j^*(d) + (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_\mu(\lambda_j) \Lambda_j^*(d) \\
&+ (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_{x\mu}(\lambda_j) \Lambda_j^*(d) + (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_{\mu x}(\lambda_j) \Lambda_j^*(d)
\end{aligned}$$

and $\hat{G}_x(d,l,m_1) = (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_x(\lambda_j) \Lambda_j^*(d).$

Theorem 2. *Suppose* y_t *is generated by (3) and Assumptions B1 and B2 hold, using* $m_1 = T^{\delta_{m_1}}$ *and* $l = T^{\delta_l}$ *for some* $0 \leq \delta_l < \delta_{m_1} < 1$*, and for* $T \to \infty$*,*

$$\hat{G}_y(d,l,m_1) \xrightarrow{p} \hat{G}_x(d,1,m_1),$$

if either

i.) $l = 1$ *and* $d_a^0 + d_b^0 > 1$*, or* $l = 1$*,* $d_a^0 + d_b^0 < 1$*, and* $\delta_{m_1} > 1 - d_a^0 - d_b^0$*.*

ii.) $d_a^0 + d_b^0 < 1$, $l = O\left(T^{(d_a^0+d_b^0-1)/(d_a^0+d_b^0-2)+\upsilon}\right)$ *for some* $\upsilon > 0$, *and* $(d_a^0 + d_b^0 - 1)/(d_a^0 + d_b^0 - 2) + \upsilon < ((d_a - d_a^0) + (d_b - d_b^0) + \delta_{m_1})/((d_a - d_a^0) + (d_b - d_b^0) + 1)$.

Here and in the rest of the paper the superscript 0 denotes the true value of a parameter, for example d_a^0 is the true memory parameter of series a.

The first part of the theorem shows that for nonstationary long memory processes the long memory component always dominates the mean component and no trimming is needed. Here, we can use the procedure by Nielsen and Shimotsu (2007) to determine the fractional cointegration rank even when low frequency contaminations are present. In the case of stationary long memory, however, this is not the case. Here, the second part of the condition gives the frequency from which onwards the long memory component becomes dominant. This depends on the true order of integration which we assumed to be known so far. If this is not the case a feasible choice would be to trim away $l = \sqrt{T}$ frequencies. We could also estimate d using univariate approaches that are robust to low frequency contaminations and then choose l based on these estimates. However, unreported simulations indicate that setting $l = \sqrt{T}$ yields superior results.

Nevertheless, we still have to estimate d for determining $\hat{G}_y(d, l, m_1)$. For this purpose, we need an estimator that is robust to low frequency contaminations and converges with a rate of $\log m$ which is the standard rate for semiparametric estimators. Moreover, as discussed in Robinson and Yajima (2002) and Nielsen and Shimotsu (2007) we cannot rely on multivariate estimators since these require knowledge of the cointegration rank and we require an estimate of d that converges faster than the estimate of G such that the effect of estimating d vanishes asymptotically. Possible estimators are those of Iacone (2010), McCloskey and Perron (2013) or Hou and Perron (2014). Denote the bandwidth for estimating d by m for which the following assumption holds.

Assumption B3. *For any* $\psi > 0$

$$\frac{m_1^{1/2-\psi} T^{\psi}}{m^{1/2}} + \frac{m^{1+2\psi} log(m)^2}{T^{2\psi}} \to 0, \text{ as } T \to \infty.$$

Theorem 3. *Suppose* y_t *is generated by (3) and Assumptions B1 to B3 hold and let* $\hat{d}(m) - d^0 = o(\log m)$, *then*

$$\hat{G}_y(\hat{d}(m), l, m_1) \xrightarrow{p} \hat{G}_y(d^0, l, m_1).$$

Theorem 3 in conjunction with 2 ii.) and Proposition 3 of Robinson and Yajima (2002) implies that G can be estimated consistently and $\hat{G}_y(\hat{d}(m), l, m_1)$ is asymptotically Gaussian given the additional assumptions made in Robinson and Yajima (2002). Note that these assumptions restrict the series to be stationary. This might seem restrictive but as discussed before for nonstationary series low frequency contaminations are not troublesome such that the standard extension by Nielsen and Shimotsu (2007) can be considered.

We can then follow the route of Robinson and Yajima (2002) to estimate the fractional cointegration rank of the series. For the sake of completeness we will briefly outline the steps.
We can test whether two series have equal memory using

$$\hat{T}_{TRE} = \frac{m^{1/2}(\hat{d}_a - \hat{d}_b)}{(1/2(1-\hat{G}_{ab}^2/(\hat{G}_{aa}\hat{G}_{ab})))^{1/2} + n(T)},$$

where G_{ab} are the respective elements of the estimated matrix $\hat{G}_y(\hat{d}(m), l, m_1)$, d_a and d_b are estimated using a robust estimator as discussed above, and $n(T) > 0$. Consistency of the test follows under the same additional assumptions as in Robinson and Yajima (2002).
If the test indicates the two series to have equal memory we can then determine whether they are fractionally cointegrated by estimating the fractional cointegration rank. To do so denote by $q_{1,G}$ the first eigenvalue of $G_y(\hat{d}(m), l, m_1)$ and let $\hat{q}_{1,G}$ denote its empirical counterpart. If rank(A) $= 2$ we have $q_{1,A} > q_{2,A} > 0$ whereas for rank(A) $= 1$ it is $q_{1,A} > q_{2,A} = 0$. Define furthermore $\sigma_{v,G} = \sum_{i=1}^{v} q_{i,G}$ and $N(T) > 0$ such that $N(T) + m_1^{-1/2} N(T)^{-1} \to 0$ as $T \to \infty$ we can estimate the fractional cointegration rank by minimizing the loss function

$$L(u) = N(T)(2-u) - \hat{\sigma}_{2-u,G}.$$

The estimator for the fractional cointegration rank is

$$\widehat{rg}_{TRE} = argmin_{u=0,1} L(u).$$

Again, consistency of this estimator follows directly from Robinson and Yajima (2002) given the same additional assumptions.

4 Robust Multivariate Local Whittle Estimator

After determining the fractional cointegration rank we aim to estimate the cointegration vector and the memory parameter robust to possible low frequency contamination. In this section we obtain a robust local Whittle estimator for the parameter θ containing of the memory parameters $d = (d_a, d_b)$ and the possible cointegration vector β.
It should be mentioned that our robust local Whittle estimator depends on the a priori specified fractional cointegration rank as much as the original local Whittle estimator in Robinson (2008a). The cointegration rank is needed to know the dimension of the parameter to be estimated. It does not enter the estimation procedure as a nuisance parameter. Therefore, we see the estimation of the cointegration rank in a first step as part of the model specification procedure and find the assumption of a known cointegration rank when it comes to parameter estimation justified. The same assumption is implicitly used by Robinson (2008a) and Shimotsu (2012) and is therefore standard

in the literature. A short Monte Carlo analysis underpinning this can be found in the Online Appendix.

The expectation of the periodogram of the contaminated multivariate long memory process has an additive structure. Consequently, a robust multivariate local Whittle estimator can be constructed in a similar fashion as for the fractional cointegration rank estimator, i.e. by trimming away the frequencies dominated by the low frequency contamination. Our estimator is then based on the univariate trimmed local Whittle estimator by Iacone (2010). In Section 2 we showed that $O_P(Tj^{-2})$ is an upper bound for the pole of the periodogram at the zero frequency for a fairly general class of processes. Therefore, trimming the periodogram by the first $l = \sqrt{T}$ frequencies eliminates the influence of the low frequency contaminations and leads to a robust estimate of the memory parameter.

We aim to estimate the parameter $\theta = (d, \beta)'$, where we restrict the series to be stationary, i.e. $-1/2 < d_a, d_b < 1/2$. As discussed in Section 3, for nonstationary time series trimming is not needed. Our trimmed multivariate local Whittle estimator is defined by the objective function using here and in what follows the superscript *tri* to indicate the trimmed version

$$R(\theta) = \log\det\hat{\Omega}^{tri}(\theta) - 2(d_a + d_b)\frac{1}{m-l+1}\sum_{j=l}^{m}\log\lambda_j$$

with

$$\Omega^{tri}(\theta) = \frac{1}{m-l+1}\sum_{j=l}^{m} Re[\Lambda_j(d) B I_y^{tri}(\lambda_j) B' \Lambda_j^*(d)]$$

and

$$B = \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix}.$$

It is

$$\hat{\theta} = argmin R(\theta).$$

To show consistency of this estimator we need to make the following assumptions.

Assumption C1. *As* $\lambda \to 0_+$

$$f_{x,ab}(\lambda) = \exp\left(i\pi\left(d_a^0 - d_b^0\right)/2\right)\lambda^{-d_a^0 - d_b^0} G_{x,ab}^0 = O\left(\lambda^{-d_a^0 - d_b^0}\right),$$

where $f_{x,ab}$ *and* $G_{x,ab}$ *are the respective element of the matrices* f_x *and* G_x *of* x_t*.*

Assumption C2. *Assumption B1 holds.*

Assumption C3. *In a neighborhood* $(0, \alpha)$ *of the origin,* $A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$ *is differentiable and*

$$\frac{\partial}{\partial \lambda}\, {}_aA(\lambda) = O\left(\lambda^{-1}\|{}_aA(\lambda)\|\right), \qquad \lambda \to 0_+,$$

where ${}_aA(\lambda)$ is the a-th row of $A(\lambda)$.

Assumption C4. *As $T \to \infty$,*

$$\frac{l}{m} + \frac{m}{T} \to 0,$$

where l is a trimming parameter with $l = max(1, [c_l T^{\delta_l}])$ and $m = max(l+1, [c_m T^{\delta_m}])$ is the bandwidth parameter with $0 \le \delta_l < \delta_m < 1$ and $c_l, c_m \in (0, \infty)$.

Assumptions C1, C2, and C3 are analogous to Assumptions A1 to A3 of Lobato (1999) respectively Assumptions 1 to 3 of Shimotsu (2007) and Assumption C4 corresponds to A4 of Shimotsu (2007). We furthermore denote in what follows $\nu^0 = d_b^0 - d_a^0$.

Theorem 4. *Suppose y_t is generated by (3) and Assumptions C1 to C4 hold with the trimming parameter $l = \sqrt{T}$ it is*

$$\hat{d} - d^0 \xrightarrow{P} 0, \qquad \hat{\beta} = \beta^0 + o_P\left(\left(\frac{m}{T}\right)^{\nu^0}\right).$$

As usual, the assumptions required to prove the normality of the estimator are somewhat stronger than those needed for consistency. Here, we assume

Assumption D1. *For $\zeta \in (0,2]$ as $\lambda \to 0_+$*

$$f_{x,ab}(\lambda) = \exp\left(i(\pi - \lambda)\left(d_a^0 - d_b^0\right)/2\right)\lambda^{-d_a^0 - d_b^0} G_{x,ab}^0 = O\left(\lambda^{-d_a^0 - d_b^0 + \zeta}\right).$$

Assumption D2. *Assumption B1 holds and in addition, it holds for $a,b,c,d = 1,2$, $t = 0, \pm 1, \pm 2, \ldots$ that*

$$E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}|\mathfrak{F}_{t-1}) = \mu_{abc} \qquad a.s.$$

and

$$E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}\varepsilon_{dt}|\mathfrak{F}_{t-1}) = \mu_{abcd} \qquad a.s.,$$

where $|\mu_{abc}| < \infty$ and $|\mu_{abcd}| < \infty$.

Assumption D3. *Assumption C3 holds.*

Assumption D4. *As $T \to \infty$ it holds for any $\tau > 0$*

$$\frac{l}{m} + \frac{m^{1+2\tau}(\log m)^2}{T^{2\tau}} + \frac{\log T}{m^{\tau}} \to 0.$$

Assumption D5. *There exists a finite real matrix Q such that*

$$\Lambda_j(d^0)^{-1}A(\lambda_j) = Q + o(1), \qquad \lambda_j \to 0.$$

These assumptions allow for non-Gaussianity. Assumption D1 and D5 are satisfied by multivariate ARFIMA processes. Assumption D4 is necessary for the Hessian of the objective function of the local Whittle function to converge. It should be mentioned that Assumption D4 gives a sharp upper bound for the number of frequencies m which can be used for the local Whittle estimator. It is $m = o(T^{0.8})$.
For the estimator we obtain the following results.

Theorem 5. *Suppose y_t is generated by (3) and Assumptions D1 to D5 hold with $T \to \infty$ and $\Delta_T = diag(\lambda_m^{-\nu^0}, 1, 1)$, then*

$$\sqrt{m}\Delta_T(\hat{\theta} - \theta^0) \xrightarrow{d} N(0, \Xi^{-1})$$

with $\Xi_{aa} = 2\tilde{\mu}[(1-2\nu^0)^{-1} - (1-\nu^0)^{-2}cos^2(\tilde{\gamma}^0)]G_{bb}/G_{aa}$, $\Xi_{ab} = \Xi_{21} = -2\tilde{\mu}\nu^0(1-\nu^0)^{-2}cos(\tilde{\gamma}^0)$ $G_{ab}/G_{aa} + (\pi/2)2\tilde{\mu}(1-\nu^0)^{-1}sin(\tilde{\gamma}^0)(G_{ab}/G_{aa})$, $\Xi_{13} = \Xi_{31} = -\Xi_{ab}$, $\Xi_{bb} = \Xi_{33} = 4 + (\pi^2/4 - 1)2\tilde{\mu}\rho^2$, $\rho = G_{ab}/(G_{aa}G_{bb})^{1/2}$, $\tilde{\mu} = (1-\rho^2)^{-1}$ and $\tilde{\gamma}^0 = (\pi/2)\nu^0$.

Consequently, the estimator is consistent and asymptotically normal with the same limiting variance as the GSE estimator in Robinson (2008b). We can therefore robustify the estimator without an asymptotic loss in efficiency. If $G_\mu = 0$, the estimator is reduced to the standard multivariate estimator of Robinson (2008b).

5 Monte Carlo

In this section we show the behavior of our proposed methods in finite samples by means of a simulation study. We first investigate the behaviour of our robust fractional cointegration rank estimator and then consider the robust local Whittle estimator. Our framework as stated in Section 2 allows for various forms of low frequency contaminations. For the ease of the presentation we present results for nonstationary random level shift processes with rare shifts in the following as this seems to be the empirical most relevant case and move the qualitatively similar results for stationary random level shift processes and deterministic trends to Tables OA.1-OA.6 in the Online Appendix.
The bivariate stationary long memory process with random level shifts considered in the following can be delineated as

$$y_{at} = \zeta_a\mu_t + \xi\tilde{\mu}_t + x_t + (1-L)^{-(d-\tilde{d})}u_t \tag{5}$$
$$y_{bt} = \zeta_b\mu_t + x_t, \tag{6}$$

with

$$\mu_t = \mu_{t-1} + \pi_t\eta_t, \qquad \pi_t \sim B(5/T), \qquad \eta_t \sim N(0,1), \tag{7}$$
$$\tilde{\mu}_t = \tilde{\mu}_{t-1} + \tilde{\pi}_t\tilde{\eta}_t, \qquad \tilde{\pi}_t \sim B(5/T), \qquad \tilde{\eta}_t \sim N(0,1), \qquad \text{and} \tag{8}$$
$$x_t = (1-L)^{-d}e_t, \qquad \begin{pmatrix} e_t \\ u_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right). \tag{9}$$

Here, L is the usual lag operator such that $(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k$, with $\binom{d}{k} = \frac{d(d-1)(d-2)...(d-(k-1))}{k!}$. This model allows for fractional cointegration, distinct structural breaks and joint structural breaks.

We present results for orders of integration of $d = 0.2, 0.4$. Concerning the fractional cointegration component, we investigate the case of no fractional cointegration with $\tilde{d} = 0$ and the case of fractional cointegration with $\tilde{d} = d$ and the cointegrating vector being $\beta = (1, -1)'$.

Concerning the low frequency contamination component, we investigate the situations of no low frequency contaminations, i.e. $\zeta_a = \zeta_b = \xi = 0$, of distinct structural breaks, i.e. $\zeta_a = 0$ but $\zeta_b = \xi = 1$, and of joint breaks with $\zeta_a = \zeta_b = 1$ and $\xi = 0$, i.e. the cobreaking vector is given by $\gamma = (1, -1)'$. Break sizes are random with mean zero and variance one and they occur with probability $5/T$. This means that on average there are five breaks in the sample.

We further present results for series whose errors are cross-sectionally uncorrelated ($r = 0$) and for series whose errors are cross-sectionally correlated with $r = 0.5$. Finally, we consider sample sizes of $T = 250, 1000$ and all presented results are the averages over $5,000$ replications.

5.1 Fractional Cointegration

We first consider the results for the estimation of the fractional cointegration rank. To put the performance of our estimator into perspective and to validate our claim in Section 3 that all existing procedures to identify fractional cointegration are seriously distorted for processes with low frequency contaminations, we also consider the performance of all other procedures applicable when the series exhibit stationary long memory. This includes the fractional cointegration rank estimator by Robinson and Yajima (2002) (RY02) and the fractional cointegration tests by Chen and Hurvich (2006) (CH06), Robinson (2008a) (R08), and Souza et al. (2018) (SRF). Parameter values are all chosen according to the authors recommendation. Note that R08 requires $r > 0$ so that we do not report results for $r = 0$

For our robust estimator, abbreviated TRE in the table, we need to choose l, m, m_1, N, and the type of robust univariate estimator. Unreported simulations indicated that $l = T^{0.5}$, $m = T^{0.75}$, $m_1 = T^{0.7}, N = m^{-0.2}$, and the univariate estimator by Iacone (2010) yielded the best trade-off between correctly and spuriously identifying fractional cointegration in our simulations. We therefore show the results for this parameter combination in the following and recommend it to be considered in empirical applications.

The results can be found in Table 1. In the table "NO" indicates no low frequency contaminations, "DIS" means that the series exhibit distinct breaks, and "COB" refers to joint breaks. Furthermore, "TRUE" means that the series are fractionally cointegrated whereas "FALSE" indicates that they are not. We then state the mean estimated cointegration rank for the rank estimators, which should be 1 for "TRUE" and 0 otherwise, and the mean rejection rate for the tests, which should be 1 for "TRUE" and 0.05

			cointegration	TRUE					FALSE				
r	d	T	breaks	TRE	RY02	CH06	R08	SRF	TRE	RY02	CH06	R08	SRF
0	0.2	250	NO	1.00	0.99	0.29	-	0.09	0.00	0.00	0.16	-	0.00
			DIS	0.96	0.35	0.24	-	0.10	0.00	0.00	0.51	-	0.16
			COB	1.00	1.00	0.81	-	0.66	0.00	0.02	0.75	-	0.37
		1000	NO	1.00	1.00	0.63	-	0.29	0.00	0.00	0.18	-	0.00
			DIS	0.99	0.23	0.44	-	0.54	0.00	0.00	0.63	-	0.61
			COB	1.00	1.00	0.97	-	0.92	0.00	0.00	0.90	-	0.72
	0.4	250	NO	1.00	1.00	0.80	-	0.39	0.00	0.00	0.23	-	0.00
			DIS	0.98	0.68	0.18	-	0.10	0.00	0.00	0.38	-	0.04
			COB	1.00	1.00	0.93	-	0.71	0.00	0.00	0.52	-	0.10
		1000	NO	1.00	1.00	1.00	-	0.90	0.00	0.00	0.19	-	0.01
			DIS	1.00	0.78	0.22	-	0.42	0.00	0.00	0.41	-	0.13
			COB	1.00	1.00	1.00	-	0.98	0.00	0.00	0.57	-	0.22
0.5	0.2	250	NO	1.00	1.00	0.18	0.01	0.10	0.35	0.07	0.08	0.09	0.01
			DIS	1.00	0.59	0.15	0.87	0.03	0.26	0.03	0.39	0.62	0.11
			COB	1.00	1.00	0.60	0.01	0.55	0.47	0.53	0.68	0.27	0.45
		1000	NO	1.00	1.00	0.46	0.14	0.25	0.01	0.00	0.07	0.08	0.01
			DIS	1.00	0.50	0.34	0.97	0.21	0.01	0.00	0.55	0.76	0.54
			COB	1.00	1.00	0.92	0.08	0.88	0.03	0.09	0.83	0.67	0.77
	0.4	250	NO	1.00	1.00	0.62	0.03	0.34	0.35	0.08	0.10	0.10	0.01
			DIS	1.00	0.83	0.10	0.68	0.03	0.26	0.03	0.26	0.42	0.03
			COB	1.00	1.00	0.82	0.03	0.60	0.46	0.34	0.38	0.08	0.15
		1000	NO	1.00	1.00	0.99	0.60	0.83	0.01	0.00	0.07	0.08	0.01
			DIS	1.00	0.88	0.16	0.79	0.16	0.01	0.00	0.30	0.51	0.10
			COB	1.00	1.00	1.00	0.54	0.95	0.02	0.02	0.43	0.20	0.30

Table 1: Mean estimated fractional cointegration rank for a bivariate fractionally integrated system. The DGP is based on Equations (5) to (9). In case of fractional cointegration $\beta = (-1,1)'$ and $b = d$. Break sizes are random with mean zero and variance one, they occur with probability $5/T$, and in case of joint breaks $\gamma = (-1,1)'$. Our estimator is considered with $l = T^{0.5}$, $m = T^{0.75}$, $m_1 = T^{0.7}$, $N = m_1^{-0.2}$, and the univariate estimator by Iacone (2010) to estimate d and for the procedures by Robinson and Yajima (2002) (RY02), Chen and Hurvich (2006) (CH06), Robinson (2008a) (R08), and Souza et al. (2018) (SRF) parameter values are chosen according to the authors recommendation.

(the significance level) otherwise as all of the procedures test the null of no fractional cointegration.

Table 1 shows that our procedure works well for all of the considered scenarios. It correctly identifies fractional cointegration respectively no fractional cointegration in most of the cases. In Section 3 we mentioned that two scenarios are of particular importance, the case when no fractional cointegration is present but joint breaks, as standard procedure might spuriously indicate fractional cointegration, and the case of fractional cointegration but distinct breaks, as standard procedure might have problems detecting the fractional cointegration relation in this case. The table shows that our estimator works well in both cases with minor distortions for the first case in a small sample size of $T = 250$. When increasing the sample size these vanish completely. In contrast to this, all other procedures show problems in at least one of the two cases. In the first case, the rejection rates of CH06, R08 (for $r = 0.5$), and SRF increase with increasing T implying that asymptotically the tests will always indicate fractional cointegration in the case of joint breaks, no matter if the series are truly fractionally cointegrated. RY02 has serious problems in the second case where at least for $d = 0.2$ the estimated cointegration rank decreases with increasing T implying that asymptotically the estimator will always indicate no fractional cointegration in the case of distinct breaks, no matter if the series are truly fractionally cointegrated.

Consequently, there is a risk in empirical applications that these procedures miss a fractional cointegration relation due to structural breaks or falsely indicate a fractional cointegration relation due to joint breaks in the series. In contrast, our procedure delivers the required robustness to correctly detect whether fractional cointegration is present in the case of low frequency contaminations while also performing well in the case that they are not. Here, the results are comparable to those achieved by RY02 which outperforms all other procedures in this setup.

The table further reveals that the performance of our estimator does not depend on d. In contrast, the performance of the other estimators improves when increasing d as implied by Theorem 2.

Lastly, it can be seen that when introducing cross-sectionally correlated errors the likelihood of all procedures, except R08, to indicate the two time series as fractionally cointegrated increases. However, the ranking of the procedures and therefore the advantages that our estimator has in comparison to the other approaches stays the same.

5.2 Order of integration

After providing evidence that we can robustly determine the fractional cointegration rank also in small samples, we now focus on robust estimation of the memory parameter. We consider the same DGPs as before and again use $m = T^{0.75}$ and $l = T^{0.5}$ for robust estimation. Furthermore, we state the results for the standard multivariate estimator by Shimotsu (2007) and Robinson (2008b). For comparison, we assume the fractional cointegration rank to be known. Tables OA.7 and OA.8 in the Online Appendix show the qualitatively similar results when first estimating the fractional cointegration rank and then estimating the order of integration.

Table 2 has the bias and RMSE of the estimators when no fractional cointegration is

			Estimator	TMLW				GSE			
				Bias		RMSE		Bias		RMSE	
r	d	T	breaks	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$
0	0.2	250	NO	-0.03	-0.02	0.19	0.20	-0.02	-0.02	0.08	0.08
			DIS	0.02	0.01	0.20	0.20	0.18	0.18	0.22	0.22
			COB	0.01	0.01	0.19	0.20	0.17	0.17	0.20	0.20
		1000	NO	-0.02	-0.01	0.09	0.09	-0.01	-0.01	0.04	0.04
			DIS	0.01	0.01	0.09	0.09	0.17	0.17	0.19	0.19
			COB	0.01	0.01	0.09	0.09	0.16	0.16	0.18	0.18
	0.4	250	NO	-0.04	-0.04	0.20	0.20	-0.02	-0.02	0.08	0.08
			DIS	-0.01	-0.01	0.19	0.20	0.08	0.08	0.12	0.12
			COB	-0.01	-0.01	0.19	0.19	0.07	0.07	0.11	0.11
		1000	NO	-0.02	-0.02	0.09	0.09	-0.01	-0.01	0.04	0.04
			DIS	-0.01	-0.01	0.09	0.09	0.06	0.06	0.09	0.09
			COB	-0.01	-0.01	0.09	0.09	0.06	0.06	0.08	0.08
0.5	0.2	250	NO	-0.02	-0.02	0.15	0.15	-0.01	-0.02	0.06	0.07
			DIS	0.03	0.03	0.16	0.16	0.20	0.20	0.23	0.23
			COB	0.00	0.00	0.15	0.15	0.13	0.13	0.17	0.17
		1000	NO	-0.01	-0.01	0.07	0.07	-0.01	-0.01	0.04	0.04
			DIS	0.02	0.02	0.08	0.08	0.19	0.19	0.21	0.21
			COB	0.00	0.00	0.07	0.07	0.13	0.13	0.15	0.15
	0.4	250	NO	-0.04	-0.04	0.16	0.16	-0.02	-0.02	0.06	0.07
			DIS	0.00	0.00	0.15	0.15	0.09	0.09	0.13	0.13
			COB	-0.02	-0.02	0.15	0.15	0.05	0.05	0.09	0.09
		1000	NO	-0.02	-0.02	0.07	0.07	-0.01	-0.01	0.04	0.04
			DIS	0.00	0.00	0.07	0.07	0.07	0.07	0.09	0.09
			COB	-0.01	-0.01	0.07	0.07	0.04	0.04	0.07	0.07

Table 2: Bias and RMSE of our trimmed multivariate local Whittle estimator and the standard multivariate local Whittle estimator in a bivariate fractionally integrated system. The DGP is based on Equations (5) to (9). Break sizes are random with mean zero and variance one, they occur with probability $5/T$, and in case of joint breaks $\gamma = (-1,1)'$. Moreover, we use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

present and Table 3 states the bias and RMSE when fractional cointegration is present.

			Estimator	TMLW						GSE					
				Bias			RMSE			Bias			RMSE		
r	d	T	breaks	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$
0	0.2	250	NO	-0.02	-0.03	-0.28	0.20	0.19	2.75	-0.02	-0.02	0.01	0.08	0.07	1.15
			DIS	0.08	0.01	0.11	0.22	0.19	2.82	0.37	0.17	1.32	0.39	0.21	3.88
			COB	-0.02	0.01	-0.21	0.20	0.19	2.65	-0.02	0.18	0.00	0.08	0.22	0.32
		1000	NO	-0.01	-0.02	-0.02	0.09	0.09	1.27	-0.01	-0.01	0.00	0.04	0.04	0.17
			DIS	0.06	0.01	0.13	0.12	0.09	1.84	0.37	0.16	1.50	0.39	0.18	3.92
			COB	-0.01	0.01	-0.03	0.09	0.09	1.18	-0.01	0.17	0.00	0.04	0.19	0.06
	0.4	250	NO	-0.02	-0.04	-0.12	0.19	0.20	1.99	-0.02	-0.01	0.00	0.08	0.08	0.18
			DIS	0.08	-0.01	0.05	0.22	0.19	2.16	0.38	0.07	0.58	0.40	0.12	3.15
			COB	-0.02	-0.02	-0.13	0.20	0.20	1.84	-0.02	0.08	0.00	0.08	0.13	0.09
		1000	NO	-0.01	-0.02	-0.01	0.09	0.09	0.25	-0.01	-0.01	0.00	0.04	0.04	0.05
			DIS	0.06	-0.01	-0.02	0.12	0.09	0.59	0.39	0.05	0.46	0.40	0.08	2.88
			COB	-0.01	-0.01	0.00	0.09	0.09	0.22	-0.01	0.06	0.00	0.04	0.09	0.03
0.5	0.2	250	NO	-0.03	-0.02	-2.58	0.18	0.18	5.28	-0.02	-0.02	-0.48	0.07	0.07	2.05
			DIS	0.12	0.04	-3.34	0.23	0.19	5.95	0.43	0.18	-2.62	0.44	0.22	5.92
			COB	-0.02	0.02	-2.29	0.18	0.18	4.98	0.03	0.18	-0.10	0.08	0.22	0.82
		1000	NO	-0.02	-0.01	-0.90	0.09	0.09	2.89	-0.01	-0.01	-0.05	0.04	0.04	0.38
			DIS	0.10	0.02	-3.09	0.15	0.09	5.51	0.42	0.17	-4.11	0.43	0.19	6.93
			COB	-0.01	0.01	-0.74	0.09	0.09	2.63	0.04	0.16	-0.03	0.06	0.19	0.10
	0.4	250	NO	-0.02	-0.03	-1.08	0.18	0.18	3.19	-0.02	-0.02	-0.03	0.08	0.07	0.17
			DIS	0.11	0.00	-2.15	0.23	0.18	4.55	0.42	0.09	-3.09	0.44	0.13	5.73
			COB	-0.01	0.00	-0.93	0.19	0.18	2.97	0.01	0.08	-0.02	0.07	0.12	0.11
		1000	NO	-0.01	-0.02	-0.07	0.09	0.09	0.38	-0.01	-0.01	-0.01	0.04	0.04	0.05
			DIS	0.10	0.00	-0.53	0.15	0.09	1.64	0.42	0.07	-4.53	0.43	0.09	6.54
			COB	-0.01	-0.01	-0.06	0.09	0.09	0.43	0.02	0.06	-0.01	0.04	0.08	0.04

Table 3: Bias and RMSE of our trimmed multivariate local Whittle estimator and the standard multivariate local Whittle estimator in a bivariate fractionally cointegrated system with cointegration vector $\beta = (1,-1)'$. The DGP is based on Equations (5) to (9). Break sizes are random with mean zero and variance one, they occur with probability $5/T$, and in case of joint breaks $\gamma = (-1,1)'$. Moreover, we use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

In the case of no fractional cointegration, we estimate the two memory parameters d_1 and d_2 of the two time series. Table 2 shows that in this case bias and RMSE of our estimator are small in all scenarios and that they decrease with increasing sample size. It is further revealed that the processes without breaks and with joint breaks typically result in a small negative bias in small samples, while the process with distinct breaks causes a small positive bias in small samples. Since increasing d seems to decrease the

bias while increasing r seems to increase it, changing these two parameters can increase or decrease the accuracy of the estimation depending on which process is considered, i.e. whether the process comes with a slightly negative or positive bias. It needs to be noted, however, that the absolute value of the bias itself and therefore also the changes in bias are of such a small magnitude in all scenarios that they barely reflect in the RMSE.
If we consider the standard estimator we can see that they are upward biased in case of low frequency contaminations. The bias is large for $d = 0.2$ and does not seem to vanish asymptotically when increasing T. However, it does decrease when increasing d as again implied by Theorem 2. The table further reveals that the variance of the standard estimator is lower than those of our procedure since more frequencies are considered for estimation. When increasing T the difference becomes smaller as implied by Theorem 5.
In the case of fractional cointegration we estimate the memory parameter d, the reduction of memory through the fractional cointegration relation $d - \tilde{d}$, and the cointegrating vector β. As Table 3 shows, for the estimation of the memory parameter the same conclusion as for the case without a cointegration relation can be drawn. Bias and RMSE are small even for $T = 250$ and they decrease with increasing sample size. The influence of the break process, d, and r on the estimation result also stays the same. Concerning the estimation of $\widehat{d - \tilde{d}}$ it can be observed that the estimator works well when considering the processes without breaks or with joint breaks. For the process with distinct breaks there is a positive bias which vanishes with increasing sample size. It should be noted that we could further decrease this bias by increasing l. This, however, comes at the cost of an increased variance of the estimator. Concerning the estimate of the cointegration vector β we can see that the estimator works well in large samples, i.e. $T \geq 1{,}000$. Here, bias and RMSE are rather small and we only observe a meaningful bias and variance for the distinct break process and cross-sectionally correlated errors. This, however, vanishes if we further increase T. For $T = 250$ and $r = 0$ the estimator performs well although it exhibits some variance. The only case where substantial distortions are present is for $T = 250$ and $r = 0.5$ where the estimator seems to be negatively biased. However, as mentioned before, the bias disappears for larger samples.
As in the case of no fractional cointegration the standard estimator shows a substantial positive bias which decreases in d but not in T when low frequency contaminations are present. The estimate of $d - \tilde{d}$ is accurate in the case of joint breaks but enormously upward biased in the case of distinct breaks. The same conclusion can also be drawn for the estimate of β with the difference that the bias is positive in case of $r = 0$ and negative for $r = 0.5$. As before it can again be observed that the variance of the standard estimator is smaller than of our procedure resulting in smaller RMSEs when no low frequency contaminations are present.
To summarize, the standard estimators are upward biased in the case of low frequency contaminations. In contrast, our estimator is robust to low frequency contaminations in the case of no fractional cointegration as well as in the case of fractional cointegration.

The price for this robustness is an increased variance which might be problematic when estimating β in small samples as the variance of the standard estimator is already large in this case. When increasing T this problem disappears making our estimator well suited for estimating the order of integration and the cointegration vector.

6 Empirical Example

To demonstrate the empirical relevance of our procedures we consider an example investigating the daily realized beta of two American stocks, namely Chevron (CVX) and ExxonMobil (XOM), relative to the $S\&P500$ between January 1996 and February 2017 ($T = 5,238$). Realized betas measure the systematic risk of a stock and are defined as the realized covariance of the stock with the market divided by the realized variance of the stock. To construct these series we use 5-minute returns obtained from the Thomson Reuters Tick History database. These returns are cleaned following the recommendations of Barndorff-Nielsen et al. (2009) to account for the typical high frequency data quality issues.

To give a first graphical impression, Figure 2 plots the realized betas and the corresponding autocorrelation function and periodogram for the two stocks we consider. It can be seen that the autocorrelation function and periodogram indicate the series to be highly persistent with significant positive correlation even after 200 lags and a pole at the origin.

Despite such evidence for persistence, realized betas have been sparsely investigated concerning their order of integration so far. A noteworthy study in this context is the one by Andersen et al. (2006), who find that quarterly betas in the time period 1969 until 1999 were best described by a process with $d \approx 0.2$. It should be noted, however, that due to the small number of observations their analysis is based on graphical investigation rather than consistent and robust estimation of the memory parameter.

The two constituents of realized beta, realized variance and realized covariance, on the other hand, have been investigated more extensively. Depending on the investigated stock and time period, it is found that realized variances can be best described by pure long memory processes (cf. e.g. Andersen et al. (2003) and Corsi (2009)) or a combination of long memory process and shift process (cf. e.g. Liu and Maheu (2007) and Choi et al. (2010)). For realized covariances, although less considered in the literature, a similar conclusion might be drawn (cf. e.g. Bertram et al. (2013) and Asai and McAleer (2015)).

To summarize, there is evidence that realized betas are fractionally integrated. Furthermore, since we investigate two companies who both mainly operate in the same industry, it seems reasonable to assume that they face the same relation to the systematic risk factor. This would imply a fractional cointegration relation between the two series. However, the series might also exhibit low frequency contaminations caused by structural breaks in the realized volatility or the realized covariance. Therefore, estimating the

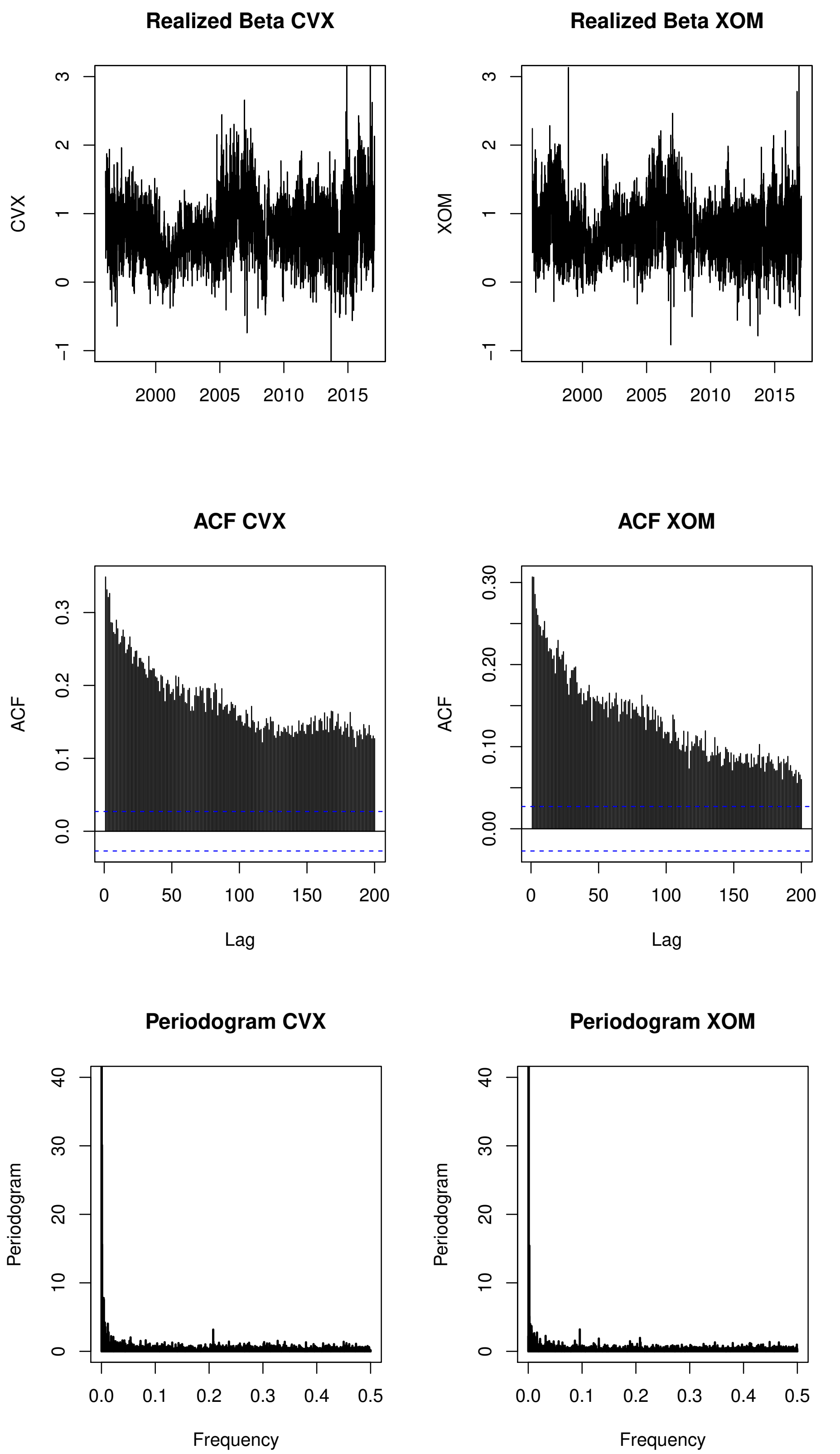

Figure 2: Top: Daily realized betas of Chevron and ExxonMobile relative to the *S&P*500 from January 1996 to February 2017. Middle: Corresponding autocorrelation functions excluding lag zero. Bottom: Corresponding periodograms.

	$\hat{T}_{RY}$	$\hat{T}_{TRE}$	$\widehat{rg}_{RY}$	$\widehat{rg}_{TRE}$	MLWS	GSE	TMLW	$\hat{\hat{d}}$	$\hat{\beta}$
Chevron ExxonMobile	0.21	0.40	0	1	3.84	0.346 0.375	0.180	0.059	0.491

Table 4: Multivariate estimation results. $\hat{T}_{RY}$ and $\widehat{rg}_{RY}$ correspond to the test for equality of d and the fractional cointegration rank estimator by Robinson and Yajima (2002) and $\hat{T}_{TRE}$ and $\widehat{rg}_{TRE}$ to our robust procedure. In analogy, GSE corresponds to the d estimate by the standard multivariate local Whittle estimator from Shimotsu (2007) and Robinson (2008b) and TMLW to the estimate by our trimmed estimator for which we also state the estimated reduction in memory $\hat{b}$ and the estimated cointegration vector $\hat{\beta}$. Finally, MLWS corresponds to the test statistic of the multivariate test for spurious long memory by Sibbertsen et al. (2018). For our procedures we employ the parameter combination used in Section 5. For the procedures by Robinson and Yajima (2002), Shimotsu (2007) and Robinson (2008b), and Sibbertsen et al. (2018) we consider the parameter combinations recommended by the authors.

memory of realized betas and thereby estimating the fractional cointegration relation should be done using our robust methods.
To demonstrate this, we test whether the two series exhibit equal order of integration and then estimate the fractional cointegration rank of the two series using the standard procedure by Robinson and Yajima (2002) and our robust procedure. As mentioned in Section 3, for the test we need to decide for a robust univarite estimator. Here, we consider the estimator by Iacone (2010), the results are, however, qualitatively similar when considering the estimators by McCloskey and Perron (2013) or Hou and Perron (2014). We are then able to compute the multivariate local Whittle estimator by Shimotsu (2007) and Robinson (2008b) and our robust multivariate local Whittle estimator. Additionally, we apply the multivariate MLWS test by Sibbertsen et al. (2018) to test for low frequency contaminations. For our methods we use the parameter combinations recommended in Section 5 and for the other methods we use the parameter combinations recommended by the authors of the procedures. The results are displayed in Table 4.
For estimating the fractional cointegration rank, we first need to test whether the series exhibit an equal order of integration. Otherwise, fractional cointegration can be excluded right away. The table shows that both, robust and standard test statistic are clearly below the five percent critical value of 1.96. Therefore, the null hypothesis that the series are equally integrated cannot be rejected meaning that it is sensible to investigate whether the two series are fractionally integrated. It can be seen that the procedure by Robinson and Yajima (2002) indicates the cointegration rank to be zero, i.e. no fractional cointegration while our robust procedure estimates it to be 1. Furthermore, the multivariate MLWS by Sibbertsen et al. (2018) shows a test statistic of 3.84 which is above the one percent critical value of 1.517 implying that the series exhibit low frequency contaminations. These then dominate the G matrix in the lower frequencies letting the matrix appear to have full rank which makes the standard estimator by

Robinson and Yajima (2002) unable to detect the cointegration relation. In contrast, our procedure trims those frequencies and is therefore robust to the contaminations.
If we then estimate the order of integration using the standard multivariate local Whittle estimator our error is twofold. First, we have a positive bias of the estimates caused by the low frequency contaminations. Second, we ignore the fact that the two series are fractionally cointegrated causing a bias as well. This reflects in the estimates made by the standard multivariate local Whittle estimator and our trimmed estimate stated in column four and five of Table 4. While the standard procedure estimates the memory parameters to be 0.35 respectively 0.38, our robust estimator yields a significantly lower value of 0.18 for both series which is in line with the considerations by Andersen et al. (2006) that realized betas have a d of approximately 0.2. The estimator further states the reduction in memory to be 0.06 and the fractional cointegration vector to be $(1, -0.49)'$.

7 Conclusion

In this paper we formally define the concept of spurious fractional cointegration, an extension to the well discussed problem of spurious long memory to the multivariate case. We show that this phenomenon can appear in practice leading to wrong inference. Spurious fractional cointegration is similar to spurious long memory caused by low frequency contaminations of the periodogram. We therefore in a first step investigate the behaviour of the pseudo-periodogram of a versatile class of time-varying mean processes. To handle spurious fractional cointegration we provide a method to estimate the rank of possible common low frequency contaminations and propose a trimmed version of the procedure of Robinson and Yajima (2002) to robustly estimate the correct fractional cointegration rank. Furthermore, a robust multivariate local Whittle estimator for the order of integration and cointegrating vector is proposed which is based on a multivariate extension of the trimmed univariate local Whittle estimator by Iacone (2010). In a Monte Carlo section the satisfying finite sample behavior of the procedures are displayed. An application to realized betas of two American stocks shows the usefulness of the methods.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Wu, G. (2006). Realized beta: Persistence and predictability. In *Econometric Analysis of Financial and Economic Time Series*, pages 1–39. Emerald Group Publishing Limited.

Asai, M. and McAleer, M. (2015). Forecasting co-volatilities via factor models with asymmetry and long memory in realized covariance. *Journal of Econometrics*, 189(2):251–262.

Barndorff-Nielsen, O. E., Hansen, P., and Shephard, N. (2009). Realised kernels in practice: Trades and quotes. *Econometrics Journal*, 12:1–33.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long memory processes: Probabilistic properties and statistical methods*. Springer London, Limited.

Bertram, P., Kruse, R., and Sibbertsen, P. (2013). Fractional integration versus level shifts: the case of realized asset correlations. *Statistical Papers*, 54(4):977–991.

Bhattacharya, R. N., Gupta, V. K., and Waymire, E. (1983). The hurst effect under trends. *Journal of Applied Probability*, 20(3):649–662.

Brown, B. (1974). Generalized wald equations in discrete time. *Stochastic Processes and their Applications*, 2(4):349–357.

Chen, W. W. and Hurvich, C. M. (2006). Semiparametric estimation of fractional cointegrating subspaces. *The Annals of Statistics*, 34(6):2939–2979.

Choi, K., Yu, W.-C., and Zivot, E. (2010). Long memory versus structural breaks in modeling and forecasting realized volatility. *Journal of International Money and Finance*, 29(5):857–875.

Christensen, B. J. and Santucci de Magistris, P. (2010). Level shifts in volatility and the implied-realized volatility relation. *Available at SSRN 1692844*.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131–159.

Engle, R. F. and Smith, A. D. (1999). Stochastic permanent breaks. *Review of Economics and Statistics*, 81(4):553–574.

Granger, C. W. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, 73(1):61–77.

Granger, C. W. and 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(3):399–421.

Guégan, D. (2005). How can we define the concept of long memory? An econometric survey. *Econometric Reviews*, 24(2):113–149.

Hendry, D. F. and Massmann, M. (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business & Economic Statistics*, 25(1):33–51.

Hou, J. and Perron, P. (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics*, 182(2):309–328.

Iacone, F. (2010). Local whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis*, 31(1):37–49.

Kellard, N. M., Jiang, Y., and Wohar, M. (2015). Spurious long memory, uncommon breaks and the implied–realized volatility puzzle. *Journal of International Money and Finance*, 56:36–54.

Künsch, H. (1986). Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability*, 23(4):1025–1030.

Liu, C. and Maheu, J. M. (2007). Are there structural breaks in realized volatility? *Journal of Financial Econometrics*, 6(3):326–360.

Lobato, I. N. (1999). A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics*, 90(1):129–153.

Marinucci, D. and Robinson, P. M. (2001). Semiparametric fractional cointegration analysis. *Journal of Econometrics*, 105(1):225–247.

McCloskey, A. and Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory*, 29(6):1196–1237.

McGown, K. J. and Parks, H. R. (2007). The generalization of faulhaber's formula to sums of non-integral powers. *Journal of Mathematical Analysis and Applications*, 330(1):571–575.

Nielsen, M. Ø. (2007). Local Whittle analysis of stationary fractional cointegration and the implied–realized volatility relation. *Journal of Business & Economic Statistics*, 25(4):427–446.

Nielsen, M. Ø. and Shimotsu, K. (2007). Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. *Journal of Econometrics*, 141(2):574–596.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics*, 29(3):423–438.

Robinson, P. M. (2008a). Diagnostic testing for cointegration. *Journal of Econometrics*, 143(1):206–225.

Robinson, P. M. (2008b). Multiple local Whittle estimation in stationary systems. *The Annals of Statistics*, 36(5):2508–2530.

Robinson, P. M. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems. *Journal of Econometrics*, 106(2):217–241.

Shimotsu, K. (2007). Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics*, 137(2):277–310.

Shimotsu, K. (2010). Supplement to exact local whittle estimation of fractionally cointegrated systems.

Shimotsu, K. (2012). Exact local whittle estimation of fractionally cointegrated systems. *Journal of Econometrics*, 169(2):266–278.

Sibbertsen, P., Leschinski, C., and Busch, M. (2018). A multivariate test against spurious long memory. *Journal of Econometrics*, 203(01):33–49.

Souza, I. V. M., Reisen, V. A., Franco, G. d. C., and Bondon, P. (2018). The estimation and testing of the cointegration order based on the frequency domain. *Journal of Business & Economic Statistics*, 36(4):695–704.

Appendix

Before proving Lemmas 1, 2 and 3 we need two auxiliary lemmas. For the structural-change processes in (1) and (2) we have the following result.

Lemma 4. *The discrete Fourier transform (DFT) of the process in (1) can be represented as*

$$w_\mu(\lambda_j) = -\frac{1}{\sqrt{2\pi T}} \sum_{k=1}^{K} \Delta\mu_k D_{T_k}(\lambda_j),$$

and that of the process in (2) can be represented as

$$w_\mu(\lambda_j) = -\frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{K} \mu_k \left(D_{T_k-1}(\lambda_j) - D_{T_{k-1}}(\lambda_j) \right),$$

where $D_{T_k}(\lambda_j) = \sum_{t=1}^{T_k} e^{i\lambda_j t}$ is a version of the Dirichlet kernel.

Note that Lemma 4 is completely algebraic and we do not impose any conditions on the $\Delta\mu_k$, μ_k, or T_k.

Proof of Lemma 4:

From (1), we have

$$w_\mu(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \left\{ \mu_0 + \sum_{k=1}^{K} 1(t \geq T_k)\Delta\mu_k \right\} e^{i\lambda_j t} = \frac{1}{\sqrt{2\pi T}} \left\{ \mu_0 \sum_{t=1}^{T} e^{i\lambda_j t} + \sum_{k=1}^{K} \Delta\mu_k \sum_{t=1}^{T} 1(t \geq T_k) e^{i\lambda_j t} \right\}.$$

Here,

$$\sum_{t=1}^{T} 1(t \geq T_k) e^{i\lambda_j t} = \sum_{t=T_k}^{T} e^{i\lambda_j t} = \sum_{t=1}^{T} e^{i\lambda_j t} - \sum_{t=1}^{T_k} e^{i\lambda_j t} = D_T(\lambda_j) - D_{T_k}(\lambda_j).$$

Therefore,

$$w_\mu(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \left\{ D_T(\lambda_j)\mu_0 + \sum_{k=1}^{K} \Delta\mu_k \left[D_T(\lambda_j) - D_{T_k}(\lambda_j) \right] \right\}$$

$$= \frac{1}{\sqrt{2\pi T}} \left\{ \left[\mu_0 + \sum_{k=1}^{K} \Delta\mu_k \right] D_T(\lambda_j) - \sum_{k=1}^{K} \Delta\mu_k D_{T_k}(\lambda_j) \right\}.$$

Furthermore, we have

$$D_T(\lambda_j) = \frac{e^{i(T+1)\lambda_j} - e^{i\lambda_j}}{e^{i\lambda_j} - 1} = e^{i(T-1)\lambda_j/2} \frac{\sin(T\lambda_j/2)}{\sin(\lambda_j/2)}, \tag{10}$$

cf. Beran et al. (2013), p. 327. Note that $\lambda_j T = 2\pi j$, $e^{i(T+1)\lambda_j} = e^{i\lambda_j T} e^{i\lambda_j}$, and $e^{i2\pi j} = \cos(2\pi j) + i\sin(2\pi j) = \cos(2\pi) + i\sin(2\pi) = 1$. Therefore, $D_T(\lambda_j) = \frac{e^{i\lambda_j} - e^{i\lambda_j}}{e^{i\lambda_j} - 1} = 0$, which proves the first part of the lemma.

Similarly, for the second part of the lemma, we have from (2) that

$$\begin{aligned} w_\mu(\lambda_j) &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \left\{ \mu_0 + \sum_{k=0}^{K} \mu_k 1(T_{k-1} \leq t < T_k) \right\} e^{i\lambda_j t} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ \mu_0 \sum_{t=1}^{T} e^{i\lambda_j t} + \sum_{k=0}^{K} \mu_k \sum_{t=1}^{T} 1(T_{k-1} \leq t < T_k) e^{i\lambda_j t} \right\}. \end{aligned}$$

Here,

$$\begin{aligned} &\sum_{t=1}^{T} 1(T_{k-1} \leq t < T_k) e^{i\lambda_j t} = \sum_{t=1}^{T} \{1(T_k > t) - 1(T_{k-1} \geq t)\} e^{i\lambda_j t} \\ &= \sum_{t=1}^{T_k - 1} e^{i\lambda_j t} - \sum_{t=1}^{T_{k-1}} e^{i\lambda_j t} = D_{T_k - 1}(\lambda_j) - D_{T_{k-1}}(\lambda_j). \end{aligned}$$

Therefore, since $D_T(\lambda_j) = 0$, we have

$$w_\mu(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{K} \mu_k \left\{ D_{T_k - 1}(\lambda_j) - D_{T_{k-1}}(\lambda_j) \right\}.$$

□

Since Lemma 4 implies that the properties of the DFT and thus the properties of the periodogram of a structural-change process are directly related to those of the Dirichlet kernel, the following lemma provides an approximation for the Dirichlet kernel at frequencies local to zero.

Lemma 5. *We have for $T_k/T = \delta_k \in (0,1)$ and $j/T \to 0$,*

$$\begin{aligned} D_{T_k}(\lambda_j) &= T j^{-1} \frac{\sin(2\delta_k \pi j)}{2\pi} + \sin^2(\pi j \delta_k) \\ &\quad + i \left[T j^{-1} \frac{\sin^2(\pi \delta_k j)}{\pi} - \frac{1}{2} \sin(2\pi \delta_k j) \right] + O_P(jT^{-1}). \end{aligned}$$

Clearly, from Lemma 5, both the real and the imaginary part of the Dirichlet kernel are $O(T j^{-1})$ for deterministic δ_k and $O_p(T j^{-1})$ if any of the δ_k are stochastic. Furthermore, the order is exact. Again, this is an approximation based on a Laurent expansion that holds irrespective of the stochastic properties of the T_k.

Proof of Lemma 5:

From the second expression in (10) in the proof of Lemma 4 we can decompose the real and the imaginary parts of the DFT at the Fourier frequencies $\lambda_j = 2\pi j/T$ as follows

$$\begin{aligned} D_{T_k}(\lambda_j) &= \frac{e^{i(T_k - 1)\lambda_j/2} \sin(T_k \lambda_j/2)}{\sin(\lambda_j/2)} \\ &= \frac{[\cos((T_k - 1)\lambda_j/2) + i \sin((T_k - 1)\lambda_j/2)] \sin(T_k \lambda_j/2)}{\sin(\lambda_j/2)} \end{aligned}$$

It follows by the sum-to-product identities that

$$
\begin{aligned}
D_{T_k}(\lambda_j) &= \frac{\sin(\frac{T_k-1}{T}\pi j+\frac{T_k}{T}\pi j)-\sin(\frac{T_k-1}{T}\pi j-\frac{T_k}{T}\pi j)}{2\sin(\frac{\pi j}{T})} \\
&\quad + i\frac{\cos(\frac{T_k-1}{T}\pi j-\frac{T_k}{T}\pi j)-\cos(\frac{T_k-1}{T}\pi j+\frac{T_k}{T}\pi j)}{2\sin(\frac{\pi j}{T})} \\
&= \frac{\sin(2\delta_k\pi j-\frac{\pi j}{T})+\sin(\frac{\pi j}{T})}{2\sin(\frac{\pi j}{T})} + i\frac{\cos(\frac{\pi j}{T})-\cos(2\delta_k\pi j-\frac{\pi j}{T})}{2\sin(\frac{\pi j}{T})}.
\end{aligned}
$$

By a Laurent series approximation around $\lambda_j = 0$, we obtain

$$
\begin{aligned}
D_{T_k}(\lambda_j) &= Tj^{-1}\frac{\sin(2\delta_k\pi j)}{2\pi} + \sin^2(\pi j\delta_k) + O_P(jT^{-1}) \\
&\quad + i\left[Tj^{-1}\frac{\sin^2(\pi\delta_k j)}{\pi} - \frac{1}{2}\sin(2\pi\delta_k j) + O_P(jT^{-1})\right],
\end{aligned}
$$

where the Laurent series is obtained from separate Taylor approximations for each of the trigonometric functions. This proves the lemma. □

We can now prove the Lemmas 1, 2 and 3.

Proof of Lemma 1:

For $\mu_t = h(t/T,T)$, by combining Lemma 4 with Lemma 5, we have

$$
\begin{aligned}
I_\mu(\lambda_j) &= \left|-\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\Delta\mu_t D_t(\lambda_j)\right|^2 \\
&= (2\pi T)^{-1}\left\{\left[T\sum_{t=1}^{T}\frac{\Delta\mu_t\sin(2\pi jt/T)}{2\pi j} + \sum_{t=1}^{T}\Delta\mu_t\sin^2(\pi jt/T) + T^{-1}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right. \\
&\quad \left. + \left[T\sum_{t=1}^{T}\Delta\mu_t\frac{\sin^2(\pi jt/T)}{\pi j} - 1/2\sum_{t=1}^{T}\Delta\mu_t\sin(2\pi jt/T) + T^{-1}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right\}
\end{aligned}
$$

Factoring out T from the square brackets gives

$$
\begin{aligned}
2\pi I_\mu(\lambda_j)T^{-1} &= \left\{\left[\sum_{t=1}^{T}\frac{\Delta\mu_t\sin(2\pi jt/T)}{2\pi j} + T^{-1}\sum_{t=1}^{T}\Delta\mu_t\sin^2(\pi jt/T) + T^{-2}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right. \\
&\quad \left. + \left[\sum_{t=1}^{T}\Delta\mu_t\frac{\sin^2(\pi jt/T)}{\pi j} - \frac{1}{2T}\sum_{t=1}^{T}\Delta\mu_t\sin(2\pi jt/T) + T^{-2}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right\}.
\end{aligned}
$$

Now, using $\Delta\mu_t = h(t/T,T) - h((t-1)/T,T)$, we have

$$
\lim_{T\to\infty}\Delta\mu_t T = \lim_{T\to\infty}\frac{\partial h(t/T,T)}{\partial(t/T)},
$$

so that

$$
\begin{aligned}
&2\pi I_\mu(\lambda_j)T^{-1}\\
&\sim \Bigg\{ \Bigg[\frac{1}{2\pi jT}\sum_{t=1}^{T}\frac{\partial h(t/T,T)}{\partial(t/T)}\sin(2\pi jt/T)+\sum_{t=1}^{T}\frac{\partial h(t/T,T)}{\partial(t/T)}\frac{1}{T^2}\sin^2(\pi jt/T)+\frac{1}{T^3}\sum_{t=1}^{T}\frac{\partial h(t/T,T)}{\partial(t/T)}O_P(j)\Bigg]^2\\
&+\Bigg[\sum_{t=1}^{T}\frac{\partial h(t/T,T)}{\partial(t/T)}\frac{\sin^2(\pi jt/T)}{\pi jT}-\frac{1}{2}\sum_{t=1}^{T}\frac{\partial h(t/T,T)}{\partial(t/T)}\frac{1}{T^2}\sin(2\pi jt/T)+\frac{1}{T^3}\sum_{t=1}^{T}\frac{\partial h(t/T,T)}{\partial(t/T)}O_P(j)\Bigg]^2\Bigg\},
\end{aligned}
$$

where $a \sim b$ means that the ratio of a and b converge to 1, as $T \to \infty$. Here and in the following this shorthand notation is used to improve the readability of the proof.
By the definition of a Riemann integral

$$
\begin{aligned}
&2\pi I_\mu(\lambda_j)T^{-1}\\
&\sim \Bigg\{\Bigg[\frac{1}{2\pi j}\int_0^1\frac{\partial h(s,T)}{\partial s}\sin(2\pi js)ds+\frac{1}{T}\int_0^1\frac{\partial h(s,T)}{\partial s}\sin^2(\pi js)ds+T^{-2}\int_0^1\frac{\partial h(s,T)}{\partial s}O_P(j)ds\Bigg]^2\\
&+\Bigg[\int_0^1\frac{\partial h(s,T)}{\partial s}\frac{\sin^2(\pi js)}{\pi j}ds-\frac{1}{2T}\int_0^1\frac{\partial h(s,T)}{\partial s}\sin(2\pi js)ds+T^{-2}\int_0^1\frac{\partial h(s,T)}{\partial s}O_P(j)ds\Bigg]^2\Bigg\}.
\end{aligned}
$$

Clearly, both parts of this expression are dominated by the first term in the respective square bracket, such that

$$
I_\mu(\lambda_j) \sim \frac{T}{8\pi^3 j^2}\left\{\left[\int_0^1\frac{\partial h(s,T)}{\partial s}\sin(2\pi js)ds\right]^2+\left[\int_0^1\frac{\partial h(s,T)}{\partial s}(1-\cos(2\pi js))ds\right]^2\right\},
$$

which finishes our proof. □

Proof of Lemma 2:

First, by (10) in the proof of Lemma 4 we have

$$
D_{T_k}(\lambda_j)D^*_{T_u}(\lambda_j)=e^{i(T_k-T_u)\lambda_j/2}\frac{\sin(T_k\lambda_j/2)\sin(T_u\lambda_j/2)}{\sin^2(\lambda_j/2)}=2e^{i\pi j(\delta_k-\delta_u)}\frac{\sin(\delta_k\pi j)\sin(\delta_u\pi j)}{1-\cos(\lambda_j)}.
$$

By a Laurent expansion around $\lambda_j = 0$, we have

$$
\begin{aligned}
D_{T_k}(\lambda_j)D^*_{T_u}(\lambda_j) &= 2e^{i\pi j(\delta_k-\delta_u)}\frac{\sin(\delta_k\pi j)\sin(\delta_u\pi j)}{1-[1-2\pi^2(j/T)^2+O((j/T)^4)]}\\
&=\frac{T^2}{\pi^2 j^2}e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j)+O_P(1).
\end{aligned}
\tag{11}
$$

In particular, for the case when $T_k = T_u = t$, we obtain

$$
D_t(\lambda_j)D^*_t(\lambda_j)=\frac{T^2}{2\pi^2 j^2}(1-\cos(2\pi tj/T))+O(1).
\tag{12}
$$

Furthermore, we have

$$
\sum_{k=1}^{K}\Delta\mu_k D_{T_k}(\lambda_j)=\sum_{t=1}^{T}\Delta\mu_t D_t(\lambda_j), \quad \text{where} \quad \Delta\mu_t=\begin{cases}\Delta\mu_k, & \text{if } t=T_k\\ 0, & \text{otherwise.}\end{cases}
$$

In addition to that,

$$I_\mu(\lambda_j) = A + B = \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}\Delta\mu_t\Delta\mu_s D_t(\lambda_j)D_s^*(\lambda_j)$$
$$= \frac{1}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2 D_t(\lambda_j)D_t^*(\lambda_j) + \frac{1}{2\pi T}\sum_{t\neq s}\Delta\mu_t\Delta\mu_s D_t(\lambda_j)D_s^*(\lambda_j).$$

Consequently, we have for term A and from (12) above

$$A = \frac{1}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2 D_t(\lambda_j)D_t^*(\lambda_j) = \frac{T}{4\pi^3 j^2}\sum_{t=1}^{T}(\Delta\mu_t)^2(1-\cos(2\pi jt/T)) + \frac{O(1)}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2$$
$$= \frac{T}{4\pi^3 j^2}\left\{\sum_{t=1}^{T}(\Delta\mu_t)^2 - \sum_{t=1}^{T}(\Delta\mu_t)^2\cos(2\pi jt/T)\right\} + O_P(1). \tag{13}$$
$$= \frac{T}{4\pi^3 j^2}\left\{\sum_{k=1}^{K}(\Delta\mu_k)^2 - \sum_{k=1}^{K}(\Delta\mu_k)^2\cos(2\pi j\delta_k)\right\} + O_P(1). \tag{14}$$

To deal with term B, we revert back to the original representation in which the sum is random and write

$$B = \frac{1}{2\pi T}\sum_{t\neq s}\Delta\mu_t\Delta\mu_s D_t(\lambda_j)D_s^*(\lambda_j) = \frac{1}{2\pi T}\sum_{k\neq u}\Delta\mu_k\Delta\mu_u D_{T_k}(\lambda_j)D_{T_u}^*(\lambda_j),$$

where $k,l = 1,\ldots,K$. Similar to the approach above, we have from (11)

$$B = \frac{T}{2\pi^3 j^2}\sum_{k\neq u}\Delta\mu_k\Delta\mu_u e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j) + \frac{1}{2\pi T}\sum_{k\neq u}\Delta\mu_k\Delta\mu_u O_P(1) \tag{15}$$

The first part of the lemma i.) follows immediately from (14) and (15).

For the second part of the lemma, from Assumption A4 we have $E[K] = E[p_t T] = \tilde{p}T^{1-\alpha}$ and, from (13) and (14)

$$E[A] = \frac{T}{4\pi^3 j^2}\left\{E\left[\sum_{k=1}^{K}(\Delta\mu_k)^2\right] - E[\Delta\mu_t^2]\sum_{t=1}^{T}\cos(2\pi jt/T)\right\} + O(1)$$
$$= \frac{T}{4\pi^3 j^2}E\left[\sum_{k=1}^{K}(\Delta\mu_k)^2\right] + O(1),$$

since $\sum_{t=1}^{T}\cos(2\pi jt/T) = 0$.
Now, from Assumption A3 we have $E[(\Delta\mu_k)^2] = \sigma_\Delta^2 T^{-\beta}$, so that by the generalized Wald identity of Brown (1974) and Assumption A4

$$E[A] = \frac{T}{4\pi^3 j^2}E[K]E[(\Delta\mu_k)^2] = \frac{\tilde{p}\sigma_\Delta^2 T^{2-\alpha-\beta}}{4\pi^3 j^2} + o(1).$$

Similarly, from (15)

$$E[B] = \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq u}\Delta\mu_k\Delta\mu_u e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j)\right] + \frac{O(1)}{2\pi T}E\left[\sum_{k\neq u}\Delta\mu_k\Delta\mu_u\right]$$

and by the generalized Wald identity of Brown (1974) in conjunction with Assumption A5

$$E[B]=\frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq u}E\left[\Delta\mu_k\Delta\mu_u\right]E\left[e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j)\right]\right]+\frac{O(1)}{2\pi T}E\left[\sum_{k\neq u}E\left[\Delta\mu_k\Delta\mu_u\right]\right].$$

Therefore,

$$\begin{aligned}|E[B]|&\leq\frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq u}\left|E\left[\Delta\mu_k\Delta\mu_u\right]\right|\left|E\left[e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j)\right]\right|\right]+\left|\frac{O(1)}{2\pi T}E\left[\sum_{k\neq u}E\left[\Delta\mu_k\Delta\mu_u\right]\right]\right|\\&\leq\frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq u}\left|E\left[\Delta\mu_k\Delta\mu_u\right]\right|\right]+\frac{|O(1)|}{2\pi T}E\left[\sum_{k\neq u}\left|E\left[\Delta\mu_k\Delta\mu_u\right]\right]\right|.\end{aligned}$$

Assumption A5 combined with Assumptions A3 and A4 implies that

$$E\left[\sum_{k\neq u}|E[\Delta\mu_k\Delta\mu_u]|\right]=2E\left[\sum_{k=2}^{K}\sum_{\tau=1}^{k-1}|E[\Delta\mu_k\Delta\mu_{k-\tau}]|\right]\leq 2E\left[K\right]Var(\Delta\mu_k)\tilde{C}=2\tilde{p}\tilde{C}\sigma_\Delta^2T^{1-\alpha-\beta},$$

so that $|E[B]|\leq\frac{\tilde{p}\sigma_\Delta^2\tilde{C}}{\pi^3 j^2}T^{2-\alpha-\beta}+\left|O(T^{-\alpha-\beta})\right|$. □

Proof of Lemma 3:

From $w_\mu(\lambda_j)=-\frac{1}{\sqrt{2\pi T}}\sum_{k=0}^{K}\mu_k[D_{T_k-1}(\lambda_j)-D_{T_{k-1}}(\lambda_j)]$, as shown in Lemma 4, we have

$$\begin{aligned}&I_\mu(\lambda_j)=\tilde{A}+\tilde{B} &(16)\\&=\frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2[D_{T_k-1}(\lambda_j)D^*_{T_k-1}(\lambda_j)-D_{T_k-1}(\lambda_j)D^*_{T_{k-1}}(\lambda_j)-D_{T_{k-1}}(\lambda_j)D^*_{T_k-1}(\lambda_j)+D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda_j)]\\&+\frac{1}{2\pi T}\sum_{k\neq u}\mu_k\mu_u[D_{T_k-1}(\lambda_j)D^*_{T_u-1}(\lambda_j)-D_{T_k-1}(\lambda_j)D^*_{T_{u-1}}(\lambda_j)-D_{T_{k-1}}(\lambda_j)D^*_{T_u-1}(\lambda_j)+D_{T_{k-1}}(\lambda_j)D^*_{T_{u-1}}(\lambda_j)].\end{aligned}$$

Denoting $(T_k-1)/T=\tilde{\delta}_k$, we have from (11) for the term in square brackets in $\tilde{A}$

$$\begin{aligned}\tilde{a}_k=&[D_{T_k-1}(\lambda_j)D^*_{T_k-1}(\lambda_j)-D_{T_k-1}(\lambda_j)D^*_{T_{k-1}}(\lambda_j)-D_{T_{k-1}}(\lambda_j)D^*_{T_k-1}(\lambda_j)+D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda)]\\=&\frac{T^2}{\pi^2 j^2}\Big[\sin^2(\tilde{\delta}_k\pi j)+\sin^2(\delta_{k-1}\pi j)-e^{i\pi j(\tilde{\delta}_k-\delta_{k-1})}\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\\&-e^{i\pi j(\delta_{k-1}-\tilde{\delta}_k)}\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\Big]+O_P(1)\\=&\frac{T^2}{\pi^2 j^2}[1-\frac{1}{2}\left[\cos(2\tilde{\delta}_k\pi j)+\cos(2\delta_{k-1}\pi j)\right]-2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\cos(\pi j(\tilde{\delta}_k-\delta_{k-1}))]+O_P(1),\end{aligned}$$

from Euler's formula. As in the proof of Lemma 1, the notation $a\sim b$ is used as a shorthand for $\lim_{T\to\infty}a/b=1$. By the sum-to-product identity for the cosine, it follows

$$\begin{aligned}\tilde{a}_k=&\frac{T^2}{\pi^2 j^2}\Big[1-\frac{1}{2}\left[2\cos(\pi j(\tilde{\delta}_k+\delta_{k-1}))\cos(\pi j(\tilde{\delta}_k-\delta_{k-1}))\right]\\&-2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\cos(\pi j(\tilde{\delta}_k-\delta_{k-1}))\Big]+O_P(1)\\=&\frac{T^2}{\pi^2 j^2}\left[1-\cos(\pi j(\tilde{\delta}_k-\delta_{k-1}))\left[\cos(\pi j(\tilde{\delta}_k+\delta_{k-1}))+2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\right]\right]+O_P(1).\end{aligned}$$

Now, by the product-to-sum identity of the sine

$$\tilde{a}_k = \frac{T^2}{\pi^2 j^2}\Big[1-\cos(\pi j(\tilde{\delta}_k-\delta_{k-1}))\Big[\cos(\pi j(\tilde{\delta}_k+\delta_{k-1}))$$
$$+\cos(\pi j(\tilde{\delta}_k-\delta_{k-1}))-\cos(\pi j(\tilde{\delta}_k+\delta_{k-1}))\Big]\Big]+O_P(1)$$
$$=\frac{T^2}{\pi^2 j^2}\Big[1-\cos^2(\pi j(\tilde{\delta}_k-\delta_{k-1}))\Big]+O_P(1).$$

Therefore, we have

$$\tilde{A}=\frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2\tilde{a}_k=\frac{T}{2\pi^3 j^2}\sum_{k=0}^{K}\mu_k^2[1-\cos^2(\pi j(\tilde{\delta}_k-\delta_{k-1}))]+\frac{(K+1)O_P(1)}{2\pi T} \tag{17}$$

For $\alpha<1$, by a Taylor expansion of the squared cosine at zero

$$\tilde{a}_k=\frac{T^2}{\pi^2 j^2}\Big[1-\Big[1-\pi^2 j^2(\tilde{\delta}_k-\delta_{k-1})^2+\pi^4 j^4 O((\tilde{\delta}_k-\delta_{k-1})^4)\Big]\Big]+O_P(1)$$
$$=T^2\Big[(\tilde{\delta}_k-\delta_{k-1})^2-\pi^2 j^2 O((\tilde{\delta}_k-\delta_{k-1})^4)\Big]+O_P(1).$$

Therefore, we obtain

$$\tilde{A}=\frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2\Big\{T^2\Big[(\tilde{\delta}_k-\delta_{k-1})^2-\pi^2 j^2 O((\tilde{\delta}_k-\delta_{k-1})^4)\Big]+O_P(1)\Big\}$$

$$=\frac{T}{2\pi}\sum_{k=0}^{K}\Big\{\mu_k^2(\tilde{\delta}_k-\delta_{k-1})^2\Big\}-\frac{T\pi j^2}{4}\sum_{k=0}^{K}\mu_k^2 O\Big((\tilde{\delta}_k-\delta_{k-1})^4\Big)+\frac{O_P(1)}{2\pi T}\sum_{k=0}^{K}\mu_k^2. \tag{18}$$

By applying the Wald identity for dependent random sums of Brown (1974) and then using Assumption A5, we obtain

$$E[\tilde{A}]=E[K+1]E[\mu_k^2]\Big\{\frac{T}{2\pi}E[(\tilde{\delta}_k-\delta_{k-1})^2]-\frac{T\pi j^2}{4}E[O((\tilde{\delta}_k-\delta_{k-1})^4)]+\frac{O(1)}{2\pi T}\Big\}$$
$$=(\tilde{p}T^{1-\alpha}+1)\sigma_\Delta^2 T^{-\beta}\Big\{\frac{T}{\pi}\frac{\tilde{D}}{\tilde{p}^2}T^{2(\alpha-1)}-\frac{T\pi j^2}{2}O(T^{4(\alpha-1)})+O(T^{-1})\Big\}$$
$$=\frac{\sigma_\Delta^2\tilde{D}}{\pi\tilde{p}}T^{\alpha-\beta}-\frac{\sigma_\Delta^2\tilde{p}\pi j^2}{2}O(T^{-2+3\alpha-\beta})+\sigma_\Delta^2\tilde{p}O(T^{-\alpha-\beta})$$
$$+\frac{\sigma_\Delta^2\tilde{D}}{\pi\tilde{p}}T^{-1+2\alpha-\beta}-\pi j^2 O(T^{-3+4\alpha-\beta})+O(T^{-1-\beta}). \tag{19}$$

For $\tilde{B}$, we have from (11),

$$\tilde{B}=\frac{T}{2\pi^2 j^2}\sum_{k\neq u}\mu_k\mu_u\Big[e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j)-e^{i\pi j(\delta_k-\delta_{u-1})}\sin(\delta_k\pi j)\sin(\delta_{u-1}\pi j)$$
$$-e^{i\pi j(\delta_{k-1}-\delta_u)}\sin(\delta_{k-1}\pi j)\sin(\delta_u\pi j)+e^{i\pi j(\delta_{k-1}-\delta_{u-1})}\sin(\delta_{k-1}\pi j)\sin(\delta_{u-1}\pi j)\Big]$$
$$+\frac{O_P(1)}{2\pi T}\sum_{k\neq u}\mu_k\mu_u. \tag{20}$$

Denote the term in the square bracket by $\tilde{b}$, and let $\tilde{b}_1$ denote the first two summands and $\tilde{b}_2$

the last two summands, so that $\tilde{b} = \tilde{b}_1 + \tilde{b}_2$. We have

$$\begin{aligned}\tilde{b}_1 &= \sin(\delta_k \pi j)\left[e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_u \pi j) - e^{i\pi j(\delta_k-\delta_{u-1})}\sin(\delta_{u-1}\pi j)\right]\\ &= \sin(\delta_k \pi j)\left[\cos(\pi j(\delta_k-\delta_u))\sin(\delta_u \pi j) - \cos(\pi j(\delta_k-\delta_{u-1}))\sin(\delta_{u-1}\pi j)\right.\\ &\left. + i\left\{\sin(\pi j(\delta_k-\delta_u))\sin(\pi j \delta_u) - \sin(\pi j(\delta_k-\delta_{u-1}))\sin(\pi j \delta_{u-1})\right\}\right].\end{aligned}$$

Now, let $\gamma_u = \delta_u - \delta_{u-1}$. Then, by a Taylor approximation at $\gamma_u = 0$

$$\tilde{b}_1 = \pi j \gamma_u \sin(\delta_k \pi j) e^{i\pi j(\delta_k - 2\delta_u)} + O_P(\gamma_u^2).$$

Similarly, we have for the third and fourth term in the square bracket

$$\tilde{b}_2 = -\sin(\delta_{k-1}\pi j)\left[e^{i\pi j(\delta_{k-1}-\delta_u)}\sin(\delta_u \pi j) - e^{i\pi j(\delta_{k-1}-\delta_{u-1})}\sin(\delta_{u-1}\pi j)\right],$$

and by a Taylor approximation at $\gamma_u = 0$,

$$\tilde{b}_2 = -\pi j \gamma_u \sin(\delta_{k-1}\pi j) e^{i\pi j(\delta_{k-1} - 2\delta_u)} + O_P(\gamma_u^2).$$

Therefore, we have

$$\tilde{b} = -\pi j \gamma_u \left[\sin(\delta_k \pi j) e^{i\pi j(\delta_k - 2\delta_u)} - \sin(\delta_{k-1}\pi j) e^{i\pi j(\delta_{k-1}-2\delta_u)}\right].$$

Defining $\gamma_k = \delta_k - \delta_{k-1}$, and approximating at $\gamma_k = 0$, we obtain

$$\tilde{b} = \pi^2 j^2 \gamma_u \gamma_k e^{2i\pi j(\delta_k - \delta_u)} + O_P(\gamma_u^2) + O_P(\gamma_k^2),$$

so that

$$\begin{aligned}\tilde{B} &= \frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_u\left[\pi^2 j^2 \gamma_u\gamma_k e^{2i\pi j(\delta_k-\delta_u)} + O_P(\gamma_u^2) + O_P(\gamma_k^2)\right] + \frac{O_P(1)}{2\pi T}\sum_{k\neq u}\mu_k\mu_u\\ &= \frac{T}{2}\sum_{k\neq u}\mu_k\mu_u\gamma_u\gamma_k e^{2i\pi j(\delta_k-\delta_u)} + \frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_u O_P(\gamma_u^2) + \frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_u O_P(\gamma_k^2) + \frac{O_P(1)}{2\pi T}\sum_{k\neq u}\mu_k\mu_u. \quad (21)\end{aligned}$$

Similar to the proof of Theorem 2, we have from the Wald identity of Brown (1974) and Assumption A5

$$\begin{aligned}&E[\tilde{B}] = \tilde{B}_1 + \tilde{B}_2 + \tilde{B}_3\\ &= \frac{T}{2}E\left[\sum_{k\neq l}E\left[\mu_k\mu_u\right]E\left[\gamma_u\gamma_k e^{2i\pi j(\delta_k-\delta_u)}\right]\right] + \frac{T}{2\pi^2 j^2}E\left[\sum_{k\neq l}E\left[\mu_k\mu_u\right]E[O_P(\gamma_k^2)]\right] + \frac{O(1)}{2\pi T}E\left[\sum_{k\neq u}E[\mu_k\mu_u]\right].\end{aligned}$$

For the first term

$$\begin{aligned}\left|E[\tilde{B}_1]\right| \le \frac{T}{2}E\left[\sum_{k\neq l}\left|E\left[\mu_k\mu_u\right]\right|\left|E\left[\gamma_u\gamma_k\right]\right|\right] &= TE\left[\sum_{k=1}^{K}\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\left|E\left[\gamma_k\gamma_{k-\tau}\right]\right|\right]\\ &\le TE\left[\sum_{k=1}^{K}\left|E\left[\gamma_k^2\right]\right|\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\right].\end{aligned}$$

Therefore, by Assumptions A4 and A5

$$
\begin{aligned}
\left|E[\tilde{B}_1]\right| &\leq TE\left[\sum_{k=1}^{K}\left|E\left[\gamma_k^2\right]\right|Var(\mu_k)\tilde{C}\right] = TE[K]E\left[\gamma_k^2\right]Var(\mu_k)\tilde{C} \\
&= \frac{2\tilde{C}\tilde{D}\sigma_\Delta^2}{\tilde{p}}T^{\alpha-\beta} + \tilde{C}\left[O(T^{2\alpha-1-\beta}) + O(T^{-\beta}) + O(T^{\alpha-\beta-1})\right]. \qquad (22)
\end{aligned}
$$

Similarly, for the second term

$$
\left|E[\tilde{B}_2]\right| \leq \frac{2T}{\pi^2 j^2}E\left[\sum_{k=1}^{K}\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]E\left[O_P(\gamma_k^2)\right]\right|\right]
$$

$$
\begin{aligned}
&\leq \frac{2T}{\pi^2 j^2}E\left[\sum_{k=1}^{K}E\left[O_P(\gamma_k^2)\right]\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\right] \leq \frac{2\tilde{C}\sigma_\Delta^2}{\pi^2 j^2}T^{1-\beta}E\left[\sum_{k=1}^{K}O_P(\gamma_k^2)\right] \\
&= \frac{2\tilde{C}\sigma_\Delta^2\tilde{p}}{\pi^2 j^2}T^{2-\alpha-\beta}\left[O(T^{2(\alpha-1)}) + O(T^{\alpha-2})\right] = \frac{2\tilde{C}\sigma_\Delta^2\tilde{p}}{\pi^2 j^2}\left[O(T^{\alpha-\beta}) + O(T^{-\beta})\right]. \qquad (23)
\end{aligned}
$$

The term $\tilde{B}_3$ is of order $O(T^{-\alpha-\beta})$ by the same arguments as in the proof of Lemma 2. Consequently, parts i.) and ii.) of the Lemma follow directly from Equations (17) and (20). Similarly, part iii.) is the direct consequence of Equation (19) and Equations (22) and (23). □

Proof of Theorem 1:

From Lemma 1 to 3, we have $I_\mu(\lambda_j) \sim Tj^{-2}\kappa$ under Assumption A2, $I_\mu(\lambda_j) \sim Tj^{-2}\tilde{\kappa}$ under Assumption A1, and $I_\mu(\lambda_j) \sim Tj^{-2}\kappa_{P,T}$ under Assumption A3 with $\beta = 0$, where κ, $\tilde{\kappa}$ and $\kappa_{P,T}$ are two finite constants and a random variable with finite variance, respectively. $E[I_\mu(\lambda_j)] = G_\mu\frac{T}{j^2}$ therefore follows immediately.

To prove that the rank of G_μ is reduced if and only if μ_t has common low frequency contaminations, we first show that co-shifting according to Definition 1 implies a reduced rank of G_μ, and then we show that G_μ has full rank, if μ_t is not co-shifting.

For the first part, note that $\Phi'\mu_t = 0 \Leftrightarrow \mu_{at} = \phi_b/\phi_a\mu_{bt}$. Let $\mu_{bt} = \omega_t$, then

$$
\mu_t = \begin{pmatrix} \mu_{at} \\ \mu_{bt} \end{pmatrix} = \begin{pmatrix} \phi_b/\phi_a \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & \phi_b/\phi_a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \omega_t \end{pmatrix}
$$

Therefore,

$$
\begin{aligned}
f_\mu(\lambda_j) &= \begin{pmatrix} 0 & \phi_b/\phi_a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & f_\omega(\lambda_j) \end{pmatrix}\begin{pmatrix} 0 & \phi_b/\phi_a \\ 0 & 1 \end{pmatrix}' \\
&= f_\omega(\lambda_j)\begin{pmatrix} (\phi_b/\phi_a)^2 & \phi_b/\phi_a \\ \phi_b/\phi_a & 1 \end{pmatrix},
\end{aligned}
$$

so that $det\left(f_\mu(\lambda_j)\right) = (\phi_b/\phi_a)^2 - (\phi_b/\phi_a)^2 = 0$.

For the second part, let $\Phi'\mu_t = c_t$, where $\exists\ c_t \neq 0$, then $\Phi'\mu_t = c_t \Leftrightarrow \mu_{at} = \phi_a^{-1} c_t - \phi_a/\phi_b \mu_{bt}$, so that for $\mu_{bt} = \omega_t$

$$\mu_t = \begin{pmatrix} \mu_{at} \\ \mu_{bt} \end{pmatrix} = \begin{pmatrix} \phi_a^{-1} & -\phi_a/\phi_b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_t \\ \omega_t \end{pmatrix}$$

Then, denoting $\tilde{\omega}_t = (c_t,\ \omega_t)'$, we have

$$\begin{aligned} f_{\mu_t}(\lambda_j) &= \begin{pmatrix} \phi_a^{-1} & -\phi_b/\phi_a \\ 0 & 1 \end{pmatrix} f_{\tilde{\omega}}(\lambda_j) \begin{pmatrix} \phi_a^{-1} & 0 \\ -\phi_b/\phi_a & 1 \end{pmatrix} \\ &= \begin{pmatrix} \phi_a^{-2} f_{c\omega}(\lambda_j) - \phi_b \phi_a^{-2} f_{\omega c}(\lambda_j) - \phi_b \phi_a^{-2} f_{c\omega}(\lambda_j) + \phi_b^2 \phi_a^{-2} f_{\omega\omega}(\lambda_j)^2 & \phi_a^{-1} f_{c\omega}(\lambda_j) - \phi_b \phi_a^{-1} f_{\omega\omega}(\lambda_j) \\ \phi_a^{-1} f_{\omega c}(\lambda_j) - \phi_a \phi_b^{-1} f_{\omega\omega}(\lambda_j) & f_{\omega\omega}(\lambda_j) \end{pmatrix}, \end{aligned}$$

so that $det\,[f_{\mu_t}(\lambda_j)] = \phi_a^{-1}\,(f_{cc}(\lambda_j) f_{\omega\omega}(\lambda_j) - f_{c\omega}(\lambda_j) f_{\omega c}(\lambda_j)) \neq 0$. □

To prove Theorem 2 we need the following Lemma.

Lemma 6. *For $e > -1$*

$$\lim_{m_1 \to \infty} \sum_{j=1}^{m_1} j^e = \zeta(-e) + \frac{m_1^{e+1}}{e+1} + O(m_1^e),$$

where ζ is the Riemann zeta function.

Proof of Lemma 6:

From the extension of the Faulhaber formula derived by McGown and Parks (2007)

$$\begin{aligned} \lim_{m_1 \to \infty} \left[(e+1) \sum_{j=1}^{m_1} j^e - m^\gamma F_e(m_1) \right] &= (e+1)\zeta(-e) \\ \lim_{m_1 \to \infty} (e+1) \sum_{j=1}^{m_1} j^e &= (e+1)\zeta(-e) + m_1^\gamma F_e(m_1) \\ \lim_{m \to \infty} \sum_{j=1}^{m_1} j^e &= \zeta(-e) + \frac{m_1^\gamma F_e(m_1)}{(e+1)}, \end{aligned}$$

where

$$\begin{aligned} F_e(m_1) &= m_1^{\lfloor e \rfloor + 2} + \sum_{k=1}^{\lfloor e \rfloor + 1} (-1)^k \binom{e+1}{k} B_k m_1^{\lfloor e \rfloor + 2 - k} \\ &= m_1^{\lfloor e \rfloor + 2} + O\left(m_1^{\lfloor e \rfloor + 1}\right), \end{aligned}$$

the constants B_k are the Bernoulli numbers, and $\gamma = -(\lfloor e \rfloor) + 1 - e = e - \lfloor e \rfloor - 1$, so that

$$\begin{aligned} \lim_{m \to \infty} \sum_{j=1}^{m_1} j^e &= \zeta(-e) + \frac{m_1^{e - \lfloor e \rfloor - 1} \left(m_1^{\lfloor e \rfloor + 2} + O\left(m_1^{\lfloor e \rfloor + 1}\right)\right)}{e+1} \\ &= \zeta(-e) + \frac{m_1^{e+1} + O\left(m_1^e\right)}{e+1}. \end{aligned}$$

$\square$

Proof of Theorem 2:

To prove the theorem we show that the difference of $\hat{G}_y(d,l,m_1)$ and $\hat{G}_x(d,1,m_1)$ vanishes in probability. Let therefore $\Delta(d) = \|\hat{G}_y(d,l,m_1) - \hat{G}_x(d,1,m_1)\|$, then

$$\Delta(d) = \|(m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_\mu(\lambda_j) \Lambda_j^*(d) - m_1^{-1} \sum_{j=1}^{l-1} \Lambda_j(d) I_x(\lambda_j) \Lambda_j^*(d) + Z + R\|,$$

where $Z = (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_{x\mu}(\lambda_j) \Lambda_j^*(d) + (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_{\mu x}(\lambda_j) \Lambda_j^*(d)$ and $R = (l-1)(m_1(m_1 - l + 1))^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_x(\lambda_j) \Lambda_j^*(d)$. Furthermore, define $\nu^- = (d_a - d_a^0) - (d_b - d_b^0)$, $\nu^+ = (d_a - d_a^0) + (d_b - d_b^0)$, and $I_x(\lambda_j) = \lambda_j^{-d_a^0 - d_b^0} \chi_j e^{i\pi(d_a^0 - d_b^0)/2}$, where χ_j is a random matrix with $E[\chi_j] = I$ and $Var[\chi_j] < \infty$. Similarly, $I_\mu(\lambda_j) = \kappa_j \lambda_j^{-2} T^{-1}$, where $E[\kappa_j] < \infty$ and $Var[\kappa_j] < \infty$. Then

$$\begin{aligned}\Delta(d) &= \|\frac{e^{i\pi(d_a - d_b)/2}}{(m_1 - l + 1)T} \sum_{j=l}^{m_1} \lambda_j^{d_a + d_b - 2} \kappa_j - \frac{e^{i\pi\nu^-/2}}{m_1} \sum_{j=1}^{l-1} \lambda_j^{\nu^+} \chi_j + Z + R\| \\ &= \|\frac{(2\pi)^{d_a + d_b - 2} e^{i\pi(d_a - d_b)/2}}{(m_1 - l + 1)T^{d_a + d_b - 1}} \sum_{j=l}^{m_1} j^{d_a + d_b - 2} \kappa_j - \frac{(2\pi)^{\nu^+} e^{i\pi\nu^-/2}}{m_1 T^{\nu^+}} \sum_{j=1}^{l-1} j^{\nu^+} \chi_j + Z + R\|.\end{aligned}$$

Due to the independence of x_t and μ_t, $Z \xrightarrow{p} 0$ as $T \to \infty$. Obviously, $R \xrightarrow{p} 0$ for $T \to \infty$ holds as well. For $l = 1$, the second sum is empty and for the first sum it holds with $A = \|(m_1 - l + 1)^{-1} T^{-d_a - d_b + 1} \sum_{j=l}^{m_1} j^{d_a + d_b - 2} \kappa_j\|$ and $d_a + d_b > 1$ from Lemma 6

$$A \leq \max \|\kappa_j\| \frac{O(m_1^{d_a + d_b - 2})}{T^{d_a + d_b - 1}} = o_p(1).$$

For $d_a + d_b < 1$, we have by definition of the Riemann ζ-function

$$A \leq \frac{\max \|\kappa_j\| \zeta(-d_a - d_b + 2)}{m_1 T^{d_a + d_b - 1}},$$

which is $o_P(1)$, for $\delta_{m_1} > 1 - d_a - d_b$.

For the second part of the theorem, we have $\sum_{j=l}^{m_1} j^{d_a + d_b - 2} \leq m_1 l^{d_a + d_b - 2}$. Therefore, with $d_a + d_b < 1$,

$$A \leq \frac{m_1 l^{d_a + d_b - 2}}{(m_1 - l + 1)T^{d_a + d_b - 1}} = O(l^{d_a + d_b - 2} T^{-(d_a + d_b - 1)}) = o_P(1),$$

for $\delta_{m_1} > \delta_l$ and $l = T^{(d_a + d_b - 1)/(d_a + d_b - 2) + \upsilon}$. Furthermore, let $B = \|m_1^{-1} T^{-\nu^+} \sum_{j=1}^{l-1} j^{\nu^+} \chi_j\|$, then

$$B \leq \frac{\max_{j<l} \|\chi_j\|}{m_1 T^{\nu^+}} \sum_{j=1}^{l-1} j^{\nu^+} = \frac{\max_{j<l} \|\chi_j\|}{m_1 T^{\nu^+}} O(l^{\nu^+ + 1}) = o_P(1),$$

for $l = o(T^{(\nu^+ + \delta_{m_1})/(\nu^+ + 1)})$. $\square$

Proof of Theorem 3:
The proof directly follows from a Taylor expansion of the matrix $\hat{G}_y(\hat{d}(m), l, m_1)$ at d^0 and is omitted here.
Proof of Theorem 4:

The proof follows ideas in Robinson (2008b). For any $c>0$ define neighbourhoods $N_\beta(c) = \{\beta : |\beta-\beta^0| < c\}$ and $N_d(c) = \{d : \|d-d^0\| < c\}$. Furthermore, fix $\varepsilon > 0$ and define $N(\varepsilon) = N_\beta(\varepsilon^{-1}(\frac{m}{T})^{\nu^0})N_d(\varepsilon)$, $\overline{N(\varepsilon)} = \Theta \backslash N(\varepsilon)$, and $\zeta_i = d_i - d_i^0$.
We split the parameter space into two. For a constant $0 < C \leq \frac{1}{8}$ define $\Theta_{da} = \{d \in \Theta_d : \zeta_a \geq -\frac{1}{2}+C; \zeta_b \geq -\frac{1}{2}+C\}$ and $\Theta_{db} = \Theta_d \backslash \Theta_{da}$. Since $P(\hat\theta \in N(\varepsilon)) \leq P(\inf_{\bar{N(\varepsilon)}}\{R(\theta)-R(\theta^0)\} \leq 0)$, the consistency of $\hat\Theta$ follows if we show

$$P\left(\inf_{\overline{N(\varepsilon)}\cap\{\Theta_\beta\times\Theta_{da}\}}\{R(\theta)-R(\theta^0)\}\leq 0\right)\to 0 \quad \text{as} \quad T\to\infty \quad \text{and} \tag{24}$$

$$P\left(\inf_{\overline{N(\varepsilon)}\cap\{\Theta_\beta\times\Theta_{db}\}}\{R(\theta)-R(\theta^0)\}\leq 0\right)\to 0 \quad \text{as} \quad T\to\infty. \tag{25}$$

First we show (24). Rewrite $R(\theta)-R(\theta^0)$ as

$$R(\theta)-R(\theta^0) = \log\det\left[\hat\Omega^{tri}(\theta)\hat\Omega^{tri}(\theta^0)^{-1}\right] - 2(\zeta_a+\zeta_b)\frac{1}{m-l+1}\sum_{j=l}^{m}\log\lambda_j,$$

with $\hat\Omega^{tri}(\theta) = \frac{1}{m-l+1}\sum_{j=l}^{m} Re[\Lambda_j(d)BI_{yy}^{tri}(\lambda_j)B^{'}\Lambda_j(d)^*]$ instead of $\hat\Omega(\theta)$ in Robinson (2008b).
Define a vector type II $I(d_a^0, d_b^0)$ process as

$$\xi_t = \begin{pmatrix}\xi_{at}\\ \xi_{bt}\end{pmatrix} = B^0, z_t = \begin{pmatrix}(1-L)^{-d_a^0} & y_{at} & 1(t\geq 1)\\ (1-L)^{-d_b^0} & y_{bt} & 1(t\geq 1)\end{pmatrix}, B^0 = \begin{pmatrix}1 & -\beta^0\\ 0 & 1\end{pmatrix}$$

Define further analogously to Robinson (2008b)

$$H_j = (h_{k_1k_2j}) = \Lambda_j(d^0)I_{\xi j}^{tri}(\lambda_j)\Lambda_j(d^0)^* \quad \text{and}$$

$$\hat G^{(1)}(d) = (\hat g^{(1)}_{k_1k_2}), \quad \text{where } \hat g^{(1)}_{k_1k_1} = \frac{1}{m-l+1}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2\zeta_{k_1}} h_{k_1k_1j} \quad \text{and}$$

$$\hat g^{(1)}_{ab} = \hat g^{(1)}_{ba} = \frac{1}{m-l+1}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{\zeta_a+\zeta_b}\left(e^{i(\pi-\lambda_j)(\zeta_b-\zeta_a)/2}h_{abj} + e^{-i(\pi-\lambda_j)(\zeta_b-\zeta_a)/2}h_{baj}\right).$$

Similar to Robinson (2008b) we obtain $R(\theta)-R(\theta^0) = U_d(d)+U_\beta(\theta)$ with

$$U(d) = \log\det\left[\Upsilon(d)\hat G^{(1)}(d)\Upsilon(d)\hat G^{(1)}(d^0)^{-1}\right] + \Phi_a(d,l) + u(d) + \Phi_b(d,l) \quad \text{and}$$

$$U_\beta(\theta) = \log\det\left[\hat\Omega^{tri*}(\theta)\hat G^{(1)}(d)^{-1}\right] - \Phi_a(d,l) + \Phi_b(d,l).$$

Here,

$$
\begin{aligned}
&\Upsilon(d) = diag\left((2\zeta_a+1)^{1/2}, (2\zeta_b+1)^{1/2}\right),\\
&\hat{\Omega}^{tri*}(\theta) = \Xi(\theta)\hat{\Omega}^{tri}(\theta)\Xi(\theta),\\
&\Xi(\theta) = diag\left(\lambda_m^{-\zeta_a}, \lambda_m^{-\zeta_b}\right),\\
&\Phi_1(d,l) = \log\left[(l-1)^2(l^{2\zeta_a+1})^{-1}(l^{2\zeta_b+1})^{-1}\right],\\
&\Phi_2(d,l) = 2(\zeta_a+\zeta_b)(l-1)^{-1}l\log l, \quad \text{and}\\
&u(d) = \sum_{i=a,b}\left[2\zeta_i - \log(2\zeta_i+1) + 2\zeta_i\left(\log m - \frac{1}{m-l+1}\sum_{j=l}^{m}\log j - 1\right)\right].
\end{aligned}
$$

The functions $\Phi_a(d,l)$ and $\Phi_b(d,l)$ control effects of taking summations from l by application of the Euler-McLaurin formula as in Lemma 2(a) of Shimotsu (2010). In contrast to Robinson (2008b) all matrices here are defined by the trimmed periodogram and we do not have the parameter γ.

Now, (24) follows if we show that, as $T \to \infty$

$$
P\left(\inf_{\overline{N(\varepsilon)}\cap\Theta_{d_a}} U_d(d) \le 0\right) \to 0 \quad \text{and} \tag{26}
$$

$$
P\left(\inf_{\overline{N_\beta}(\frac{1}{\varepsilon}(\frac{T}{m})^{\nu_0})\times\Theta_d} U_\beta(d) \le 0\right) \to 0. \tag{27}
$$

The proof of (26) is similar to Robinson (2008b). Define the population analogue of $\hat{g}^{(1)}_{k_1k_2}$ as

$$
g^{(1)}_{k_1k_2} = \omega_{k_1k_1}\frac{1}{l}\int_l^1 x^{2\zeta_{k_1}}dx
$$

and

$$
g^{(1)}_{ab} = g^{(1)}_{ba} = \omega_{ab}\frac{1}{l}\int_l^1 x^{\zeta_a+\zeta_b}dx\cos\tau,
$$

where

$$
\tau = (\zeta_b - \zeta_a)\frac{\pi}{2}.
$$

Then (26) holds if

$$
\sup_{\Theta_{da}}\left|\left|\Upsilon(d)[\hat{G}^{(1)}(d) - G^{(1)}(d)]\Upsilon(d)\right|\right| \xrightarrow{P} 0, \tag{28}
$$

$$
\sup_{\Theta_{da}}\left|\left|\left[\Upsilon(d)G^{(1)}(d)\Upsilon(d)\right]^{-1}\right|\right| < \infty, \tag{29}
$$

$$
\inf_{\overline{N_d(\varepsilon)}\cap\Theta_{d_a}}\left[\log\det\left[\Upsilon(d)G^{(1)}(d)\Upsilon(d)G^{(1)}(d^0)^{-1}\right] + \Phi_a(\theta,l)\right] \ge 0, \quad \text{and} \tag{30}
$$

$$\lim_{T\to\infty} \inf_{\overline{N_d(\varepsilon)\cap\Theta_{d_a}}} [u(d)-\Phi_b(d,l)] > 0. \tag{31}$$

These conditions correspond to (7.5) - (7.8) in Robinson (2008b).
The proof of (28) follows from observing that

$$\begin{aligned}
&\left|\left|\frac{1}{m-l+1}\sum_{j=l}^{m} Re\left[\Lambda_j(d)BI_y^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right|\right| \\
=&\left|\left|\frac{1}{m-l+1}\sum_{j=l}^{m} Re\left[\Lambda_j(d)B(I_x^{tri}(\lambda_j)+I_{x\mu}^{tri}(\lambda_j)+I_{\mu x}^{tri}(\lambda_j)+I_\mu^{tri}(\lambda_j))B'\Lambda_j(d)^*\right]\right|\right|.
\end{aligned}$$

Now,

$$\begin{aligned}
&\frac{1}{m-l+1}\sum_{j=l}^{m} Re\left[\Lambda_j(d)BI_\mu^{tri}(\lambda_j)B'\Lambda_j(d)^*\right] \\
=&O_P\left(\frac{1}{m}\left(\frac{\delta_M}{T}\right)^{2\left(\frac{d_a+d_b}{2}-\frac{1}{2}\right)}\right) \\
=&o_P(1).
\end{aligned}$$

In addition to this we have because $||I_{\mu x}(\lambda_j)||^2 = I_\mu(\lambda_j)I_x(\lambda_j)$

$$\begin{aligned}
&\left|\left|\frac{1}{m-l+1}\sum_{j=l}^{m} Re\left[\Lambda_j(d)BI_{\mu x}^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right|\right| \\
\leq&\frac{1}{m-l+1}\sum_{j=l}^{m}\left|\left|Re\left[\Lambda_j(d)BI_{\mu x}^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right|\right| \\
\leq&\frac{1}{m-l+1}\sum_{j=l}^{m}\left(Re\left[\Lambda_j(d)BI_x^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right)^{\frac{1}{2}} \\
&\cdot\left(Re\left[\Lambda_j(d)BI_\mu^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right)^{\frac{1}{2}} \\
\leq&\underbrace{\left(\frac{1}{m-l+1}\sum_{j=l}^{m} Re\left[\Lambda_j(d)BI_x^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right)^{\frac{1}{2}}}_{O_P(1)} \\
&\cdot\underbrace{\left(\frac{1}{m-l+1}\sum_{j=l}^{m} Re\left[\Lambda_j(d)BI_\mu^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right)^{\frac{1}{2}}}_{o_P(1)}.
\end{aligned}$$

Applying now the same arguments as in the proof of (17) in Shimotsu (2012) gives (28).
Also the proof of (29), (30) and (31) is equal to Shimotsu (2012) proving his Equations (18), (19) and (20).
We proceed to show (27). Define $\hat{g}_{k_1k_2}^{(i)}$ similarly to Robinson (2008b) but using $\frac{1}{m-l+1}\sum_{j=l}^{m}$ and setting $\tau = (\zeta_b-\zeta_a)\frac{\pi}{2}$ and $\gamma^0 = (d_b^0-d_a^0)\frac{\pi}{2}$. Let further

$$\begin{aligned}
\hat{\alpha}_a &= \left(\hat{g}_{aa}^{(2)}\hat{g}_{bb}^{(1)} - 2\hat{g}_{ab}^{(1)}\hat{g}_{ab}^{(2)}/\det(\hat{G}^{(1)}(d)\right), \quad \text{and} \\
\hat{\alpha}_b &= \left(\hat{g}_{aa}^{(3)}\hat{g}_{bb}^{(1)} - (\hat{g}_{ab}^{(2)})^2)/\det(\hat{G}^{(1)}(d)\right).
\end{aligned}$$

Define $g^{(i)}_{k_1k_2}$, the population counterpart of $\hat{g}^{(i)}_{k_1k_2}$, analogously to $g^{(1)}_{k_1k_2}$: for example,

$$g^{(2)}_{ab} = g^{(2)}_{ba} = \frac{1}{l}\omega_{bb}\cos\gamma\int_l^1 x^{d_a-d_b^0+\zeta_b}dx, \quad \text{and}$$
$$g^{(3)}_{aa} = \frac{1}{l}\omega_{bb}\int_l^1 x^{2(d_a-d_b^0)}dx,$$

where $\gamma=(d_b-d_a)\frac{\pi}{2}$.
Using summation by parts and Lemma 1(b) of Shimotsu (2012), we obtain

$$\sup_{\Theta_d}|\hat{g}^{(i)}_{k_1k_2}-g^{(i)}_{k_1k_2}| \xrightarrow{P} 0$$

for $i=1,2,3$, $k_1,k_2=a,b$ as $T\to\infty$.
Rewrite $U_\beta(d)=\log Q(b_n(\beta))-\Phi_a(d,l)+\Phi_b(d,l)$, where $Q(s)=1+\hat{a}_1 s+\hat{a}_2 s^2$ and $b_n(\beta)=\lambda_j^{-\nu^0}(\beta^0-\beta)$. Define

$$\alpha_a = \left(g^{(2)}_{aa}g^{(1)}_{bb}-2g^{(1)}_{ab}g^{(2)}_{ab}/\det(G^{(1)}(d)\right) \quad \text{and}$$
$$\alpha_b = \left(g^{(3)}_{aa}g^{(1)}_{bb}-(g^{(2)}_{ab})^2)/\det(G^{(1)}(d)\right).$$

Following Robinson (2008b), the probability in (27) is bounded by, with $\rho=\sup_{\Theta_d}|\Phi_a(d,l)-\Phi_b(d,l)|<\infty$,

$$P\left(\log\left[1-\left(\sup_{\Theta_d}\frac{|\hat{\alpha}_a|}{\varepsilon}+\inf_{\Theta_d}\frac{|\hat{\alpha}_b|}{\varepsilon^2}\right)\right]\leq\rho\right)+P\left(\sup_{\Theta_d}\frac{|\hat{\alpha}_a|}{2|\hat{\alpha}_b|}>\frac{1}{\varepsilon}\right)$$
$$\leq 2P\left(\sup_{\Theta_d}|\hat{\alpha}_a-a_a|+\frac{2}{\varepsilon}\sup_{\Theta_d}|\hat{\alpha}_b-\alpha_b|+\varepsilon\rho\geq\frac{1}{\varepsilon}\inf_{\Theta_d}\alpha_b-\sup_{\Theta_d}|\alpha_a|\right),$$

which has an additional term $\varepsilon\rho$ compared to (7.13) in Robinson (2008b). (27) follows now exactly as in Shimotsu (2012).

It remains to show (25). Write

$$R(\theta)-R(\theta^0)=U_d^*(d)+U_\beta^*(\theta),$$

where

$$U_d^*(d)=\log\,\det\left[\Xi(d)\hat{G}^{(1)}(d)\Xi(d)\hat{G}^{(1)}(d^0)^{-1}\right]$$
$$-2(\zeta_a+\zeta_b)\frac{1}{m-l+1}\sum_{j=l}^m\log\,\lambda_j$$

and

$$U_\beta^*(\theta)=\log\,\det[\hat{\Omega}^{tri*}(\theta)\hat{G}^{(1)}(d)^{-1}]=U_\beta(\theta)+\Phi_a(d,l)-\Phi_b(d,l).$$

Then

$$P\left(\inf_{N_\beta(\varepsilon^{-1})(\frac{T}{m})^{\nu^0})\times\Theta_d}U_\beta^*(\theta)\leq 0\right)\to 0$$

follows from the proof of (27), so it suffices to show

$$P\left(\inf_{\Theta_{db}} U_d^*(d) \leq 0\right) \to 0$$

Rewrite $U_d^*(d)$ as

$$U_d^*(d) = \log\det \hat{D}(d) - \log\det \hat{D}(d^0),$$

where

$$\hat{D}(d) = \frac{1}{m-l+1}\sum_{j=l}^{m}\begin{pmatrix} \left(\frac{j}{q}\right)^{2\zeta_a} h_{aaj} & \left(\frac{j}{q}\right)^{\zeta_a+\zeta_b} Re\left(e^{i(\pi-\lambda_j)\frac{(\zeta_b-\zeta_a)}{2}} h_{abj}\right) \\ \left(\frac{j}{q}\right)^{\zeta_a+\zeta_b} Re\left(e^{i(\pi-\lambda_j)\frac{(\zeta_b-\zeta_a)}{2}} h_{abj}\right) & \left(\frac{j}{q}\right)^{2\zeta_b} h_{bbj} \end{pmatrix}$$

and

$$q = \exp\left(\frac{1}{m-l+1}\sum_{j=l}^{m}\log j\right)$$

Define K(d) as $\hat{D}(d)$ but $h_{k_1k_2j}$ is replaced with $\omega_{k_1k_2}$.
Now,

$$\sup_{\Theta_{db}} |D(d) - K(d)| \xrightarrow{P} 0$$

follows from Lemma 1 of Shimotsu (2010) and the proof of Theorem 1 of Shimotsu (2007). Furthermore, it follows from Shimotsu (2007) and Shimotsu (2012) that there exists an $\varepsilon \in (0, 0.1)$ and $l < m$ such that

$$\inf_{\Theta_{db}} \det K(d) \geq (1+\varepsilon)\det G^0 + o(1).$$

Therefore,

$$\det \hat{D}(d) \geq (1+\varepsilon)\det G^0 + o(1).$$

Since

$$\det \hat{D}(d^0) = \det \hat{G}^{(1)}(d^0) = \det \Omega^0 + o_P(1)$$

from (28), we establish (25). □

Proof of Theorem 5:

$\hat{\theta}$ has now the stated limiting distribution if for any $\tilde{\theta}$ such that $\tilde{\theta} - \theta^0 = O_P(m^{-1/2})$,

$$\sqrt{m}(\Delta_T)^{-1}\frac{dR(\theta^0)}{d\theta}|_{\theta^0} \overset{d}{\Rightarrow} N(0, \Xi)$$

and

$$(\Delta_T)^{-1}\frac{d^2R(\tilde{\theta})}{d\theta d\theta'}(\Delta_T)^{-1}\xrightarrow{P}\Xi.$$

For the score vector approximation denote by $s_k(\theta)$ the k-th element of $\frac{dR(\theta)}{d\theta}$ and by E_{ij} the matrix of zeros where the ij-th element has been replaced by a one. Following exactly the lines of Robinson (2008b), Theorem 4 we have that

$$\begin{aligned}
s_1(\theta^0) &= -tr\frac{1}{m-l+1}\sum_{j=l}^{m}\lambda_j^{d_b^0-d_a^0}(E_{12}Re[\Lambda_j(d^0)B^0I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]e^{i\pi(d_b-d_a)/2} \\
&\quad +Re[\Lambda_j(d^0)B^0I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{21}e^{-i\pi(d_b-d_a)/2})\hat{G}(d^0)^{-1} \\
&= -tr\frac{1}{m-l+1}\sum_{j=l}^{m}\lambda_j^{d_b^0-d_a^0}(E_{12}Re[\Lambda_j(d^0)B^0I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]e^{i\pi(d_b-d_a)/2} \\
&\quad +Re[\Lambda_j(d^0)B^0I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{21}e^{-i\pi(d_b-d_a)/2})\hat{G}(d^0)^{-1}+o_P(1),
\end{aligned}$$

$$\begin{aligned}
s_2(\theta^0) &= itr\Bigg[\frac{1}{m-l+1}\sum_{j=l}^{m}(Re[\Lambda_j(d^0)B^0I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{22} \\
&\quad -E_{22}Re[\Lambda_j(d^0)B^0I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*])\hat{G}(d^0)^{-1}\Bigg] \\
&= itr\Bigg[\frac{1}{m-l+1}\sum_{j=l}^{m}(Re[\Lambda_j(d^0)B^0I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{22} \\
&\quad -E_{22}Re[\Lambda_j(d^0)B^0I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*])\hat{G}(d^0)^{-1}\Bigg]+o_P(1),
\end{aligned}$$

$$\begin{aligned}
s_{2+k}(\theta^0) &= tr\frac{1}{m-l+1}\sum_{j=l}^{m}(\log\lambda_j-\frac{1}{m-l+1}\sum_{j=l}^{m}\log\lambda_j) \\
&\quad (E_{kk}Re[\Lambda_j(d^0)B^0I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]+Re[\Lambda_j(d^0)B^0I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{kk})\hat{G}(d^0)^{-1} \\
&= tr\frac{1}{m-l+1}\sum_{j=l}^{m}(\log\lambda_j-\frac{1}{m-l+1}\sum_{j=l}^{m}\log\lambda_j) \\
&\quad (E_{kk}Re[\Lambda_j(d^0)B^0I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]+Re[\Lambda_j(d^0)B^0I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{kk})\hat{G}(d^0)^{-1},
\end{aligned}$$

for $k=1,2$ and by the same arguments as in the consistency proof. The score vector approximation follows now directly as in Robinson (2008b).
The Hessian approximation is also similar to the arguments in Robinson (2008b) and therefore omitted here. □