Essays on Long Memory Time Series

Von der Wirtschaftswissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

– Doctor rerum politicarum –

genehmigte Dissertation

von

M.Sc. Christian Hendrik Leschinski

geboren am 2. Juni 1987 in Peine

2016

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover

Koreferent: Prof. Dr. Uwe Hassler, Goethe-Universität Frankfurt am Main

Tag der Promotion: 05.07.2016

Acknowledgements

First, I want to thank my advisor and co-author Prof. Dr. Philipp Sibbertsen for his encouragement and guidance, for hours and hours of inspiring discussions and for providing me with the right work environment to pursue my research projects. I also want to thank Prof. Dr. Uwe Hassler for taking the time to be my second examiner and Prof. Dr. Marcel Prokopczuk for chairing my examination board.
Moreover, I am grateful to my co-authors: Robinson Kruse, for his advice and many inspiring discussions, and Marie Busch and Michael Will, for their effort and companionship. I truly learned a lot from each of them.
During the first year I was fortunate to share the office with Dr. Hendrik Kaufmann and to work with Dr. Philip Bertram. Both of them shared their experience with me and helped me to get started, which made everything thereafter much easier. Thanks are also due to all other colleagues and former colleagues, who always kept an open door and were happy to help.
Finally, I want to thank my friends and family - especially my parents, for their everlasting support, and my wife Melina, for her patience and love, for always being understanding, and for believing in me during all the little setbacks along the way.

Berlin, March 2016
Christian Leschinski

Kurzzusammenfassung

Diese Dissertation beinhaltet vier Aufsätze über Zeitreihen mit Langem Gedächtnis. Nach der Einleitung in Kapitel 1, stellt der erste Aufsatz in Kapitel 2 einen multivariaten Score-artigen Misspezifikationstest für Prozesse mit Langem Gedächtnis vor. Die Teststatistik basiert auf der gewichteten Summe der partiellen Ableitungen der multivariaten lokalen Whittle-Likelihood-Funktion und kann benutzt werden, um zwischen wahrem und scheinbarem Langen Gedächtnis zu unterscheiden. Konsistenz beweisen wir insbesondere für die Alternativen stochastischer Niveauverschiebungen und glatter Trends. Um den Test auf ein fraktional kointegriertes System anwenden zu können, wird die Teststatistik nach der Schätzung der Kointegrationsmatrix für das linear transformierte System berechnet. Wir zeigen, dass die Grenzverteilung von diesem Vorgehen unberührt bleibt. Eine Monte Carlo Analyse zeigt, dass der Test gute Size- und Power-Eigenschaften in endlichen Stichproben hat. Um die Nützlichkeit des Tests in der Praxis hervorzuheben, wenden wie ihn auf zwei Maße für die Variation von Aktienindizes an - die log-absolute Rendite und die realisierte Volatilität - und zwar für den S&P 500, den DAX, den FTSE und den NIKKEI.
In Kapitel 3 schlagen wir eine automatische Modellselektionsprozedur für k-Faktor-Gegenbauer-Prozesse vor. Die Prozedur basiert auf sequentiellen Tests des Maximums des Periodograms und semiparametrischen Schätzern der Modellparameter. Als Nebenprodukt führen wir eine generalisierte Version von Walkers „Large sample g-test“ ein, die es erlaubt in stationären Prozessen mit kurzem Gedächtnis auf persistente Periodizität zu testen. Unsere Simulationsstudien zeigen, dass die Prozedur die korrekte Modellordnung unter verschiedensten Umständen korrekt identifiziert. Eine Anwendung auf kalifornische Stromlast-Daten verdeutlicht den Wert der Prozedur in empirischen Analysen und liefert neue Erkenntnisse über die Periodizität dieses Prozesses.

Im darauf folgenden Kapitel 4 leite ich das Gedächtnis der Produktzeitreihe $x_t y_t$ her, wobei x_t und y_t stationäre Zeitreihen mit Langem Gedächtnis der Ordnungen d_x und d_y sind. Besondere Aufmerksamkeit wird auf quadrierte Reihen gelegt und auf Produkte von Reihen die von einem gemeinsamen stochastischen Faktor getrieben werden. Es stellt sich heraus, dass das Gedächtnis von Produkten von Zeitreihen mit von null verschiedenen Mittelwerten vom maximalen Gedächtnis der Faktorreihen bestimmt wird. Dahingegen wird das Gedächtnis reduziert, wenn die Mittelwerte der Reihen null sind.
Zum Schluss beschäftigt sich Kapitel 5 mit Prognose-Vergleichen mittels des populären Diebold-Mariano Tests unter Langem Gedächtnis. Es wird gezeigt, dass Langes Gedächtnis von Prognosen oder der zu prognostizierenden Variablen auf die Prognosefehler-Verlustdifferentiale übertragen werden kann und dass der konventionelle Diebold-Mariano Test in diesem Fall nicht mehr valide ist. Zwei robuste Statistiken basierend auf einem Gedächtnis- und Autokorrelations - konsistenten Schätzer und einem erweiterten fixed-b Ansatz werden diskutiert und ihre relative Performance wird in einer Monte Carlo Studie evaluiert. Anschließend werden die robusten Statistiken benutzt, um die relative Prognosegüte aktueller Erweiterungen des heterogenen autoregressiven

Modells für realisierte Volatilitäten des S&P 500 Index zu evaluieren. Wir kommen zu dem Ergebnis, dass sich Prognosen signifikant verbessern, wenn Sprünge im log-Preis-Prozess separat von kontinuierlichen Komponenten behandelt werden. Im Gegensatz dazu stellt sich heraus, dass Verbesserungen die durch die Inklusion von implizierter Volatilität erreicht werden insignifikant sind.
Schlagworte: Langes Gedächtnis · Semiparametrische Schätzung · Zeitreihenanalyse

Short Summary

This thesis contains four essays on long memory time series. After an introduction in Chapter 1, the first essay in Chapter 2 provides a multivariate score-type misspecification test for long memory processes. The test statistic is based on the weighted sum of the partial derivatives of the multivariate local Whittle likelihood function and can be used to distinguish between true and spurious long memory. In particular, we prove the consistency of our test against the alternatives of random level shifts or smooth trends. To apply the test to fractionally cointegrated systems, the test statistic is calculated for the linearly transformed system after estimating the cointegrating matrix. We show that the limit distribution is asymptotically unaffected by this procedure. A Monte Carlo analysis shows good finite sample properties of the test in terms of size and power. To highlight the usefulness of the test in practice, we apply it to two measures for the variation of stock market indices: the log-absolute returns and log-realized volatilities of the S&P 500, the DAX, the FTSE, and the NIKKEI.

In Chapter 3, we propose an automatic model order selection procedure for k-factor Gegenbauer processes. The procedure is based on sequential tests of the maximum of the periodogram and semiparametric estimators of the model parameters. As a byproduct, we introduce a generalized version of Walker's "large sample g-test" that allows to test for persistent periodicity in stationary short memory processes. Our simulation studies show that the procedure performs well in identifying the correct model order under various circumstances. An application to electricity load data from the state of California illustrates its value in empirical analyses and allows new insights into the periodicity of this process.

In the following Chapter 4, I derive the memory of the product series $x_t y_t$, where x_t and y_t are stationary long memory time series of orders d_x and d_y, respectively. Special attention is paid to the case of squared series and products of series driven by a common stochastic factor. I find that the memory of products of series with non-zero means is determined by the maximal memory of the factor series, whereas the memory is reduced, if the series are mean zero.

Finally, Chapter 5 deals with forecast comparisons using the popular Diebold-Mariano test under long memory. It is shown that long memory can be transmitted from forecasts or the forecast objective to the forecast error loss differential and that the conventional Diebold-Mariano test is invalidated under these circumstances. Two robust statistics, based on a memory and autocorrelation consistent estimator and an extended fixed-bandwidth approach, are discussed and their relative performance is compared in a Monte Carlo study. The robust statistics are used to evaluate the forecast performance of recent extensions of the heterogeneous autoregressive model for the realized volatility of the S&P 500 index. It is found that forecasts improve significantly if jumps in the log-price process are considered separately from continuous components. In contrast to that, improvements achieved by the inclusion of implied volatility turn out to be insignificant.

Keywords: Long Memory · Semiparametric Estimation · Time Series Analysis

Contents

Chapter 1

Introduction

Introduction

In the frequency domain long memory is typically defined by a pole in the spectral density $f(\lambda)$ that obeys the power law $f(\lambda) \sim G\lambda^{-2d}$, as $\lambda \to 0_+$. Here G is a positive constant, λ denotes the frequency and d is the memory parameter. Equivalently, long memory time series can also be characterized by a hyperbolical decay of the autocorrelation function $\gamma(\tau)$ at large lags τ, so that $\gamma(\tau) \sim \tilde{G}\tau^{2d-1}$, as $\tau \to \infty$, for another positive constant $\tilde{G}$. In addition to applications to economic time series such as inflation, unemployment or interest rates, long memory has become a popular concept for the modelling of financial time series - in particular for volatility processes (cf., e.g. Deo et al. (2006), Martens et al. (2009) or Chiriac and Voev (2011)).
Parametrically, long memory processes are predominantly modelled by ARFIMA processes introduced in Granger (1980) and Hosking (1981). Here, the time series X_t is given by

$$\phi(L)(1-L)^d X_t = \theta(L)\varepsilon_t,$$

where $-1/2 < d \leq 1$, L is the lag-operator, $\phi(L)$ and $\theta(L)$ are the usual autoregressive and moving average lag-polynomials and ε_t is a mean-zero martingale difference sequence.
Consistent maximum likelihood estimation of the parameters of these processes requires a correct specification of the ARMA components that is unknown in practice. This is one of the reasons why semiparametric estimators are widely used. These rely on the spectral definition of long memory to estimate the parameter d from the periodogram local to the origin and do not require any specification of the short run dynamics of the process. The most popular ones among these estimators are the log-periodogram regression of Geweke and Porter-Hudak (1983) and Robinson (1995b) and the local Whittle estimator of Künsch (1987) and Robinson (1995a). Among these two, the log-periodogram estimate has the advantage of greater conceptual simplicity, whereas the local Whittle estimator has a smaller asymptotic variance.
The essays in this thesis address issues in specification testing, model selection and inference procedures from this semiparametric perspective. The first chapter is concerned with spurious long memory in multivariate time series. It is well known from the univariate time series literature that asymptotically hyperbolically decaying autocovariance structures and poles in the spectrum (similar to those of long memory processes) can be generated by other processes that feature trends and structural breaks. Examples of this literature include Granger and Ding (1996), Lobato and Savin (1998), Diebold and Inoue (2001), Granger and Hyung (2004) and Mikosch and Stărică (2004), among others. Recently, Perron and Qu (2010) showed differences in the spectral behavior and autocorrelation functions of long memory processes and processes with low frequency contaminations that cause spurious long memory. These results were used by Qu (2011) to construct a periodogram based test for the null hypothesis of pure long memory. According to the simulation results of Leccadito et al. (2015) this is the most powerful test available in many situations.

For multivariate time series however, no such methods exist. Chapter 2 therefore introduces a score-type misspecification test for multivariate systems with long memory, that generalizes the test of Qu (2011). The test statistic is based on the weighted sum of the partial derivatives of the multivariate local Whittle likelihood function and can be used to distinguish between true and spurious long memory. In particular, we prove the consistency of our test against the alternatives of random level shifts or smooth trends. By choosing the weighting scheme accordingly, one can either test the complete spectral density matrix for a misspecification local to the origin, or one can focus on particular rows and columns. For the first weighting scheme we obtain a pivotal limit distribution, whereas the second weighting scheme can be used to evaluate which series of the multivariate system might cause a possible rejection. To apply the test to fractionally cointegrated series, the test statistic is calculated for the linearly transformed system after estimating the cointegrating matrix. We show that the limit distribution is asymptotically unaffected by this procedure.
A Monte Carlo analysis is conducted to assess the finite sample properties of the test in various situations. It is found that it has good size and power properties and that it is robust to complications - such as unconditional variance- or covariance breaks, conditional heteroscedasticity and perturbations - and it has good power against an array of low frequency contaminations.
To highlight the usefulness of the test in practice, we apply it to two measures for the variation of stock indices - the log-absolute returns, and log-realized volatilities - of the S&P 500, the DAX, the FTSE, and the NIKKEI. It is found that the log-absolute return of the S&P 500 is not correctly specified as a pure long memory process, which is particularly interesting because it has often been quoted as an example of a series that may exhibit spurious long memory (cf. for example Granger and Ding (1996), Granger and Hyung (2004), Lu and Perron (2010), Varneskov and Perron (2011) and Xu and Perron (2014)). The available univariate tests, however, usually fail to reject the null hypothesis of a pure long memory process. For the realized volatility series, on the other hand, we find no evidence for contaminations.

The subject of Chapter 3 are processes with seasonal long range dependence. Parametric seasonal long memory models generalize seasonal ARIMA models so that the stochastic seasonal effects become more persistent and the magnitude of the autocovariance function declines hyperbolically at large lags. In this case, the poles in the spectrum that characterize long memory do not appear at the origin, but at the cyclical frequencies. Examples of these models include the (rigid) SARFIMA model of Porter-Hudak (1990), the flexible SARFIMA of Hassler (1994) and the k-factor GARMA processes of Woodward et al. (1998) and Giraitis and Leipus (1995) which nests the former two. An open issue in this literature is the model specification - in particular that of the number of poles and their location. If a possible specification can be inferred from theoretical considerations, the LM tests of Robinson (1994) and Hassler et al. (2009) allow to test for the null hypothesis that a given model is correct. Without a priori information however, the specification of the model order and the location of the cyclical frequencies with these methods is not possible.

We therefore propose an automatic model order selection procedure for k-factor Gegenbauer processes. The procedure is based on sequential tests of the maximum of the periodogram and semiparametric estimators of the model parameters. As a byproduct, we introduce a generalized version of Walker's large sample g-test that allows to test for persistent periodicity in stationary short memory processes. This test is combined with a result of Hidalgo and Soulier (2004), who show that the maximum of the periodogram is a consistent estimator for the location of a pole and the generalized local Whittle estimator of Arteche and Robinson (2000). Since this estimator only makes assumptions about the behavior of the spectrum in a degenerating neighborhood of the respective pole, it allows for a consistent estimation of the memory of the corresponding cycle without any knowledge about possible other persistent cyclical effects.
In each iteration of the proposed model selection procedure our modified G^*-test is applied to test for the presence of an omitted pole. If there is no rejection, the procedure terminates. Otherwise, the location of the pole is estimated by the maximum of the periodogram and the respective memory parameter is estimated using the aforementioned local Whittle estimator. Subsequently, the cycle is removed by applying the Gegenbauer filter with the estimated parameters and the next iteration is conducted on the filtered process.
The consistency of the procedure is proved and the effect of seasonal demeaning prior to the application of our procedure is also considered. Monte Carlo simulations show that the procedure performs well in identifying the correct model order under various circumstances.
As an empirical application, we consider electricity load data from the state of California. Electricity loads have frequently been discussed in the forecasting literature, cf. for example Ramanathan et al. (1997), Soares and Souza (2006) or Weron and Misiorek (2008). Using our procedure shows that the series is driven by a complex 14-factor Gegenbauer model, with poles at weekly and at daily harmonics with different memory parameters.

Chapter 4 provides a result on the memory of products of long memory time series. It turns out that the persistence of products of time series critically depends on the means of the processes. On the one hand, if both factor series have non-zero means, the memory of the product is determined by the maximal memory of the factor series. On the other hand, if both series are mean zero, the memory of the product is reduced. If only one of the series has a zero-mean, it is the memory of this series that determines that of the product series. Corollaries are derived, that characterize the memory of squared processes and products of fractionally cointegrated series.

Finally, Chapter 5 extends the popular Diebold-Mariano test to situations in which the forecast error loss differential exhibits long memory. Using the results from Chapter 4, it is shown that this situation can arise frequently, since long memory can be transmitted from forecasts and the forecast objective to forecast error loss differentials. The nature of this transmission mainly depends on the (un)biasedness of the forecasts and whether the involved series share common long memory. Further results show that the conventional Diebold-Mariano test is invalidated under these circumstances. The reason is that the presence of long memory can

complicate inference on the sample mean since the convergence rate is reduced to $T^{1/2-d}$ and the autocovariance function is no longer absolutely summable. Therefore, standard estimators of the long run variance (such as HAC) do not work without modifications.

We then consider two robust statistics based on a memory and autocorrelation consistent estimator of Robinson (2005) and the extended fixed-bandwidth approach of McElroy and Politis (2012). A Monte Carlo study is conducted to provide a comparison of these robust statistics. It is found that tests based on the MAC estimator have better power properties, whereas tests using extended fixed-bandwidth asymptotics allow for a better size control - especially if the modified quadratic spectral kernel is used.

As an empirical application, we conduct forecast comparison tests for the realized volatility of the S&P 500 index among recent extensions of the heterogeneous autoregressive model of Corsi (2009). We find that forecasts improve significantly if jumps in the log-price process are considered separately from continuous components. Improvements achieved by the inclusion of implied volatility turn out to be insignificant.

Chapter 2

A Multivariate Test Against Spurious Long Memory

A Multivariate Test Against Spurious Long Memory

Co-authored with Philipp Sibbertsen and Marie Busch.
Under revision for the Journal of Econometrics.

2.1 Introduction

Distinguishing between true and spurious long memory is of major importance for the empirical modeling of many macroeconomic and financial time series. Usually, fractionally integrated processes are used to model long memory. Nevertheless, several authors point out that other data generating processes can have similar autocovariance features. Examples of this literature include Granger and Ding (1996), who argue that time varying coefficient models and other non-linear models can generate spurious long memory or Diebold and Inoue (2001) who derive parameter constellations for which random level shift processes, STOPBREAK models, and markov switching models can generate spurious long memory and support these arguments with extensive simulation studies. Similar results are obtained by Granger and Hyung (2004), who also report that the evidence for long memory in absolute returns of the S&P 500 vanishes if breaks are accounted for. Further contributions that give similar evidence are Lobato and Savin (1998) and Mikosch and Stărică (2004), among others.

Motivated by these findings, several tests have been proposed to distinguish true long memory from spurious long memory. Berkes et al. (2006) or Yau and Davis (2012), among others, suggest tests for the null hypothesis of spurious long memory. Here, we focus on the literature with the null hypothesis of true long memory. Dolado et al. (2005) propose a time domain test based on the testing principle of previously derived fractional Dickey Fuller tests. Shimotsu (2006) suggests two tests. The first one is based on the observation that for a true fractionally integrated process the memory parameter d must be the same across all subsamples. On the other hand, the implied d of a spurious long-memory process depends on the number of shifts and their location in the sample. Splitting the sample into subsamples, therefore, changes the implied d in every subsample. Shimotsu's second approach is to test whether the d-th difference of a process is $I(0)$ or its partial sum $I(1)$, using KPSS and Phillips-Perron tests. Another property of fractionally integrated processes is that the parameter d remains unchanged if the process is temporally aggregated. Ohanissian et al. (2008) use this property to test for true long memory by testing the equality of ds estimated from different aggregation levels of the same process. A similar test based on periodic sub-sampling, also known as skip-sampling, is presented by Davidson and Rambaccussing (2015). Their testing procedure compares the memory estimator of the skip-sampled data with the estimated d from the original data. Haldrup and Kruse (2014) propose a test based on the fact that nonlinear transformations of an $I(d)$ process will reduce the order of memory when the Hermite rank of the transformation is greater than one.

Perron and Qu (2010) derive the properties of the periodogram of processes with short memory and level shifts. They find that for low frequencies the effect of the shifts dominates the behavior of the spectral density and the implied value of d is one. For larger frequencies, on the other

hand, the short memory component is dominant and the implied d is zero. These findings explain the sensitivity of semiparametric d-estimators with respect to the bandwidth choice. Therefore, Perron and Qu (2010) propose a test statistic based on the difference between memory parameters estimated with different bandwidths. The same results on the spectral density of level shift processes are used by Qu (2011), who derives a score-type test that is based on the derivative of the local Whittle likelihood function. Simulation studies conducted by Qu (2011) and Leccadito et al. (2015) show that among the tests suggested so far, overall the Qu test has the best power against a wide range of alternatives.
All of these approaches discussed so far are univariate. There is, however, a strand of literature that has considered multivariate extensions of fractionally integrated processes. Early examples include the work of Sowell (1989) and Lobato (1997). In particular, Lobato (1999) and Shimotsu (2007) extend the local Whittle estimator to a multivariate framework. Extensions of the local Whittle estimator to fractionally cointegrated systems have been considered by Nielsen (2007a), Robinson (2008) and Shimotsu (2012).

We contribute to this literature by generalizing the approach of Qu (2011) to test for true long memory in multivariate processes. The test statistic is based on the weighted sum of the partial derivatives of the multivariate local Whittle likelihood function in the form introduced by Shimotsu (2007). In this specification the cross-spectral densities contain information on the phase and coherence of the process. As Kechagias and Pipiras (2015) show, the assumed form of the spectral density matrix local to the origin is specific to causal filters with hyperbolically decaying coefficients. Therefore, our test can be interpreted as a general test on the correct specification of a multivariate series as a causal long memory process. If one is willing to assume that the process is causal, a rejection of the test can be interpreted as evidence for low frequency contaminations.
The limit distribution of the test statistic is derived for general weights. However, by choosing the weighting scheme accordingly, one can obtain a pivotal distribution that coincides with that of the univariate Qu test. Furthermore, it is also possible to choose the weights so one can gain further insights into which components of a multivariate process cause a rejection.
To our knowledge, this is the first multivariate test against spurious long memory. The idea behind the test is that under the null hypothesis the derivative of the local Whittle likelihood function evaluated at $\hat{d}$ for the first $\lfloor mr \rfloor < m$ Fourier frequencies with $r \in [\varepsilon, 1]$ is approximately equal to zero. Under the alternative the derivative diverges if it is evaluated for a lower number of Fourier frequencies than used for the estimation of d, since it is based on a wrong assumption about the shape of the spectral density.
Our test statistic is derived in a multivariate long memory framework which excludes fractional cointegration. Nevertheless, we show that the test can easily be modified for the situation of fractionally cointegrated data. This modified test statistic has the same asymptotic properties as the original test statistic, including the same limiting distribution.
In the empirical example we apply our test to the log-absolute returns and realized volatilities

of four stock market indices, the Standard & Poor 500, the DAX, the FTSE and the NIKKEI. Even though especially the log-absolute values of S&P 500 returns have been studied in many of the aforementioned contributions on the possibility of spurious long memory, the tests proposed so far often fail to reject the null hypothesis of a true long-memory process. We, therefore, reconsider this example by extending it to a multivariate framework and we can clearly reject the null hypothesis of a pure long memory process for the S&P 500. For realized volatility series of these four stock market indices on the other hand, we do not find any evidence of spurious long memory.

The rest of the paper is structured as follows. After stating the model and the assumptions in Section 2.2, the test statistic is derived in Section 2.3. Some Monte Carlo simulations are given in Section 2.4. The empirical application is presented in Section 2.5 and Section 2.6 concludes. Proofs can be found in the appendix and additional results are provided in a supplementary appendix.

2.2 Model Specification and Assumptions

The spectral density of a multivariate long-memory process X_t, with $d = (d_1, d_2, ..., d_q)'$ and $-1/2 < d_1, \dots, d_q < 1/2$ being the memory vector, is local to the origin given by

$$f(\lambda_j) \sim \Lambda_j(d) G \Lambda_j^*(d), \tag{2.1}$$

with $\Lambda_j(d) = diag(\Lambda_{ja}(d))$ and $\Lambda_{ja}(d) = \lambda_j^{-d_a} e^{i(\pi - \lambda_j)d_a/2}$, where $\lambda_j = 2\pi j/T$ denotes the j-th Fourier frequency, and $j = 1, \dots, \lfloor T/2 \rfloor$. G is a real, positive definite, symmetric and finite matrix and the asterix A^* denotes the conjugate transpose of the matrix A. Further, the imaginary number is denoted by i and d_a is the memory parameter in dimension a.

The assumptions on G exclude fractional cointegration as they stand. We first derive our test statistic under this assumption and consider the case of fractionally cointegrated series afterwards. It turns out that the asymptotic properties of the test statistic remain unchanged if the test is slightly modified to accommodate fractional cointegration.

The spectral density representation in (2.1) accounts for phase shifts in the spectrum. Phase shifts occur as the covariance function $\gamma(h)$ of the process is no longer necessarily time-reversible in the multivariate setting, that is $\gamma(h) \neq \gamma(-h)$. Therefore, the off-diagonal elements of the spectral matrix of X_t contain complex valued elements which are not vanishing even at $\lambda = 0$ and which depend on the memory parameter d_a. These complex valued elements vanish if and only if the matrix G in (2.1) is diagonal or $d_a = d$ for all dimensions a.

One possible example is the multivariate q-dimensional FIVARMA model

$$\begin{pmatrix} (1-L)^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-L)^{d_q} \end{pmatrix} \begin{pmatrix} X_{1t} - EX_{1t} \\ \vdots \\ X_{qt} - EX_{qt} \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{qt} \end{pmatrix},$$

with $-1/2 < d_1, \ldots, d_q < 1/2$ and $t = 1, \ldots, T$. This can alternatively be written as

$$D(d_1, ..., d_q)(X_t - EX_t) = u_t, \tag{2.2}$$

where X_t is a $(q \times 1)$ column vector and $u_t = (u_{1t}, u_{2t}, ..., u_{qt})'$ is a covariance stationary process with spectral density $f_u(\lambda)$ which is bounded and bounded away from zero in a matrix sense at the zero frequency, $\lambda = 0$. The operator $D(d_1, ..., d_q) = diag((1-L)^{d_1}, ..., (1-L)^{d_q})$ is a $(q \times q)$ matrix polynomial with zeros on the non-diagonal elements.
In a univariate framework a type II fractionally integrated process (e.g., Marinucci and Robinson, 1999) is defined by $(1-L)^d x_t = u_t 1(t \geq 0)$, where u_t is an $I(0)$ process having the Wold representation $u_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}$ with $\sum_{j=0}^{\infty} \|\theta_j\|^2 < \infty$. The innovations ϵ_t are assumed to be a martingale difference sequence satisfying $E(\epsilon_t | \mathfrak{F}_{t-1}) = 0$ and $E(\epsilon_t^2 | \mathfrak{F}_{t-1}) < \infty$ with $\mathfrak{F}_t = \sigma(\{\epsilon_s, s \leq t\})$. Furthermore, it is $u_t = 0$ for $t \leq 0$. The order of fractional integration is given by d and $(1-L)^d$ is defined by its binomial expansion

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

with $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$. L denotes the Backshift operator, i.e. $Le_t = e_{t-1}$. Details about recent developments on long-memory time series can be found in Beran et al. (2013) or Giraitis et al. (2012).

The spectral density of the process u_t in (2.2) is assumed to fulfill the local condition

$$f_u(\lambda) \sim G, \qquad \lambda \to 0.$$

This condition is fulfilled whenever u_t has the Wold decomposition $u_t = C(L)\varepsilon_t$, where $C(1)$ is finite and has full rank, and $C(L)$ is a polynomial in the lag operator with absolute summable weights.
Furthermore, define the periodogram of X_t evaluated at frequency λ as

$$I(\lambda) = w(\lambda) w^*(\lambda), \qquad \text{with} \qquad w(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t e^{it\lambda}.$$

In the rest of the paper the superscript 0 denotes the true value of a parameter, for example d^0 is the true memory parameter.
We need to state the following assumptions which follow those in Shimotsu (2007):

Assumption 2.1. *For $\beta \in (0, 2]$ and $a, b = 1, \ldots, q$ as $\lambda \to 0+$*

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

Here and in the following f_{ab} and G_{ab} are the respective elements of the matrices f and G.

Assumption 2.2. *It holds that*

$$X_t - EX_t = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j},$$

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* $K > 0$ *it is* $P(\|\varepsilon_t\|^2 > \tau) \leq KP(\varepsilon^2 > \tau)$*. 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 2.3. *In a neighborhood* $(0, \delta)$ *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 2.4. *As* $T \to \infty$ *it holds for any* $\gamma > 0$

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{T^{2\beta}} + \frac{\log T}{m^{\gamma}} \to 0,$$

where m *is the bandwidth parameter.*

Assumption 2.5. *There exists a finite real matrix* Q *such that*

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

These assumptions are multivariate versions of the assumptions in Qu (2011). They allow for non-Gaussianity. Assumption 2.1 and 2.5 are satisfied by multivariate FIVARMA processes. Assumption 2.4 is slightly stronger than the assumption used in Qu (2011) for the univariate local Whittle estimator. However, this stronger assumption is necessary for the Hessian of the objective function of the local Whittle estimator to converge, which is needed in our proof. It should be mentioned that Assumption 2.4 gives a sharp upper bound for the number of frequencies m which can be used for the local Whittle estimator and thus for our test statistic. It is $m = o(T^{0.8})$.

2.3 Testing for Spurious Long Memory

In this section we consider multivariate testing for pure causal long memory. Our test is spectral based and investigates whether the data under consideration can be well described by the spectral density (2.1). However, it is motivated by using the different properties of the periodogram of long-memory processes and processes with structural breaks or trends. Special use will be made from the fact that the slope of the spectral density of a process with structural breaks is nearly zero for $j > \sqrt{T}$.

2.3.1 The MLWS Statistic

To be specific, we are interested in testing the hypothesis that the spectral density local to the origin has the shape given in equation (2.1):

$$H_0 : f(\lambda_j) \sim \Lambda_j(d) G \Lambda_j^*(d)$$

as $\lambda_j \to 0+$ with $d_a \in (-1/2, 1/2)\ \forall\ a = 1, \ldots, q$. Thus, under the null hypothesis X_t is a multivariate causal long-memory process with phase $(d_a - d_b)\pi/2$. The alternative is that the data cannot be described by this spectral density:

$$H_1 : f(\lambda_j) \nsim \Lambda_j(d) G \Lambda_j^*(d).$$

To motivate our test statistic, we focus on considering the properties of the periodogram under the alternative of low frequency contaminations. We consider two alternative models. The first one is a multivariate random level shift model defined by

$$\begin{aligned} X_t &= \mu_t + \kappa_t \quad \text{with} \\ \mu_t &= (I_q - \phi \Pi_t)\mu_{t-1} + \Pi_t e_t, \end{aligned} \tag{2.3}$$

where κ_t, $\Pi_t = diag(\pi_{1t}, \ldots, \pi_{qt})$ and e_t are mutually independent. The Bernoulli variables π_{it} and π_{jt} for the different dimensions of the q-dimensional process X_t are correlated with correlation matrix Σ_π for $i, j = 1, ..., q$. We consider a shift probability that is defined by $p = \tilde{p}/T$, where $\tilde{p}$ is the expected number of shifts. Furthermore, the magnitude of the shifts is characterized by the q-dimensional column vector e_t, with $e_t \sim N(0, \Sigma_e)$, and the noise process κ_t, $\kappa_t \sim N(0, \Sigma_\kappa)$. The pairwise correlation coefficients of π_{it} and π_{jt}, e_{it} and e_{jt}, and κ_{it} and κ_{jt} are labeled as $\rho_{\pi,ij}$, $\rho_{e,ij}$ and $\rho_{\kappa,ij}$, $\forall\ i, j = 1, ..., q$.

The autoregressive coefficient $0 \leq \phi \leq 1$ determines the persistence of the level shifts. This allows us to consider stationary as well as non-stationary multivariate random level shift processes. This formulation of our random level shift model is a multivariate version of the autoregressive random level shift process suggested in Xu and Perron (2014).

The second model under the alternative is the smooth trend model:

$$X_t = H\left(\frac{t}{T}\right) + \kappa_t, \tag{2.4}$$

where all variables are q-dimensional column vectors, $H(t/T) = (h_1(t/T), \ldots, h_q(t/T))'$ and $h_i(t/T)$ is a Lipschitz continuous function on $[0,1]$, $\forall\ i = 1, ..., q$. The noise term κ_t is defined as in equation (2.3).

In analogy to Perron and Qu (2010), the periodogram of X_t in (2.3) or (2.4) can be decomposed in four components by

$$\begin{aligned} I_X(\lambda_j) &= \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mu_t\mu_s' \exp\left\{i(t-s)\lambda_j\right\} \\ &+ \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_t\kappa_s' \exp\left\{i(t-s)\lambda_j\right\} \\ &+ \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_t\mu_s' \exp\left\{i(t-s)\lambda_j\right\} \\ &+ \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mu_t\kappa_s' \exp\left\{i(t-s)\lambda_j\right\}. \end{aligned}$$

In contrast to the univariate case, we have four terms instead of three, due to the asymmetry of the autocovariance matrices. By similar arguments as in Proposition 3 of Perron and Qu (2010) for $\lambda_j = o(1)$ the first summand is of order $O_P(1)$, the second is of order $O_P(T^{-1}\lambda_j^{-2})$, and the third and fourth term are of order $O_P(T^{-1/2}\lambda_j^{-1})$. Therefore, for each component in X_t the level shifts affect the periodogram only up to $j = O(T^{1/2})$. The stochastic orders are exact in the case of level shifts as in equation (2.3) and approximate for slowly varying trends in (2.4).

This decomposition of the periodogram can now be used to construct a multivariate local Whittle score-type test (MLWS test). It is based on the difference between the spectral density of a fractionally integrated process and the periodogram of a series contaminated by mean shifts or smooth trends that is almost flat for frequencies $m > \sqrt{T}$. This property also explains why the bias of the estimate $\hat{d}$ of the memory parameter depends heavily on the bandwidth choice if a local semiparametric estimator is used.

The test statistic is based on the derivative of the local Whittle likelihood function evaluated at $\hat{d}$, where $\hat{d}$ is the local Whittle estimate obtained using the first m Fourier frequencies. Qu (2011) now evaluates the derivative of the local Whittle likelihood function at the first $\lfloor mr \rfloor$ Fourier frequencies, where $r \in [\varepsilon, 1]$ with $\varepsilon > 0$. For $r = 1$ the derivative is exactly zero and for smaller r the derivative should be close to zero as long as the estimate of d remains stable when the bandwidth is decreased. This is the case under the null hypothesis. If the alternative is true, the non-uniform behavior of the spectral density leads to a divergence of the derivative. The test statistic is obtained by taking the supremum of the derivative over all r.

Our test statistic extends this idea to the multivariate case. It is based on the weighted sum of

the partial derivatives of the multivariate local Whittle likelihood as defined in Shimotsu (2007), which is given in (2.1). As the Gaussian log-likelihood of X_t and G are real, the local Whittle likelihood localized to the origin can be written as

$$Q_m(G,d) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log\det\Lambda_j(d)G\Lambda_j^*(d) + tr\left[G^{-1}Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]\right]\right\}. \tag{2.5}$$

The first order condition with respect to G gives

$$G = \frac{1}{m}\sum_{j=1}^{m} Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right].$$

Substituting this into $Q_m(G,d)$ and

$$\begin{aligned}\log\det\Lambda_j(d) + \log\det\Lambda_j^*(d) &= \log\det\Lambda_j(d)\Lambda_j^*(d)\\ &= \log\det\left(diag\left(\lambda_j^{-2d_a}\right)\right)\\ &= -2\sum_{a=1}^{q} d_a\log\lambda_j\end{aligned}$$

gives the objective function of the multivariate Gaussian semiparametric estimate (GSE) of Shimotsu (2007):

$$R(d) = \log\det\hat{G}(d) - 2\sum_{a=1}^{q} d_a\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j \tag{2.6}$$

with

$$\hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m} Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right].$$

To state our test statistic, we need to introduce some algebra on the first derivative of the objective function $R(d)$ which is condensed in Lemma 2.1 below. Denote by $\eta = (\eta_1,\ldots,\eta_q)'$ a $(q\times 1)$ vector of real numbers and $\nu_j = log\lambda_j - 1/m\sum_{j=1}^{m} log\lambda_j$. Furthermore, set ${}_aG^{-1}$ to be the a-th row of G^{-1} and set i_a to be the $(q\times q)$ matrix with a one on the a-th diagonal element and zeros elsewhere. Additionally, M_a denotes the a-th column of the matrix M. Then, we can write:

Lemma 2.1. *Under Assumptions 2.1 to 2.5 we have*

$$\begin{aligned}\sum_{a=1}^{q}\eta_a\sqrt{m}\frac{\partial R(d)}{\partial d_a} &= \frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\nu_j\left({}_aG^{-1}Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]_a - 1\right)\\ &\quad + \frac{1}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}\,{}_aG^{-1}Im\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]_a + o_P(1)\end{aligned}$$

The right hand side of Lemma 2.1 is the main ingredient of our test statistic which is asymptoti-

cally equivalent to the weighted sum of the components of the gradient vector. The test statistic is given by:

$$MLWS = \frac{1}{2}\sup_{r\in[\varepsilon,1]}\Bigg\|\frac{2}{\sqrt{\sum_{j=1}^{m}\nu_j^2}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{[mr]}\nu_j\Big({}_aG^{-1}(\hat{d})\Big[Re\Big[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\Big]\Big]_a-1\Big) \tag{2.7}$$
$$+\frac{1}{\sqrt{\sum_{j=1}^{m}\nu_j^2}}\sum_{a=1}^{q}\eta_a\Big({}_aG^{-1}(\hat{d})\Big)\sum_{j=1}^{[mr]}\frac{\lambda_j-\pi}{2}Im\Big[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\Big]_a\Bigg\|.$$

Remark 1: The factor 1/2 is added in order to obtain comparability with the univariate case.
Remark 2: As usual, a small sample correction is applied by replacing $m^{-1/2}$ with $(\sum_{j=1}^{m}\nu_j^2)^{-1/2}$ which improves the size of the test and is asymptotically equivalent.

In the univariate case our test reduces exactly to that of Qu (2011). The imaginary part in our test statistic accounts for the phase shifts in the multivariate spectrum. By combining the results of Shimotsu (2007) with those of Qu (2011) we are able to derive the limiting distribution of the test statistic (2.7). It is stated in the following theorem, where B(s) denotes standard one-dimensional Brownian motion, $\odot$ is the Hadamard product and $\Rightarrow$ denotes weak convergence:

Theorem 2.1. *Under Assumptions 2.1 to 2.5 and denoting* $i=\sqrt{-1}$ *we have for* $T\to\infty$

$$\begin{aligned} MLWS \Rightarrow\ & \frac{1}{2}\sup_{r\in[\epsilon,1]}\Bigg\|\int_0^r\Bigg[(1+\log s)\Big(2\eta'\eta+2\eta'\Big(G^0\odot\Big(G^0\Big)^{-1}\Big)\eta\Big)^{1/2} \\ & +i\Bigg[\frac{\pi^2}{2}\Big(\eta'\Big(G^0\odot\Big(G^0\Big)^{-1}\Big)\eta-\eta'\eta\Big)\Bigg]^{1/2}\Bigg]dB(s) \\ & -2\eta'B(1)\int_0^r(1+\log s)ds\eta \\ & -\int_0^1\Bigg[(1+\log s)\Big(2\eta'F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}+I_q)\Omega^{-1'}F(r)'\eta\Big)^{1/2} \\ & +i\Big(\frac{\pi^2}{2}\eta'F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}-I_q)\Omega^{-1'}F(r)'\eta\Big)^{1/2}\Bigg]dB(s)\Bigg\|, \end{aligned} \tag{2.8}$$

where $\Omega=2\Big[G^0\odot(G^0)^{-1}+I_q+\frac{\pi^2}{4}(G^0\odot\Big(G^0\Big)^{-1}-I_q)\Big]$ *and*

$$F(r)=2\int_0^r\Bigg[(1+\log s)^2\Big(G^0\odot\Big(G^0\Big)^{-1}+I_q\Big)+\frac{\pi^2}{4}\Big(G^0\odot\Big(G^0\Big)^{-1}-I_q\Big)\Bigg]ds.$$

The test statistic as it stands and its limiting distribution in Theorem 2.1 hold for any choice of the weight vector η. However, the test statistic is not pivotal as the limiting distribution depends on G^0 and thus on the unknown memory parameter d^0. Furthermore, the limiting distribution depends on the dimension q.

To overcome this problem, we fix the weighting scheme η to $\eta_a = 1/\sqrt{q}$, $\forall a = 1,\ldots,q$, to obtain a pivotal test independent of the unknown parameter d^0. Furthermore, this choice guarantees that for every dimension q the limiting distribution is exactly the same as in Qu (2011). This is stated in the following lemma:

Lemma 2.2. *Under Assumptions 2.1 to 2.5 and setting $\eta_1 = \ldots = \eta_q = 1/\sqrt{q}$ we have for $T \to \infty$*

$$\begin{aligned} MLWS \quad \Rightarrow \quad & \sup_{r\in[\epsilon,1]} \Bigg\| \int_0^r (1+\log s) dB(s) \\ - \quad & B(1) \int_0^r (1+\log s) ds \\ - \quad & F(r) \int_0^1 (1+\log s) dB(s) \Bigg\|, \end{aligned}$$

where

$$F(r) = \int_0^r (1+\log s)^2 ds.$$

Remark 3: For $\varepsilon = 0.02$, the asymptotic critical values of the MLWS test are given by 1.118, 1.252, 1.374, and 1.517 for a 10%, 5%, 2.5%, and 1% significance level respectively. The corresponding critical values for a larger trimming parameter, $\varepsilon = 0.05$, equal 1.022, 1.155, 1.277, and 1.426, as shown by Qu (2011).

After deriving the limiting distribution of the test, we have to prove its consistency under the alternatives (2.3) and (2.4). This is done in the following theorem:

Theorem 2.2. *Suppose that the process X_t is generated by (2.3) or (2.4) with at least one non-constant mean. Furthermore, assume that $m/T^{1/2} \to \infty$, $P(\hat{d}_j - d_j^0 > \epsilon) \to 1$, $P(m^{-1}\sum_{j=1}^m I_j(\lambda)\lambda_j^{2\hat{d}} > \epsilon) \to 1$ for each j, as $T \to \infty$ with ϵ being some arbitrary small constant, and assume that Assumptions 2.1 to 2.5 hold. Then, MLWS $\xrightarrow{p} \infty$, as $T \to \infty$.*

Note that (2.3) or (2.4) nest the cases, where only a subvector of X_t is subject to low frequency contaminations. Theorem 2.2 therefore does not assume, that all components of X_t are affected.

2.3.2 Testing for Low Frequency Contaminations in a Component of a Multivariate System

A rejection of the MLWS statistic indicates misspecifications in at least one of the components of the process. To gain further insights into which of the components of X_t cause the rejection, one can use the limit distribution derived in Theorem 2.1 to test the hypothesis

$$H_0(a): \; S(a) \odot f(\lambda_j) \sim S(a) \odot \Big(\Lambda_j(d) G \Lambda_j^*(d)\Big), \tag{2.9}$$

as $\lambda_j \to 0$, where $S(a)$ is a selection matrix with ones in its a-th row and a-th column and zeros in all other elements. Such a test is obtained by setting the a-th element of η to one and all others to zero.
In this case the limit distribution simplifies slightly to

$$\begin{aligned} MLWS \quad \Rightarrow \quad & \frac{1}{2}\sup_{r\in[\epsilon,1]}\left\| \int_0^r \left[\sqrt{2}(1+\log s)\left(\frac{g_{aa}\det(G^0_{-aa})}{\det(G^0)}+1\right)^{1/2} \right.\right. \\ & \left.\left. +i\left[\frac{\pi^2}{2}\left(\frac{g_{aa}\det(G^0_{-aa})}{\det(G^0)}-1\right)\right]^{1/2}\right] dB(s) \\ - \quad & 2B(1)\int_0^r (1+\log s)ds \\ - \quad & \int_0^1 \left[(1+\log s)\left(2F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}+I_q)\Omega^{-1'}F(r)'\right)_{aa}^{1/2} \right. \\ & \left.\left. +i\left(\frac{\pi^2}{2}F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}-I_q)\Omega^{-1'}F(r)'\right)_{aa}^{1/2}\right] dB(s)\right\|. \end{aligned} \tag{2.10}$$

Since the distribution depends on G^0, it is not pivotal and the implementation of the test statistic requires the simulation of critical values for each $\hat{G}(\hat{d})$.
Under the null hypothesis the a-th row and column of the spectral density matrix correspond to those of a multivariate long memory process. In case of a low frequency contamination in component $b \neq a$, only one of the off-diagonal elements in the a-th row and the a-th column is affected, whereas all elements in the b-th row and column differ from the null hypothesis.
A rejection of $H_0(a)$ might therefore be due to a low frequency contamination in the b-th component. However, a non-rejection of $H_0(a)$ and a rejection of $H_0(b)$ can be interpreted as evidence for a contamination in component b only. Furthermore, the ordering of the p-values of the test statistics can be used as an indication for the relative probability of a contamination in the respective components.

2.3.3 MLWS Test for Fractionally Cointegrated Series

So far, fractional cointegration has been ruled out by our assumptions on the matrix G, which has reduced rank if components of X_t are cointegrated. However, our test can be robustified against fractional cointegration without altering the limiting distribution. Let there be p_G cointegrating relationships between the components of X_t, where $1 \leq p_g < q$, and assume without loss of generality that these involve the first p_G components of X_t. Then $rank(G) = q - p_G$, the memory order of the first p_G components is $d_1 = d_2 = ... = d_{p_G}$, and there exists a cointegrating matrix B, such that

$$BX_t = w_t,$$

and w_t has a spectral density matrix as specified in (2.1), but with $\tilde{G} = BGB'$ and $\tilde{d} = (d_{p_G} - b_1, ..., d_{p_G} - b_{p_G}, d_{p_G+1}, ..., d_q)'$ instead of G and d, for $0 < b_a \leq d_a$. The matrix B is such that the first p_G elements of X_t are replaced by the cointegrating residual series so that $\tilde{G}$ has full rank. This is achieved, if the first p_G rows of B contain the cointegrating vectors, normalized so that the diagonal elements of B equal 1 and the remaining rows contain zeros on all the off-diagonal elements. In the bivariate case, for example, B takes the form

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

Consequently, the MLWS test for the hypothesis $H_0: \; f_{BX_t}(\lambda_j) \sim \Lambda_j(d)\tilde{G}'\Lambda_j^*(d)$, as $\lambda_j \to 0+$ can simply be carried out on the transformed series BX_t, if the cointegrating matrix B is known. To obtain a feasible procedure for unknown B, a consistent estimator for B has to be applied, that converges with a faster rate than $\sqrt{m}$, where m is the bandwidth used for the MLWS statistic. If a multivariate local Whittle estimator $\hat{B}_{MLW}$ is used, such as those of Robinson (2008) or Shimotsu (2012), this is equivalent to constructing the test statistic as in (2.7), but using the concentrated local Whittle likelihood of the cointegrated system. The resulting test statistic is

$$\begin{aligned} \widetilde{MLWS} \;=\; & \frac{1}{2} \sup_{r\in[\varepsilon,1]} \Bigg\| \frac{2}{\sqrt{\sum_{j=1}^m \nu_j^2}} \sum_{a=1}^q \eta_a \sum_{j=1}^{[mr]} \nu_j \Big({}_a\tilde{G}^{-1}(\hat{d},\hat{B}) \Big[Re\Big[\Lambda_j(\hat{d})^{-1} \tilde{I}(\lambda_j,\hat{B}) \Lambda_j^*(\hat{d})^{-1} \Big] \Big]_a - 1 \Big) \\ & + \frac{1}{\sqrt{\sum_{j=1}^m \nu_j^2}} \sum_{a=1}^q \eta_a \Big({}_a\tilde{G}^{-1}(\hat{d},\hat{B}) \Big) \sum_{j=1}^{[mr]} \frac{(\lambda_j - \pi)}{2} Im\Big[\Lambda_j(\hat{d})^{-1} \tilde{I}(\lambda_j,\hat{B}) \Lambda_j^*(\hat{d})^{-1} \Big]_a \Bigg\|, \end{aligned}$$

where $\tilde{G}(d,B) = \frac{1}{m}\sum_{j=1}^m Re\Big[\Lambda_j(d)^{-1} BI(\lambda_j) B' \Lambda_j^*(d)^{-1}\Big]$ and $\tilde{I}(\lambda_j, B) = BI(\lambda_j)B'$. As Robinson (2008) shows for the bivariate example, $\hat{B}_{MLW}$ converges to B^0 with rate $\sqrt{m}\Delta_m$, where $\Delta_m = diag(\lambda_m^{-b_1}, ..., \lambda_m^{-b_{p_G}}, 1, ..., 1)$. Since this is faster than $\sqrt{m}$, we can state the following lemma for our test under fractional cointegration:

Lemma 2.3. *Let Assumption 2.1 to 2.5 hold. Then the test statistic $\widetilde{MLWS}$ has the same null limiting distribution as MLWS in Theorem 2.1, but with G^0 replaced by $\tilde{G} = B^0 G^0 B^{0\prime}$.*

Consequently, if fractional cointegration is present, the cointegrated variables can be removed from X_t and replaced by the cointegration residuals w_t, without altering the limit distribution (apart from G^0).

2.3.4 Pre-Whitening

Although the MLWS statistic in (2.7) is asymptotically independent of short memory dynamics, we need to apply a pre-whitening procedure to avoid negative effects on the size and power properties of the test in finite samples if short memory dynamics are present. Similar to Qu

(2011), we do so by approximating the short memory dynamics of u_t in (2.2) with a low order VARMA process that is given by

$$u_t = \mathcal{A}(L)^{-1}\mathcal{M}(L)v_t,$$

where $\mathcal{A}(L) = I_q - \mathcal{A}_1 L - ... - \mathcal{A}_{p_A} L^{p_A}$ and $\mathcal{M}(L) = I_q - \mathcal{M}_1 L - ... - \mathcal{M}_{q_M} L^{q_M}$ are matrix lag polynomials and v_t is a $(q \times 1)$ white noise with zero mean and variance-covariance matrix Σ_v.
Since $\mathcal{A}(L)$ and $D(d_1,\dots,d_q)$ are not commutative in the multivariate case, there are two ways to generalize the univariate ARFIMA process to a vector process. This has been pointed out by Lobato (1997). For the specification used here, Sela and Hurvich (2009) coined the acronym FIVARMA because the process $X_t = D(d_1,\dots,d_q)^{-1}\mathcal{A}(L)^{-1}\mathcal{M}(L)v_t$ can be interpreted as a fractionally integrated VARMA process. The process $X_t = \mathcal{A}(L)^{-1}D(d_1,\dots,d_q)^{-1}\mathcal{M}(L)v_t$ is a vector autoregression of a fractionally integrated vector moving average process and is referred to as VARFIMA. The FIVARMA model has recently been applied in Chiriac and Voev (2011) and has first been studied by Sowell (1989). As Lütkepohl (2007) points out, the parameters of the unrestricted VARMA model are not identified. Thus, we follow Chiriac and Voev (2011) and estimate the model in its final equation form, where $\mathcal{A}(L)$ is a scalar operator.
The estimation is carried out using the conditional sum of squares estimator (cf. Beran (1995), Hualde et al. (2011) and Nielsen (2015)) which is based on an approximation of the AR(∞) representation of the FIVARMA process.[1]
Before we can introduce our pre-whitening procedure we have to sharpen Assumption 2.2. This is necessary as our test is then applied to the filtered rather than the original series. The sharpened assumption is a multivariate version of Assumption F in Qu (2011).

Assumption 2.6. *Assume that in addition to Assumption 2.2* $A_j = O(j^{-1/2-c})$ *with* $c > 0$ *as* $j \to \infty$.

Since the short memory dynamics can be approximated well with a low order model and the estimation of FIVARMA models is computationally very demanding, we restrict the model order to be $p_A = q_M = 1$. In analogy to Qu (2011), we then apply the filter $\hat{\mathcal{A}}(L)^{-1}\hat{\mathcal{M}}(L)(X_t - EX_t) = \tilde{X}_t$ to the original series X_t. To test for spurious long memory we subsequently apply the MLWS test in (2.7) to the filtered series $\tilde{X}_t$.

Lemma 2.4. *Assume that* X_t *satisfies Assumptions 2.1 to 2.5 and 2.6. Then, the MLWS test applied on the filtered series* $\tilde{X}_t$ *has the same limiting distribution as given in Theorem 2.1.*

Note that we do not assume that the short memory dynamics follow a VARMA(1,1)-process. We use it as a reasonable approximation to the true short memory dynamics in small samples. Asymptotically the test is unaffected by any form of short memory dependence because we only

[1] For the sample sizes considered here, this approach turns out to be faster than the method of Sela and Hurvich (2009). For larger samples the computing time can be further reduced by conducting the ML estimation on subsamples and using the average of these subsample estimators, as suggested by Beran and Terrin (1994).

use the periodogram ordinates at Fourier frequencies local to the pole. The short memory dynamics have no influence on the shape of the pole. This is also why the pre-whitening procedure leaves the limiting distribution of the test unaffected.

2.4 Monte Carlo Study

To analyze the finite sample properties of the MLWS test, we conduct a Monte Carlo analysis that consists of three parts. In the first part, we consider a bivariate setup and conduct experiments to determine the influence of the bandwidth choice, $m = \lfloor T^{\delta} \rfloor$, and the choice of the trimming parameter ε on the size and the power of the test. Then, we turn to higher dimensional applications to analyze how the size and power depend on the dimension q of the multivariate process. Finally, in the third part we analyze the performance of the test if short memory dynamics exist. To disentangle the performance of the multivariate test from that of the pre-whitening procedure, the latter is only applied in the last part when short memory dynamics are present.

The simulation studies of Qu (2011) and Leccadito et al. (2015) show that the Qu test has good power against a wide range of different alternatives, such as non-stationary random level shifts, smooth trends, markov switching models, or the STOPBREAK proces of Engle and Smith (1999). Therefore, we focus on analyzing the properties that are specific to the multivariate case and use a stationary random level shift process for all power DGPs. Further simulation studies covering other forms of deterministic trends and conditional heteroscedasticity are included in a supplementary appendix, available online. All results presented hereafter are based on $M = 5000$ Monte Carlo replications and all tests are carried out with a nominal significance level of $\alpha = 0.05$.

2.4.1 Size and Power Comparison in a Bivariate Setup

The size study for the bivariate case is based on the multivariate fractionally integrated process from equation (2.2), where the short memory component $u_t = v_t$ with $v_t \sim N(0, \Sigma_v)$ is specified to be a bivariate white noise

$$D(d_1, d_2)X_t = v_t.$$

In this setup we want to investigate two aspects. First, we evaluate whether the size depends on the correlation ρ_v between the components of the innovation vector v_t, or whether it depends on the (possibly different) degrees of memory d_1 and d_2 in the two series. Second, we want to determine the effect of the bandwidth m and the trimming parameter ε. Since the trimming parameter ε can be chosen discretionary, we follow Qu (2011) and conduct our simulations for $\varepsilon \in \{0.02, 0.05\}$.

Table 2.1 shows the results. We find that the test is generally conservative in finite samples - a feature which it shares with its univariate version. For all parameter constellations, the size is better with $\varepsilon = 0.05$ than with $\varepsilon = 0.02$ and it is increasing in m. The results also improve as the

			$\rho_v = -0.8$				$\rho_v = 0$				$\rho_v = 0.4$				$\rho_v = 0.8$			
		d_2	0		0.4		0		0.4		0		0.4		0		0.4	
T	d_1	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	0	**0.60**	0.006	0.007	0.007	0.012	0.005	0.010	0.006	0.009	0.005	0.008	0.007	0.008	0.004	0.008	0.007	0.007
		0.65	0.009	0.014	0.011	0.013	0.008	0.015	0.010	0.015	0.009	0.014	0.007	0.013	0.011	0.015	0.012	0.012
		0.70	0.012	0.017	0.012	0.015	0.011	0.016	0.014	0.018	0.013	0.017	0.011	0.016	0.010	0.021	0.011	0.019
		0.75	0.013	0.015	0.016	0.023	0.014	0.019	0.014	0.019	0.016	0.023	0.012	0.018	0.015	0.021	0.013	0.021
	0.4	**0.60**	0.007	0.010	0.008	0.008	0.007	0.009	0.007	0.010	0.005	0.008	0.006	0.009	0.008	0.008	0.006	0.009
		0.65	0.011	0.016	0.007	0.013	0.009	0.016	0.011	0.015	0.010	0.015	0.011	0.014	0.009	0.018	0.011	0.013
		0.70	0.013	0.014	0.017	0.021	0.014	0.018	0.011	0.017	0.012	0.016	0.014	0.019	0.015	0.017	0.012	0.018
		0.75	0.015	0.025	0.017	0.023	0.016	0.024	0.019	0.019	0.014	0.022	0.021	0.022	0.021	0.019	0.016	0.022
1000	0	**0.60**	0.013	0.019	0.015	0.02	0.013	0.021	0.014	0.020	0.014	0.022	0.014	0.021	0.014	0.021	0.017	0.022
		0.65	0.021	0.027	0.019	0.025	0.019	0.028	0.017	0.028	0.020	0.025	0.022	0.028	0.018	0.028	0.019	0.024
		0.70	0.023	0.028	0.031	0.032	0.024	0.029	0.021	0.027	0.025	0.029	0.025	0.029	0.023	0.029	0.025	0.031
		0.75	0.030	0.036	0.032	0.043	0.025	0.034	0.029	0.034	0.029	0.038	0.030	0.035	0.026	0.035	0.030	0.041
	0.4	**0.60**	0.016	0.020	0.015	0.023	0.013	0.019	0.015	0.022	0.014	0.018	0.014	0.020	0.015	0.022	0.015	0.019
		0.65	0.023	0.024	0.022	0.024	0.019	0.022	0.020	0.024	0.017	0.026	0.018	0.025	0.019	0.024	0.021	0.024
		0.70	0.024	0.037	0.022	0.031	0.024	0.029	0.020	0.028	0.022	0.032	0.023	0.031	0.024	0.032	0.023	0.035
		0.75	0.033	0.033	0.029	0.040	0.026	0.035	0.032	0.033	0.027	0.038	0.030	0.038	0.028	0.041	0.032	0.036
2000	0	**0.60**	0.022	0.031	0.023	0.027	0.022	0.025	0.021	0.028	0.021	0.033	0.016	0.027	0.028	0.030	0.023	0.028
		0.65	0.028	0.035	0.025	0.035	0.026	0.029	0.022	0.035	0.023	0.037	0.024	0.033	0.020	0.029	0.026	0.032
		0.70	0.029	0.029	0.031	0.038	0.025	0.034	0.026	0.031	0.026	0.035	0.024	0.035	0.028	0.033	0.030	0.034
		0.75	0.031	0.041	0.043	0.042	0.036	0.040	0.034	0.040	0.032	0.040	0.031	0.038	0.033	0.039	0.042	0.039
	0.4	**0.60**	0.021	0.034	0.022	0.027	0.018	0.025	0.021	0.026	0.020	0.032	0.018	0.028	0.021	0.028	0.018	0.029
		0.65	0.024	0.028	0.026	0.038	0.024	0.035	0.023	0.034	0.028	0.031	0.028	0.032	0.023	0.035	0.026	0.032
		0.70	0.035	0.033	0.028	0.033	0.028	0.033	0.030	0.033	0.028	0.034	0.027	0.034	0.026	0.034	0.031	0.028
		0.75	0.036	0.044	0.035	0.042	0.032	0.036	0.033	0.043	0.034	0.043	0.038	0.039	0.037	0.046	0.038	0.043

Table 2.1: Size of MLWS test for FIVARMA (0,d,0): $D(d_1,d_2)X_t = v_t$ with $v_t \sim N(0,\Sigma_v)$ and $\sigma_v^2 = 1$. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

sample size increases. With a sample size of $T = 2000$, $m = \lfloor T^{0.75} \rfloor$ and $\varepsilon = 0.05$ for example, we find that the size is between 3.6 and 4.6 percent for all combinations of ρ_v, d_1, and d_2. Thus, in larger samples the MLWS test achieves good size properties with the right choice of m and ε.
With regard to the correlation ρ_v between the innovations, the size tends to improve as the correlation increases, since the MLWS test makes use of the coherence information. Overall, even though the test is quite conservative in small samples, the size is good in larger samples and it is remarkably stable for different degrees of memory in the components of the series and correlations among the innovation sequences.
We will now turn to the effect of m and ε on the power of the test. In contrast to the true long-memory processes under the null hypothesis, that we denote by X_t, the DGP in the power study will be denoted by Y_t. Here, Y_t it is the sum of the white noise sequence v_t and the multivariate random level shift process μ_t from equation (2.3):

$$\begin{aligned} Y_t &= \mu_t + v_t \\ \mu_t &= (I_q - \phi\Pi_t)\mu_{t-1} + \Pi_t e_t. \end{aligned} \tag{2.11}$$

For $\phi = 1$ the process is stationary and for $\phi = 0$ it is non-stationary. The shift probability is

		Stationary ($\phi = 1$)						**Non-Stationary** ($\phi = 0$)					
	$\rho_\pi = \rho_e$	0		0.5		1		0		0.5		1	
T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.114	0.120	0.110	0.128	0.199	0.196	0.184	0.209	0.192	0.218	0.296	0.302
	0.65	0.264	0.308	0.259	0.299	0.350	0.377	0.395	0.441	0.407	0.465	0.485	0.502
	0.70	0.447	0.475	0.445	0.463	0.494	0.509	0.598	0.610	0.613	0.623	0.641	0.634
	0.75	0.632	0.640	0.638	0.645	0.634	0.633	0.766	0.764	0.764	0.768	0.743	0.730
1000	**0.60**	0.674	0.710	0.688	0.720	0.734	0.756	0.817	0.832	0.816	0.838	0.832	0.834
	0.65	0.897	0.910	0.898	0.908	0.872	0.873	0.952	0.949	0.942	0.948	0.918	0.919
	0.70	0.967	0.971	0.971	0.967	0.933	0.934	0.983	0.983	0.981	0.978	0.948	0.945
	0.75	0.990	0.988	0.986	0.985	0.952	0.952	0.991	0.992	0.988	0.986	0.960	0.956
2000	**0.60**	0.918	0.922	0.910	0.921	0.912	0.908	0.957	0.963	0.954	0.961	0.936	0.944
	0.65	0.984	0.984	0.982	0.982	0.959	0.958	0.991	0.993	0.988	0.986	0.963	0.964
	0.70	0.995	0.996	0.994	0.993	0.970	0.969	0.996	0.997	0.996	0.994	0.977	0.970
	0.75	0.998	0.996	0.995	0.994	0.976	0.974	0.999	0.998	0.995	0.994	0.979	0.977

Table 2.2: Power of MLWS test against stationary random level shifts: $Y_t = \mu_t + v_t$ with $v_t \sim N(0, \Sigma_v)$ and $\mu_t = (I_q - \phi\Pi_t)\mu_{t-1} + \Pi_t e_t$. The bandwidth m is determined by $m = \lfloor T^\delta \rfloor$.

always kept at $p = 5/T$, so that in expectation there are five shifts in every sample and the standard deviation of the shifts is $\sigma_e = 1$. Since a different behavior of the breaks can imply different coherence information, we consider different values for the correlation between the occurrence of shifts ρ_π and the correlation of the shift sizes ρ_e. For simplicity, we always set $\rho_\pi = \rho_e$. If $\rho_\pi = \rho_e = 0$ shifts occur independently in each of the components of the series, whereas shifts always coincide in timing and size if $\rho_\pi = \rho_e = 1$.[2]

The results of this experiment are shown in Table 2.2. We find that the power is always increasing in the bandwidth and it is higher against non-stationary level shifts. For small sample sizes with weakly correlated shifts the test has better power with $\varepsilon = 0.05$, but in larger samples $\varepsilon = 0.02$ leads to a higher power if m is also relatively large. With regard to the correlation of the shifts, we find that the power of the test increases in small samples if shifts show a stronger correlation. In large samples the power slightly decreases if shifts are perfectly correlated.

Overall, the test shows good size and power properties and for an increasing bandwidth both size and power improve. Note however that a larger bandwidth also makes the test more prone to errors if short memory dynamics are present. Therefore, we will address the choice of ε and m in practice later.

2.4.2 The Effect of Increasing Dimensionality

Since the proposed MLWS test is multivariate and its limiting distribution is independent of the dimension q of the process, we now consider how its finite sample properties depend on the

[2] Since the presence of spurious long memory depends on the location of the shifts in the sample, we discard all samples for which a test, for $H_0: d = 0$ based on the local Whittle estimate $\hat{d}_{LW}$, is not rejected for all components.

		Size							Power					
		MLWS			Qu				MLWS			Qu		
T	q/ρ_v	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**	$q/\rho_\pi,\rho_e$	**0**	**0.5**	**1**	**0**	**0.5**	**1**
100	**1**	0.011	0.010	0.011	0.014	0.013	0.012	**1**	0.098	0.093	0.095	0.093	0.098	0.094
	2	0.011	0.013	0.015	0.008	0.010	0.010	**2**	0.173	0.172	0.205	0.121	0.125	0.105
	3	0.013	0.015	0.011	0.007	0.007	0.007	**3**	0.243	0.243	0.296	0.125	0.118	0.111
	4	0.013	0.014	0.009	0.007	0.010	0.008	**4**	0.295	0.309	0.328	0.128	0.126	0.120
	5	0.011	0.010	0.011	0.006	0.006	0.007	**5**	0.356	0.357	0.366	0.135	0.133	0.118
250	**1**	0.020	0.020	0.019	0.017	0.022	0.023	**1**	0.411	0.393	0.392	0.402	0.415	0.414
	2	0.018	0.023	0.025	0.018	0.018	0.012	**2**	0.645	0.643	0.621	0.534	0.539	0.486
	3	0.020	0.022	0.020	0.013	0.012	0.014	**3**	0.801	0.784	0.712	0.617	0.610	0.506
	4	0.019	0.021	0.022	0.016	0.011	0.015	**4**	0.884	0.877	0.763	0.686	0.683	0.528
	5	0.020	0.017	0.023	0.012	0.013	0.009	**5**	0.946	0.930	0.773	0.729	0.714	0.559
500	**1**	0.027	0.026	0.025	0.027	0.027	0.026	**1**	0.751	0.742	0.752	0.742	0.743	0.747
	2	0.028	0.029	0.026	0.025	0.021	0.026	**2**	0.922	0.919	0.865	0.911	0.890	0.813
	3	0.026	0.028	0.029	0.018	0.021	0.021	**3**	0.979	0.973	0.906	0.963	0.951	0.834
	4	0.029	0.029	0.029	0.025	0.026	0.021	**4**	0.996	0.987	0.917	0.983	0.974	0.849
	5	0.026	0.031	0.028	0.021	0.023	0.022	**5**	0.999	0.994	0.923	0.994	0.987	0.857

Table 2.3: Size and power of MLWS test and repeated Qu test with Simes correction for increasing dimensions q. **Left panel**: Size for FIVARMA (0,d,0): $D(d_1,\ldots,d_q)X_t = v_t$. **Right panel**: Power for $Y_t = \mu_t + v_t$ with $v_t \sim N(0,\Sigma_v)$.

dimension q.

As before, our size DGP, $D(d_1,...,d_q)X_t = v_t$, is a fractionally integrated white noise. Motivated by our previous findings, we set $m = \lfloor T^{0.75} \rfloor$ and $\varepsilon = 0.05$ and consider only the effect of increasing the dimension q.

Since there is no other multivariate test against spurious long memory available in the literature, a practitioner has no other choice but to apply the Qu test to each of the q components of the process separately. We will use this approach as a benchmark procedure. To avoid Bonferroni errors, we apply the correction of Simes (1986) that consists in ordering the p-values in ascending order and then comparing them with α/q, $2\alpha/q$, ..., α. The null hypothesis is rejected if any of the ordered p-values exceeds its respective threshold.

Note that for $q = 1$ the MLWS test and the Qu test are identical. The left panel of Table 2.3 contains the results. We can observe that the MLWS test is quite conservative in small samples, but the size improves if the sample size increases. It also maintains approximately the same size independent of the dimension q and independent of the correlation among the components of v_t. For the repeated application of the Qu test we find that similar to the MLWS test it is conservative in small samples. In addition, the size tends to further decrease with increasing q and with increasing correlation ρ_v between the noise components, which is an effect of the Simes correction.

As in the bivariate setup, the power DGP, $Y_t = \mu_t + v_t$, is the sum of the q-dimensional white noise

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
DGP		1		2		3		4		1		2	
T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.006	0.007	0.006	0.009	0.004	0.007	0.010	0.008	0.007	0.009	0.005	0.010
	0.65	0.012	0.013	0.006	0.009	0.003	0.007	0.007	0.009	0.012	0.017	0.011	0.024
	0.70	0.010	0.013	0.005	0.006	0.002	0.004	0.013	0.013	0.018	0.017	0.037	0.072
	0.75	0.009	0.010	0.001	0.002	0.002	0.003	0.011	0.012	0.018	0.020	0.127	0.209
500	**0.60**	0.012	0.015	0.013	0.013	0.009	0.008	0.013	0.012	0.012	0.015	0.007	0.013
	0.65	0.013	0.015	0.015	0.016	0.009	0.011	0.011	0.012	0.014	0.025	0.017	0.038
	0.70	0.013	0.017	0.013	0.015	0.004	0.004	0.016	0.015	0.021	0.021	0.061	0.105
	0.75	0.013	0.013	0.006	0.006	0.002	0.003	0.010	0.013	0.023	0.030	0.238	0.340
1000	**0.60**	0.014	0.021	0.016	0.024	0.016	0.018	0.017	0.018	0.017	0.018	0.011	0.025
	0.65	0.021	0.019	0.021	0.024	0.015	0.017	0.021	0.023	0.023	0.029	0.023	0.033
	0.70	0.020	0.019	0.025	0.019	0.011	0.007	0.016	0.018	0.031	0.032	0.075	0.112
	0.75	0.018	0.018	0.014	0.011	0.003	0.004	0.019	0.016	0.039	0.038	0.351	0.452
2000	**0.60**	0.016	0.027	0.017	0.024	0.016	0.028	0.017	0.026	0.020	0.027	0.017	0.027
	0.65	0.019	0.030	0.023	0.035	0.022	0.029	0.023	0.034	0.023	0.031	0.023	0.039
	0.70	0.024	0.022	0.029	0.024	0.018	0.017	0.026	0.027	0.030	0.036	0.075	0.106
	0.75	0.028	0.022	0.025	0.020	0.009	0.006	0.020	0.024	0.043	0.040	0.433	0.535

Table 2.4: Size of the MLWS test for FIVARMA(1,d,1): $(1-\phi_1 L)D(d_1,d_2)X_t = (I_q - \mathcal{M}_1 L)v_t$ with and without pre-whitening. Parameters values of the respective DGPs are given above. The bandwidth m is determined by $m = \lfloor T^\delta \rfloor$.

v_t and the q-dimensional multivariate random level shift model from equation (2.11). Similar to the size DGP, we restrict the correlations of shifts in the components as well as the correlation of the shift sizes to be the same among all components such that $\rho_{\pi,ab} = \rho_{e,ab} = \rho_\pi = \rho_e$ for all $a \neq b$.

If we consider the results on the right hand side in Table 2.3, we find that there are indeed large power gains compared to the repeated application of the Qu test. For $T = 100$ these can be more than 24 percentage points. We find that the power is increasing in q and T. While correlated shifts increase the power in smaller samples, the power reduction observed in the bivariate simulations for correlated shifts in large samples increases with increasing q.

2.4.3 Short Memory Dynamics

So far we have considered the MLWS test applied directly to the observed series X_t. As discussed in Section 2.4, the performance of the local Whittle based methods can be negatively affected in finite samples if short memory dynamics are present. This is why we suggested a pre-whitening procedure based on the FIVARMA(1,d,1). Subsequently, the test is applied to the filtered series $\tilde{X}_t$. The performance of this procedure is analyzed in the following.

We consider four different DGPs for the size. These are given by:

$$\begin{aligned}
&\textbf{DGP1:} \quad D(0.2,0.3)X_t = v_t \\
&\textbf{DGP2:} \quad (1-0.4L)D(0.2,0.3)X_t = v_t \\
&\textbf{DGP3:} \quad (1-0.6L)D(0.2,0.3)X_t = v_t \\
&\textbf{DGP4:} \quad (1-0.4L)D(0.2,0.3)X_t = \left[I_2 - \begin{pmatrix} 0.5 & 0 \\ 0 & 0.3 \end{pmatrix} L\right] v_t.
\end{aligned}$$

For all processes we set $\sigma_v^2 = 1$, $\rho_v = 0.5$ and $p = 5/T$. All of these processes are special cases of the FIVARMA(1,d,1). DGP1 is a simple bivariate fractional white noise, while DGP2 and DGP3 contain autoregressive dynamics. Finally, DGP4 contains both autoregressive and moving average dynamics.

For the power studies we combine the respective size DGP X_t with the stationary multivariate random level shift μ_t:

$$\begin{aligned}
Y_t &= X_t + \mu_t \\
\mu_t &= (I_q - \Pi_t)\mu_{t-1} + \Pi_t e_t.
\end{aligned}$$

In the multivariate random level shift process we use $\rho_\pi = \rho_e = 0.9$ and $\sigma_e = 2$.

To investigate the costs and benefits of the pre-whitening procedure, the simulations for DGP1 and DGP2 are conducted with and without pre-whitening. The results for the size simulations are given in Table 2.4.

DGP1 is the baseline case. By comparing the results for the tests applied to the pre-whitened series $\tilde{X}_t$ with the results of the unfiltered series X_t, we can observe that the empirical size becomes a bit more conservative. However, for an increasing T the size approaches its nominal level. By considering DGP2, we see that there are considerable over-rejections if moderate autoregressive dynamics are present and no pre-whitening is applied, whereas with pre-whitening these distortions are successfully removed.

The results for DGP3 and DGP4 show that the pre-whitening procedure works well in controlling the size for different forms of short memory dynamics. As before, the size is generally better if $\varepsilon = 0.05$ is used. With regard to the bandwidth selection, the best size is observed most often for $\delta = 0.65$ or $\delta = 0.7$. It is no longer strictly increasing in m.

Table 2.5 considers the power results. For the baseline case there is a considerable power reduction caused by the additional flexibility introduced through the pre-whitening procedure. But when comparing the results for DGP1 and DGP2 without pre-whitening, we observe that the power also suffers severely if there are short memory dynamics but no pre-whitening is applied. Using the filtering procedure reduces this effect substantially. The power is increasing in T and generally also in m, but we observe some power drops for larger bandwidths, especially if the autoregressive dynamics become more persistent. Similarly to Section 2.4.1, in small samples

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
DGP		1		2		3		4		1		2	
T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.022	0.026	0.032	0.032	0.012	0.014	0.023	0.030	0.037	0.045	0.014	0.020
	0.65	0.052	0.066	0.050	0.062	0.017	0.020	0.053	0.077	0.099	0.102	0.024	0.028
	0.70	0.099	0.088	0.068	0.062	0.016	0.016	0.107	0.110	0.183	0.199	0.020	0.021
	0.75	0.154	0.159	0.066	0.059	0.020	0.016	0.161	0.168	0.315	0.322	0.025	0.048
500	**0.60**	0.083	0.076	0.092	0.081	0.044	0.041	0.080	0.081	0.130	0.121	0.075	0.063
	0.65	0.146	0.158	0.139	0.154	0.049	0.061	0.155	0.177	0.243	0.287	0.077	0.102
	0.70	0.254	0.266	0.194	0.194	0.054	0.052	0.284	0.300	0.462	0.508	0.096	0.087
	0.75	0.372	0.361	0.203	0.187	0.049	0.037	0.399	0.399	0.662	0.677	0.067	0.067
1000	**0.60**	0.226	0.231	0.241	0.234	0.134	0.128	0.232	0.245	0.308	0.339	0.232	0.223
	0.65	0.398	0.398	0.369	0.354	0.186	0.158	0.396	0.413	0.564	0.588	0.324	0.307
	0.70	0.543	0.518	0.463	0.393	0.177	0.099	0.566	0.569	0.766	0.772	0.327	0.273
	0.75	0.634	0.591	0.475	0.381	0.121	0.069	0.685	0.651	0.889	0.894	0.255	0.181
2000	**0.60**	0.502	0.514	0.524	0.526	0.345	0.345	0.511	0.534	0.611	0.639	0.509	0.518
	0.65	0.709	0.710	0.685	0.668	0.423	0.379	0.721	0.730	0.841	0.855	0.651	0.632
	0.70	0.777	0.749	0.749	0.698	0.404	0.285	0.829	0.807	0.931	0.930	0.679	0.604
	0.75	0.779	0.703	0.765	0.651	0.316	0.146	0.845	0.793	0.961	0.958	0.569	0.441

Table 2.5: Power of the MLWS test for $Y_t = \mu_t + X_t$, with $(1-\phi_1 L)D(d_1,d_2)X_t = (I_q - \mathcal{M}_1 L)v_t$ with and without pre-whitening. Parameters values of the respective DGPs are given above. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

$\varepsilon = 0.05$ tends to give better results, whereas $\varepsilon = 0.02$ gives better power results in large samples. In view of the size and power results presented here, the rule of thumb to choose $\varepsilon = 0.05$ for $T \leq 500$, that is suggested by Qu (2011), still works well if short memory dynamics are present. Similarly, using $m = \lfloor T^{0.7} \rfloor$ for the bandwidth remains a good rule as well.

2.4.4 Testing Against Breaks in Fractionally Cointegrated Systems

In Section 2.3.3, we showed that the limit distribution of the MLWS statistic is asymptotically unaffected by the estimation of the cointegrating matrix B, so that the test can be applied to the linearly transformed system $\hat{B}X_t$. To explore the finite sample performance of this approach, we conduct a simulation study where the DGP is

$$D(0,d_2)\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} X_t = v_t,$$

with $v_t \sim N(0,\Sigma_v)$ and $\Sigma_v = ((1,\rho_v),(\rho_v,1))'$. Here the components of X_t are fractionally cointegrated with cointegrating vector $(1,-1)'$. The parameter d_2 determines the memory of both components in X_t and since $d_1 = 0$, the memory in the linear combination is reduced to zero. By increasing

			Size						Power					
		ρ_v	0		0.4		0.8		0		0.4		0.8	
d_2	T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
0.1	250	**0.60**	0.005	0.007	0.005	0.008	0.006	0.007	0.076	0.088	0.075	0.091	0.077	0.102
		0.65	0.008	0.016	0.010	0.010	0.009	0.013	0.179	0.223	0.178	0.198	0.187	0.227
		0.70	0.010	0.014	0.012	0.014	0.009	0.014	0.321	0.327	0.303	0.322	0.337	0.355
		0.75	0.012	0.021	0.011	0.019	0.011	0.016	0.471	0.465	0.459	0.455	0.509	0.517
	1000	**0.60**	0.017	0.018	0.012	0.018	0.012	0.020	0.584	0.602	0.574	0.606	0.600	0.620
		0.65	0.017	0.025	0.020	0.024	0.017	0.027	0.817	0.819	0.818	0.813	0.833	0.837
		0.70	0.023	0.027	0.020	0.029	0.022	0.030	0.910	0.902	0.906	0.903	0.932	0.935
		0.75	0.026	0.032	0.025	0.035	0.025	0.032	0.953	0.946	0.950	0.939	0.960	0.963
0.25	250	**0.60**	0.006	0.007	0.003	0.009	0.005	0.007	0.053	0.061	0.049	0.058	0.071	0.076
		0.65	0.007	0.015	0.006	0.010	0.007	0.015	0.121	0.139	0.115	0.141	0.151	0.178
		0.70	0.011	0.016	0.009	0.015	0.013	0.016	0.206	0.218	0.214	0.216	0.278	0.281
		0.75	0.016	0.019	0.011	0.021	0.012	0.018	0.334	0.337	0.328	0.330	0.423	0.427
	1000	**0.60**	0.017	0.021	0.012	0.017	0.014	0.016	0.440	0.455	0.475	0.494	0.525	0.565
		0.65	0.020	0.025	0.017	0.022	0.017	0.022	0.669	0.687	0.705	0.713	0.780	0.789
		0.70	0.020	0.032	0.019	0.029	0.025	0.029	0.828	0.818	0.844	0.835	0.894	0.899
		0.75	0.027	0.038	0.023	0.031	0.025	0.037	0.904	0.889	0.916	0.903	0.950	0.941
0.4	250	**0.60**	0.006	0.009	0.004	0.009	0.004	0.007	0.034	0.041	0.034	0.039	0.060	0.065
		0.65	0.008	0.013	0.007	0.013	0.008	0.012	0.075	0.091	0.077	0.102	0.145	0.156
		0.70	0.013	0.016	0.011	0.017	0.009	0.015	0.140	0.139	0.163	0.168	0.257	0.255
		0.75	0.018	0.022	0.014	0.021	0.015	0.022	0.235	0.227	0.262	0.262	0.401	0.400
	1000	**0.60**	0.015	0.019	0.011	0.017	0.011	0.022	0.286	0.302	0.355	0.370	0.475	0.489
		0.65	0.014	0.024	0.016	0.025	0.018	0.028	0.514	0.532	0.603	0.611	0.718	0.728
		0.70	0.018	0.029	0.021	0.025	0.022	0.028	0.716	0.704	0.777	0.775	0.877	0.878
		0.75	0.024	0.033	0.022	0.035	0.029	0.038	0.851	0.836	0.883	0.863	0.941	0.938

Table 2.6: Size and power of the MLWS test in a bivariate fractionally cointegrated system, where $D(0,d_2)BX_t = v_t$ with $B = ((1,0)',(-1,1)')$, $v_t \sim N(0,\Sigma_v)$ and $\Sigma_v = ((1,\rho_v),(\rho_v,1))'$.

d_2 the cointegration strength is increased. The correlation between the innovations to the linear combination and the common fractional trend is determined by ρ_v.

The results of this experiment are shown in Table 2.6. First, the size remains conservative for all parameter constellations. With respect to the power, one can observe that the power is decreasing with increasing strength of the cointegrating relationship. Since the convergence rate of the local Whittle estimator for the B matrix is faster if the cointegrating relationship is stronger, this effect cannot be attributed to the effect of the estimation error. Instead, the MLWS test has lower power to detect contaminations if the memory is stronger. For the Qu test this was pointed out by Kruse (2015), who advocates to apply the test to the fractionally differenced process. In addition to that, one can observe that the power is higher, the higher

		$\eta = \frac{1}{\sqrt{2}}(1,1)'$				$\eta = (1,0)'$				$\eta = (0,1)'$			
	ζ	0		2		0		2		0		2	
T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.007	0.012	0.059	0.042	0.013	0.016	0.219	0.121	0.014	0.015	0.012	0.017
	0.65	0.006	0.017	0.107	0.095	0.015	0.022	0.355	0.247	0.016	0.023	0.017	0.022
	0.70	0.011	0.018	0.177	0.150	0.019	0.024	0.511	0.380	0.018	0.026	0.023	0.024
	0.75	0.016	0.019	0.219	0.194	0.024	0.029	0.533	0.463	0.021	0.029	0.028	0.028
1000	**0.60**	0.014	0.019	0.249	0.244	0.022	0.026	0.642	0.582	0.020	0.027	0.022	0.029
	0.65	0.022	0.028	0.466	0.455	0.028	0.035	0.888	0.848	0.032	0.030	0.032	0.036
	0.70	0.023	0.030	0.703	0.598	0.034	0.036	0.976	0.938	0.033	0.046	0.040	0.045
	0.75	0.025	0.039	0.844	0.761	0.040	0.047	0.996	0.985	0.038	0.052	0.040	0.053

Table 2.7: MLWS test for breaks in components using different weight vectors η in $Y_t = (\zeta,0)'\text{SD}(X_t)\mathbb{I}_{(t/T>1/2)} + X_t$, with $v_t \sim N(0,\Sigma_v)$ and $\Sigma_v = ((1,0),(0,1))'$.

the correlation ρ_v.

2.4.5 Testing for Breaks in Components of a Multivariate System

In Section 2.3.2 we introduced a variation of the MLWS test where all components of the weight vector η are set to zero and only one takes the value 1. This allows to test for misspecifications in components of the spectral density matrix and can be used to gain further insights about which components of X_t cause a rejection of the MLWS test with equal weights.
The performance of the MLWS test using the proposed weighting scheme is evaluated in Table 2.7. Here the DGP is a bivariate fractionally integrated process with a single deterministic break in the first component. It is given by

$$Y_t = \begin{pmatrix} \zeta \\ 0 \end{pmatrix} \text{SD}(X_t)\mathbb{I}_{(t/T>1/2)} + X_t,$$

with $v_t \sim N(0,\Sigma_v)$ and $\Sigma_v = ((1,0),(0,1))'$. The parameter ζ controls the magnitude of the breaks. To determine the critical values based on an estimate $\hat{G}(\hat{d})$, we approximate the integrals in (2.10) by sums over 500 increments and we draw 1000 values. One can observe for $\zeta = 0$, the size is well controlled. Also, as one would expect, the test generates better power if we specifically test for a contamination in the first component, compared to the baseline case with equal weights. Furthermore, if one specifically tests for a contamination in the second component, the test does not generate power due to the misspecification of the off-diagonal components alone. A rejection therefore gives a strong indication for a low frequency contamination in the respective component.

2.4.6 Further Simulations

A number of further simulations are provided in a supplementary appendix. Here we explore the impact of perturbations, heteroscedasticity, breaks in the variance-covariance matrix of the innovations, the power against other alternative processes and the performance of the test, if the pre-whitening is conducted using univariate estimators. It is found, that the MLWS test is remarkably stable under all these complications. However, power against non-causal alternatives is only developed very slowly.

2.5 Empirical Example

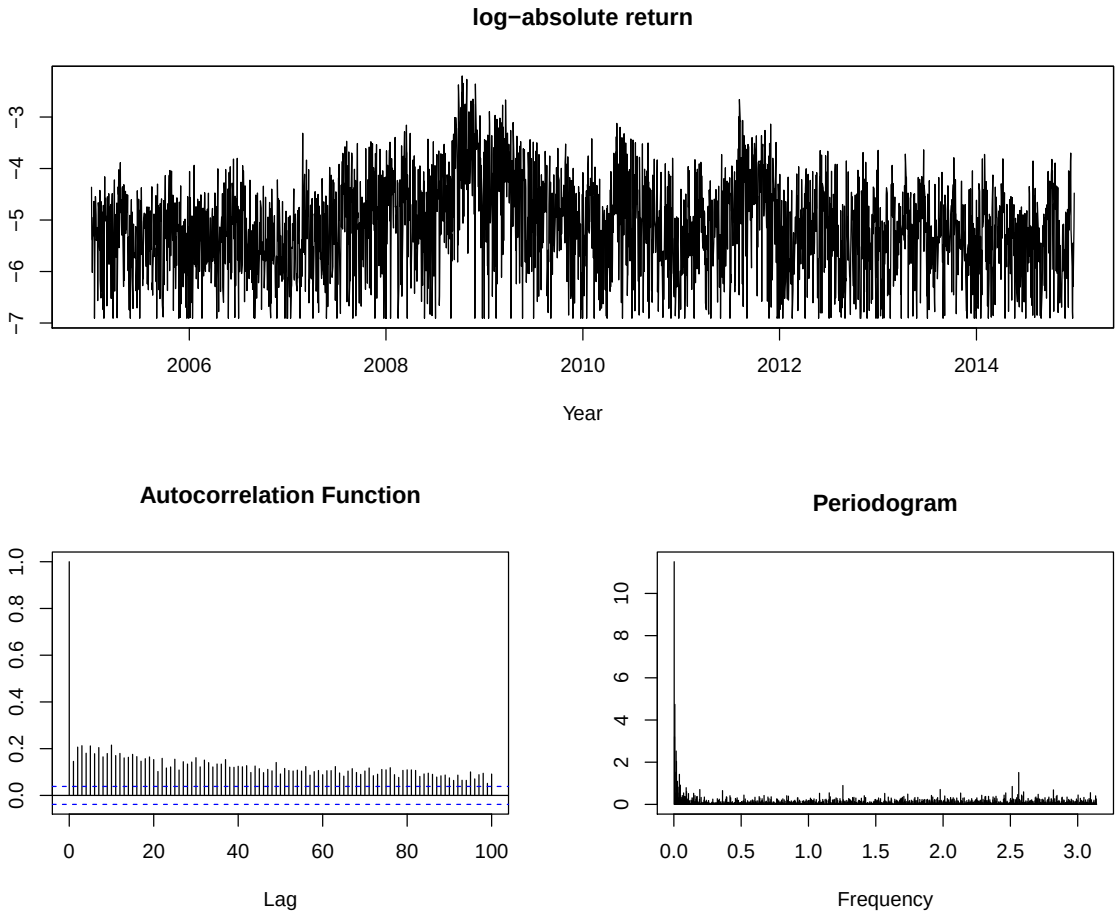

Figure 2.1: The **log-absolute return** of the S&P 500 series with the corresponding autocorrelation function and periodogram.

Log-absolute returns of stock market indices are a typical example in the spurious long memory literature - in particular that of the Standard & Poor's 500 (hereafter S&P 500). The series is examined by Granger and Ding (1996) who find that it seems to follow a long-memory process. Nevertheless, they argue that long memory properties can be generated by other models than

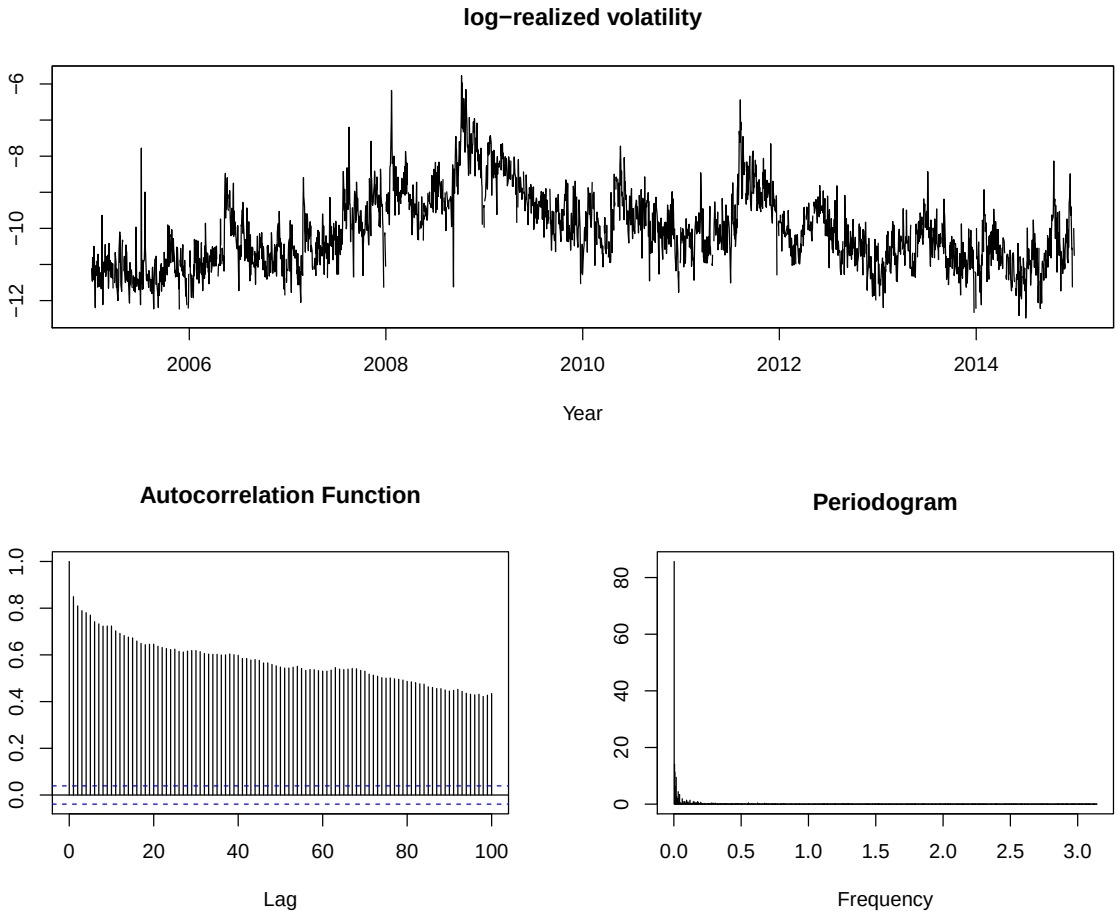

Figure 2.2: The **log-realized volatility** of the S&P 500 series with the corresponding autocorrelation function and periodogram.

the standard $I(d)$ process. Granger and Hyung (2004) obtain a reduction of the estimated memory parameter by considering structural breaks in the series. Similarly, Varneskov and Perron (2011) consider a model allowing for both random level shifts and ARFIMA effects. Lu and Perron (2010) and Xu and Perron (2014) analyze the forecast performance of random level shift processes for the log-absolute returns of the S&P 500. In most cases, random level shift processes clearly outperform GARCH, FIGARCH and HAR models.

All these findings indicate spurious long memory in log-absolute return series. However, univariate tests are often not able to reject the null hypothesis of true long memory. Dolado et al. (2005), for example, apply their test to absolute and squared returns of the S&P 500, without being able to indicate spurious long memory.

Due to the increased availability of high frequency data, the focus in the more recent literature has shifted to the modelling of realized volatility. Especially the heterogenous autoregressive model of Corsi (2009) and its extensions have become very popular. As for the log-absolute return series existing tests against spurious long memory tend not to reject their null hypothesis

	$\log(\lvert r_t\rvert + 0.001)$				$\log RV_t$			
	DAX	**NIKKEI**	**S&P 500**	**FTSE**	**DAX**	**NIKKEI**	**S&P 500**	**FTSE**
mean	-4.977	-4.924	-5.156	-5.107	-9.789	-9.432	-10.055	-9.665
median	-4.920	-4.803	-5.133	-5.051	-9.946	-9.543	-10.183	-9.741
std.dev.	0.924	1.004	0.937	0.900	1.126	0.964	1.017	0.885
skewness	-0.131	-0.307	0.053	-0.087	0.744	0.662	0.639	0.747
kurtosis	2.475	2.419	2.528	2.573	3.666	3.751	3.295	4.301

Table 2.8: Summary statistics of the log-absolute returns and the log-realized volatility series.

if applied to these realized volatility series. Examples include the application in Qu (2011), who finds no evidence for low frequency contaminations in the realized volatility of the exchange rate between Japanese Yen and US Dollar and Kruse et al. (2016) who apply the Qu test to S&P 500 log-realized volatility.

In view of the power gains of the multivariate procedure demonstrated in Section 2.4.2, we revisit these variation series of the S&P 500 and additionally consider those of the DAX, FTSE and NIKKEI in a multivariate setup to test for spurious long memory using the MLWS test. The analysis is conducted for both - the log-absolute return and the log-realized volatility. Note however, that both of these measures follow different variation concepts. It is therefore not to expect that features found in one of them will translate to the other.

We analyze the period from 2005/01/03 to 2014/12/31 (T=2608 observations). Data on daily stock price indices is obtained from Thomson Reuters Datastream. The log-returns are computed by first differencing the logarithm of the price index, $r_t = \ln(P_t) - \ln(P_{t-1})$. Subsequently, the log-absolute returns are calculated as $\ln(|r_t| + 0.001)$.[3] Realized volatilities calculated from 5 minute returns are obtained from the Oxford-Man Realized Library.

As an example, Figures 2.1 and 2.2 depict the log-absolute return and log-realized volatility of the S&P 500 series. Both, the autocorrelation functions and the periodograms show the typical characteristics of long-memory processes. Since the series of the DAX, FTSE and NIKKEI are highly correlated with that of the S&P 500, we omit plots of these series. Descriptive statistics for the dataset are given in Table 2.8. It can be seen that all four series have similar locations and standard deviations if the same variation measure is used. With the exception of the S&P 500, the distributions of the log-absolute return series are slightly negatively skewed and all log-absolute return series have lighter tails than the normal distribution. The realized volatility series on the other hand are positively skewed and have excess kurtosis.

Since the specification of the MLWS test depends on whether or not the series are fractionally cointegrated, we proceed by applying the semiparametric cointegrating rank estimation method

[3]The constant 0.001 is added to avoid infinite values for zero returns, by following Lu and Perron (2010) and Xu and Perron (2014).

δ	DAX	NIKKEI	S&P 500	FTSE	partitions	coint.rank	$\hat{\beta}$	$\hat{\mathbf{d}}_{\mathbf{w}}$
0.60	0.379	0.295	0.472	0.393	(1,1,1,1)	1	(0.155, 0.066, -1.446)	0.260
0.65	0.338	0.290	0.411	0.362	(1,1,1,1)	1	(-0.119, 0.079, -1.014)	0.236
0.70	0.328	0.285	0.359	0.303	(1,1,1,1)	1	(-0.043, -0.144, -1.153)	0.194
0.75	0.264	0.252	0.300	0.290	(1,1,1,1)	1	(-0.074, -0.037, -0.954)	0.139

Table 2.9: Fractional cointegration analysis for the **log-absolute return series** based on local Whittle estimates of d with different bandwidths $m = \lfloor T^{\delta} \rfloor$.

δ	DAX	NIKKEI	S&P 500	FTSE	partitions	coint.rank	$\hat{\beta}_{13}$	$\hat{\beta}_{23}$	$\hat{\mathbf{d}}_{\mathbf{w}_{13}}$	$\hat{\mathbf{d}}_{\mathbf{w}_{23}}$
0.60	0.642	0.631	0.635	0.637	(1,1,1,1)	2	-0.931	-1.084	0.464	0.596
0.65	0.605	0.612	0.642	0.570	(1,1,1,1)	2	-0.728	-0.978	0.463	0.514
0.70	0.594	0.611	0.633	0.568	(1,1,1,1)	2	-0.868	-1.068	0.400	0.483
0.75	0.563	0.573	0.588	0.540	(1,1,1,1)	2	-0.910	-1.172	0.368	0.447

Table 2.10: Fractional cointegration analysis for the **log-realized volatility series** based on local Whittle estimates of d with different bandwidths $m = \lfloor T^{\delta} \rfloor$.

of Robinson and Yajima (2002). The method consists of two steps. First, the vector series X_t is partitioned into subvectors with equal memory parameters using sequential tests for the equality of the d_a in each subvector. In the second step, the cointegrating rank of the relevant subvectors is estimated.

All results of this procedure are given in Tables 2.9 and 2.10. The analysis is carried out for different bandwidths $m = \lfloor T^{\delta} \rfloor$ using the local Whittle estimator. For both variation measures it can be observed, that the estimates tend to decrease as the bandwidth increases, which indicates that the series indeed might be contaminated by level shifts.

Since the log-absolute return series is considered to be a noisy estimate of the underlying absolute variation process and perturbations cause a downward bias in the local Whittle estimator, we include further results using different specifications of the LPWN estimator of Frederiksen et al. (2012) and the robust estimator of Hou and Perron (2014) in Table 2.12 in the supplementary appendix. It can be observed that there is a downward bias for the log-absolute return series. Nevertheless, as discussed in Section 2.4 the MLWS test is fairly robust to perturbations. Also the Hou-Perron estimator is lower for the S&P 500, which is a further indication of spurious long memory. Apart from that, all estimates turn out to be remarkably stable.

Using the $\hat{T}_0$ statistic of Robinson and Yajima (2002) to test for the equality of the memory parameters, the null hypothesis cannot be rejected for any of the bandwidths, so that no further partitioning of X_t is necessary. Subsequently, the cointegrating rank of X_t is estimated. Again,

the results are remarkably stable for different bandwidth choices. We find that there is one cointegrating relationship between the four log-absolute return series and there are two relationships between the realized volatility series.

As described in Section 2.3.3, the analysis than proceeds by estimating the cointegrating matrix B using the multivariate local Whittle estimator of Robinson (2008) with the phase set to $(d_a - d_b)\pi/2$.

In the case of the log-absolute return series the DAX series is specified to be the variable that is replaced by the linear combination. For the log-realized volatility series we assume pairwise relationships of the DAX and the NIKKEI with the S&P 500. Subsequently, the transformed series $\hat{B}X_t$ are obtained. Additionally, we report the estimate $\hat{d}_w$ of the noise term in the last column of Table 2.9 and the last two columns of Table 2.10 to show that the memory in the linear combination is reduced. When the cointegrating rank analysis is repeated on the transformed series there is no evidence for a cointegrating relationship anymore, supporting the selection of the estimated cointegrating relations.

It should be noted, that the rank-estimation procedure of Robinson and Yajima (2002) operates under the assumption of a multivariate long memory series. In the presence of low frequency contaminations on the hand, it will no longer be consistent. The estimates of the cointegrating relations should therefore not be interpreted unless the MLWS test fails to reject.

To formally test for true long memory, we then apply the robustified multivariate local Whittle score-type test ($\widetilde{MLWS}$) to our system $\hat{B}X_t$. The test for contaminations in components of the system discussed in Section 2.3.2 is applied to analyze which components of the series might cause a rejection of the MLWS test. As a benchmark, we also apply the univariate test of Qu (2011) to each series separately. Because of the large number of observations the trimming parameter is set to $\varepsilon = 0.02$ for both tests. The corresponding test statistics are given in Table 2.11, where the p-values are displayed in brackets.[4]

As one can see, Qu's univariate test fails to reject the null hypothesis of true long memory for each country, all bandwidth specifications, and both variation measures. The only exception is the log-absolute return series of the S&P 500, if the bandwidth is set to $m = \lfloor T^{0.75} \rfloor$. This would lead to the conclusion that there are no low frequency contaminations in the variation of stock returns. The MLWS statistic on the other hand rejects for the log-absolute return series for all but one bandwidth. In the case of the realized volatility series however, it also fails to reject - except for $m = \lfloor T^{0.75} \rfloor$.

If one considers the tests for contaminations in components of the spectral density matrix, we find that the test rejects for the S&P 500 series if the bandwidth parameter is $\delta \in \{0.65, 0.70, 0.75\}$ for the log-absolute return series, but not for the realized volatility series. The application of the MLWS test therefore gives formal support to the arguments of Granger and Ding (1996) and Granger and Hyung (2004), among others, who argued that the memory in the log-absolute

[4] Due to the large number of free parameters in the 4-dimensional example, the pre-whitening is carried out for each series separately. Monte Carlo results supporting the validity of this approach are provided in the supplementary appendix.

	Qu test				Components				MLWS
δ	DAX	NIKKEI	S&P 500	FTSE	DAX	NIKKEI	S&P 500	FTSE	ALL
				$\log(\lvert r_t\rvert + 0.001)$					
0.60	0.521	0.860	0.515	0.446	1.245	0.877	0.765	1.085	1.233
	(0.862)	(0.314)	(0.871)	(0.949)	(0.494)	(0.665)	(0.854)	(0.912)	(0.054)
0.65	0.505	0.749	1.078	0.443	1.052	0.793	1.681	0.619	1.470
	(0.886)	(0.474)	(0.118)	(0.953)	(0.220)	(0.495)	(0.013)	(0.979)	(0.014)
0.70	0.395	0.519	1.100	0.739	1.370	0.814	1.726	0.557	1.448
	(0.983)	(0.865)	(0.107)	(0.492)	(0.179)	(0.395)	(0.004)	(0.998)	(0.016)
0.75	0.640	0.469	1.477	0.547	1.322	0.918	1.843	0.598	1.413
	(0.662)	(0.929)	(0.013)	(0.824)	(0.057)	(0.257)	(0.002)	(0.948)	(0.019)
				$\log RV_t$					
0.60	0.317	0.425	1.179	0.544	0.700	0.452	1.283	0.548	0.662
	(0.999)	(0.966)	(0.071)	(0.829)	(0.600)	(0.978)	(0.173)	(0.992)	(0.621)
0.65	0.445	0.641	0.807	0.929	1.241	0.965	1.140	0.985	1.465
	(0.950)	(0.661)	(0.387)	(0.236)	(0.136)	(0.187)	(0.279)	(0.425)	(0.014)
0.70	0.406	0.657	0.700	0.670	0.617	0.539	1.039	0.807	0.643
	(0.977)	(0.632)	(0.554)	(0.607)	(0.729)	(0.881)	(0.331)	(0.635)	(0.656)
0.75	0.597	1.062	0.724	1.022	0.488	0.397	0.614	1.199	0.683
	(0.740)	(0.129)	(0.513)	(0.152)	(0.946)	(0.999)	(0.919)	(0.082)	(0.585)

Table 2.11: Test statistics of the Qu test applied to each series separately and the MLWS test applied to the multivariate series for different bandwidths $m = \lfloor T^\delta \rfloor$. p-values are given in brackets. Critical values are 1.252 and 1.374 for $\alpha = 5\%$ and $\alpha = 1\%$, respectively.

returns of the S&P 500 might be spurious.

We therefore find that one would falsely conclude that the process is not contaminated, if only the univariate test is used. In contrast to that, there is no evidence for low frequency contaminations in the log-realized volatility series, which are therefore well modelled as long memory processes.

2.6 Conclusion

This paper provides a multivariate score-type test for spurious long memory based on the objective function of the local Whittle estimator. The test statistic consists of a weighted sum of the partial derivatives of the concentrated local Whittle likelihood function. By introducing a suitable weighting scheme, the test statistic becomes pivotal and the limiting distribution becomes independent of the dimension of the data generating process. Consistency against multivariate random level shift processes and smooth trends is shown.

Our test encompasses the test of Qu (2011) as a special case for scalar processes. Apart from

the generalization to vector valued series, we consider several issues that are unique to the multivariate case. First, we provide a modification of the test statistic in the case of fractionally cointegrated series which has the same asymptotic properties as the original test statistic. Second, by changing the weighting scheme, the multivariate test statistic can be used to gain insights about which components of the multivariate series cause a rejection.

A Monte Carlo study shows that the test has good size and power properties in finite samples. These properties hold for different bandwidths, $m = \lfloor T^{\delta} \rfloor$, as well as for different trimming parameters ε. Furthermore, the size and power remain good if the dimensions of the data generating process increase and the test is robust against short memory dynamics if a pre-whitening procedure is applied.

In our empirical example we revisit the log-absolute returns of the S&P 500 together with the DAX, FTSE and NIKKEI stock indices in a multivariate framework. By applying our multivariate test, we find evidence of spurious long memory in the log-absolute returns of the S&P 500. A simple application of the univariate Qu test to the log-absolute returns, on the other hand, cannot reject the null hypothesis of true long memory. As discussed in Section 2.5, several authors have pointed out that the log-absolute returns might follow a spurious long-memory process. Our empirical application adds to this literature by providing a formal rejection of pure long memory in the sense of a statistically significant test decision. For realized volatilities on the other hand, no such evidence is found.

Appendix

Proof of Lemma 2.1:

To prove the lemma, note the following arguments in Shimotsu (2007)

$$\begin{aligned}
\sum_{a=1}^{q}\eta_a\sqrt{m}\frac{\partial R(d)}{\partial d_a} &= \sum_{a=1}^{q}\eta_a\sqrt{m}\left[-\frac{2}{m}\sum_{j=1}^{m}\log\lambda_j + tr\left[\hat{G}(d)^{-1}\frac{\partial\hat{G}(d)}{\partial d_a}\right]\right] \\
&= \sum_{a=1}^{q}\eta_a\sqrt{m}\left[-\frac{2}{m}\sum_{j=1}^{m}\log\lambda_j + tr\left[\hat{G}(d)^{-1}\left[\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j\right.\right.\right. \\
&\quad \times Re\left[(\Lambda_j(d))^{-1}(i_a I(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}\right] \\
&\quad \left.\left.\left. +\frac{1}{m}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j(d))^{-1}(-i_a I(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}\right]\right]\right]\right] \\
&= \sum_{a=1}^{q}\eta_a\sqrt{m}\left[-\frac{2}{m}\sum_{j=1}^{m}\log\lambda_j + tr\left[\hat{G}(d)^{-1}\times H_{1a}+H_{2a}\right]\right]
\end{aligned}$$

with $H_{1a}=\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j Re[(\Lambda_j(d))^{-1}(i_a I(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}]$ and $H_{2a}=\frac{1}{m}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}$ $Im[(\Lambda_j(d))^{-1}(-i_a I(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}]$.
Therefore, we can write with $\nu_j=\log\lambda_j-1/m\sum_{j=1}^{m}\log\lambda_j$, ${}_aM$ denoting the a-th row of the matrix M, and M_a denoting the a-th column of the matrix M

$$\begin{aligned}
\sum_{a=1}^{q}\eta_a\sqrt{m}\frac{\partial R(d)}{\partial d_a} &= \sum_{a=1}^{q}\eta_a\left[tr\left[\hat{G}(d)^{-1}\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\nu_j Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]i_a\right.\right. \\
&\quad \left.\left. +\hat{G}(d)^{-1}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j(d))^{-1}(-i_a I(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}\right]\right]\right] \\
&= \frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\nu_j\left({}_a\left(G^0\right)^{-1}\left[Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]\right]_a-1\right)+o_P(1) \\
&\quad +tr\left[\hat{G}(d)^{-1}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j(d))^{-1}(-i_a I(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}\right]\right]
\end{aligned}$$

which proves our lemma. □

Proof of Theorem 2.1:

To prove the theorem we start with the Taylor expansion

$$\sqrt{m}\eta' \frac{\partial R^r(d)}{\partial d}\Big|_{\hat{d}} = \sqrt{m}\eta' \frac{\partial R^r(d)}{\partial d}\Big|_{d^0} + \sqrt{m}\eta' \frac{\partial^2 R^r(d)}{\partial d \partial d'}\Big|_{\bar{d}} (\hat{d} - d^0) \tag{2.12}$$

where $\bar{d}$ fulfills $\|\bar{d} - d^0\| \leq \|\hat{d} - d^0\|$ and the notation $R^r(d)$ indicates that the summation is done until $[mr]$ rather than m. For the first part of the right hand side of equation (2.12) we can write:

$$\begin{aligned}
\sum_{a=1}^{q} \eta_a \sqrt{m} \frac{\partial R^r(d)}{\partial d_a}\Big|_{d^0} &= \frac{2}{\sqrt{m}} \sum_{a=1}^{q} \eta_a \sum_{j=1}^{[mr]} \nu_j \Big({}_a\big(G^0\big)^{-1} \Big[Re\Big[\Lambda_j^0(d)^{-1} I(\lambda_j) \Lambda_j^{0*}(d)^{-1} \Big] \Big]_a - 1 \Big) \\
&\quad + o_P(1) + tr\Bigg[\hat{G}\big(d^0\big)^{-1} \frac{1}{\sqrt{m}} \sum_{j=1}^{[mr]} \frac{\lambda_j - \pi}{2} Im\Big[(\Lambda_j^0(d))^{-1} (-i_a I(\lambda_j) \\
&\quad + I(\lambda_j) i_a) (\Lambda_j^{0*}(d))^{-1} \Big] \Big] \\
&= \frac{2}{\sqrt{m}} \sum_{a=1}^{q} {}_aG^0 \eta_a \sum_{j=1}^{[mr]} \nu_j \Big[{}_a\big(G^0\big)^{-1'} {}_a\big(G^0\big)^{-1} \Big[Re\Big[\Lambda_j^0(d)^{-1} I(\lambda_j) \Lambda_j^{0*}(d)^{-1} \Big] \Big]_a - 1 \Big] \\
&\quad - \frac{2_aG^0}{m^{3/2}} \Bigg(\sum_{j=1}^{[mr]} \nu_j \Bigg) \sum_{j=1}^{m} \Big[{}_a\big(G^0\big)^{-1'} {}_a\big(G^0\big)^{-1} \Big[Re\Big[\Lambda_j^0(d)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j) \\
&\quad \times \Lambda_j^{0*}(d)^{-1} \Big] \Big]_a - 1 \Big] + o_P(1) + tr\Bigg[\hat{G}(d_0)^{-1} \frac{1}{\sqrt{m}} \sum_{j=1}^{[mr]} \frac{\lambda_j - \pi}{2} Im\Big[(\Lambda_j^0(d))^{-1} \\
&\quad \times (-i_a I(\lambda_j) + I(\lambda_j) i_a) (\Lambda_j^{0*}(d))^{-1} \Big] \Big].
\end{aligned} \tag{2.13}$$

By arguments as in Shimotsu (2007), we can write the first term plus the imaginary part as $\sum_{t=1}^{T} z_t + o_P(1)$ with $z_1 = 0$ and $z_t = \varepsilon_t' \sum_{s=1}^{t-1} [\Theta_{t-s} + \tilde{\Theta}_{t-s}] \varepsilon_s$. Here $\Theta_s = \frac{1}{\pi\sqrt{m}T} \sum_{j=1}^{m} \nu_j Re[\Psi_j + \Psi_j'] \cos(s\lambda_j)$, $\tilde{\Theta}_s = \frac{\pi}{2} \frac{1}{\pi\sqrt{m}T} \sum_{j=1}^{m} Re[\Psi_j - \Psi_j'] \sin(s\lambda_j)$, Ψ_j is defined by $\Psi_j = \sum_{a=1}^{q} \eta_a$ $[A^*(\lambda_j) \Lambda_j^{0*}(d)^{-1}]_{aa} (G^0)^{-1} \Lambda_j^0(d)^{-1} A(\lambda_j)$, $A(\lambda) = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$ and A_j is given in Assumption 2.2. The asymptotic normality of z_t follows from Theorem 2 of Robinson (1995a). To obtain the covariance of the z_t we have for $0 \leq r_1 \leq r_2 \leq 1$

$$\begin{aligned}
Cov\Bigg(\sum_{t=1}^{T} z_{t,r_1}, \sum_{t=1}^{T} z_{t,r_2} \Bigg) &= E\Bigg(\sum_{t=1}^{T} \sum_{s=1}^{t-1} \Big(\Theta_{t-s,r_1} + \tilde{\Theta}_{t-s,r_1} \Big)' \varepsilon_s' \sum_{l=1}^{t-1} \Big(\Theta_{t-s,r_2} + \tilde{\Theta}_{t-s,r_2} \Big) \varepsilon_l \Bigg) \\
&= \sum_{t=2}^{T} \sum_{s=1}^{t-1} tr\Big[\Big(\Theta_{t-s,r_1} + \tilde{\Theta}_{t-s,r_1} \Big)' \Big(\Theta_{t-s,r_2} + \tilde{\Theta}_{t-s,r_2} \Big) \Big] + o_P(1)
\end{aligned}$$

by using Lemma 2 and 3 from Lobato (1999). Now we have

$$\begin{aligned}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\Theta'_{t-s,r_1}\tilde{\Theta}_{t-s,r_2} &= \frac{1}{2\pi mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\sum_{j=1}^{[mr_1]}\sum_{l=1}^{[mr_2]}\nu_j\nu_l tr\left[Re\left[\Psi'_j+\Psi_j\right]Re\left[\Psi_l-\Psi'_l\right]\right]\\ &\quad\times\cos(s\lambda_j)\sin(s\lambda_j)\\ &= 0\end{aligned}$$

as $tr[(A'+A)(B-B')]=0$ for any real matrices A and B. Furthermore, we have

$$\begin{aligned}&\sum_{t=2}^{T}\sum_{s=1}^{t-1}tr\left[\Theta'_{t-s,r_1}\Theta_{t-s,r_2}+\tilde{\Theta}'_{t-s,r_1}\tilde{\Theta}_{t-s,r_2}\right]\\ =\;&\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}\nu_j^2\times tr\left[Re\left[\Psi'_j+\Psi_j\right]\,Re\left[\Psi_j+\Psi'_j\right]\right]\cos^2(s\lambda_j)\\ +\;&\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}\sum_{l=1,l\neq j}^{[mr_2]}\nu_j\nu_l tr\left[Re\left[\Psi'_j+\Psi_j\right]\,Re\left[\Psi_l+\Psi'_l\right]\right]\cos(s\lambda_j)\cos(s\lambda_l)\\ +\;&\frac{\pi^2}{4}\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}tr\left[Re\left[\Psi'_j-\Psi_j\right]\,Re\left[\Psi_j-\Psi'_j\right]\right]\sin^2(s\lambda_j)\\ +\;&\frac{\pi^2}{4}\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}\sum_{l=1,l\neq j}^{[mr_2]}tr\left[Re\left[\Psi'_j\Psi_j\right]Re\,\left[\Psi_l-\Psi'_l\right]\right]\sin(s\lambda_j)\sin(s\lambda_l).\end{aligned}$$

The second and fourth term of this sum are $o_P(1)$ by Lemma 3b) and 3d) in Shimotsu (2007). Applying Lemma 3a) in Shimotsu (2007) for the first term, we obtain for $\lambda_j\to 0$

$$\begin{aligned}&\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}\nu_j^2 tr\left[Re\left[\Psi'_j+\Psi_j\right]Re\left[\Psi_j+\Psi'_j\right]\right]\cos^2(s\lambda_j)\\ =\;&\frac{1}{m}\sum_{j=1}^{[mr_1]}\nu_j^2\frac{1}{4\pi^2}tr\left[Re\left[\Psi'_j+\Psi_j\right]Re\left[\Psi_j+\Psi'_j\right]\right]\\ =\;&\frac{1}{m}\sum_{j=1}^{[mr_1]}\nu_j^2\left(2\sum_{a=1}^{q}\eta_a^2+2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}\right).\end{aligned}$$

For the third term we have again for $\lambda_j\to 0$ by Shimotsu (2007) Lemma 3c)

$$\begin{aligned}&\frac{\pi^2}{4}\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}tr\left[Re\left[\Psi'_j-\Psi_j\right]Re\left[\Psi_j-\Psi'_j\right]\right]\sin^2(s\lambda_j)\\ =\;&\frac{\pi^2}{4m}\sum_{j=1}^{[mr_1]}\frac{1}{4\pi^2}tr\left[Re\left[\Psi'_j-\Psi_j\right]Re\left[\Psi_j-\Psi'_j\right]\right]\\ =\;&\frac{\pi^2}{4m}\sum_{j=1}^{[mr_1]}\left(2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}-2\sum_{a=1}^{q}\eta_a^2\right).\end{aligned}$$

From the Euler-Mc Laurin equality and Lemma B.1 in Qu (2011) it follows that $1/m\sum_{j=1}^{[mr_1]}\nu_j^2 \to \int_0^{r_1}(1+\log s)^2 ds$. We thus obtain altogether

$$\begin{aligned} Cov\left(\sum_{t=1}^{T} z_{t,r_1}, \sum_{t=1}^{T} z_{t,r_2}\right) \quad \to \quad & \int_0^{r_1}\left[(1+\log s)^2\left(2\sum_{a=1}^{q}\eta_a^2 + 2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G_{ab}^0\left(G^0\right)_{ab}^{-1}\right)\right. \\ + \quad & \left.\frac{\pi^2}{4}\left(2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G_{ab}^0\left(G^0\right)_{ab}^{-1} - 2\sum_{a=1}^{q}\eta_a^2\right)\right] ds. \end{aligned}$$

For the second term of the second equality of (2.13) we have by similar arguments as in Qu (2011) that

$$\frac{{}_aG^0}{m^{1/2}}\sum_{j=1}^{m}\left[{}_a\left(G^0\right)^{-1'}{}_a\left(G^0\right)^{-1}\left[Re\left[\Lambda_j^0(d)^{-1}I(\lambda_j)\Lambda_j^{0*}(d)^{-1}\right]\right]_a - 1\right] \Rightarrow B(1),$$

where $B(s)$ denotes a standard Brownian motion. As before we have from Lemma B.1 in Qu (2011) that

$$\frac{1}{m}\sum_{j=1}^{[mr_1]}\nu_j \to \int_0^{r_1}(1+\log s)ds.$$

It remains the last part of the Taylor decomposition. For the second derivative of the objective function $R(d)$ we obtain

$$\frac{\partial^2 R(d)}{\partial d_a \partial d_b} = tr\left[-\hat{G}^{-1}(d)\frac{\partial\hat{G}(d)}{\partial d_a}\hat{G}^{-1}(d)\frac{\partial\hat{G}(d)}{\partial d_b} + \hat{G}^{-1}(d)\frac{\partial^2\hat{G}(d)}{\partial d_a\partial d_b}\right].$$

Thus,

$$\frac{\partial\hat{G}^r(d)}{\partial d_a} = \frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)\hat{G}_{1a}(d) + o_P\left((logT)^{-1}\right)$$

with

$$\hat{G}_{1a}(d) = i_a\hat{G}(d) + \hat{G}(d)i_a$$

and the superscript r denoting again that the sum goes up to $[mr]$ rather than m. Furthermore, it is

$$\frac{\partial^2\hat{G}^r(d)}{\partial d_a\partial d_b} = \frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)^2\hat{G}_{2ab}(d) + \frac{\pi^2}{4}\hat{G}_{3ab}(d) + o_P(1),$$

with

$$G_{2ab}(d) = i_a i_b\hat{G}(d) + i_a\hat{G}(d)i_b + i_b\hat{G}(d)i_a + \hat{G}(d)i_a i_b$$

and

$$G_{3ab}(d) = -i_a i_b\hat{G}(d) + i_a\hat{G}(d)i_b + i_b\hat{G}(d)i_a - \hat{G}(d)i_a i_b.$$

It also holds that

$$tr\left[\hat{G}(d)^{-1}\hat{G}_{1a}(d)\hat{G}(d)^{-1}\hat{G}_{1b}(d)\right] = tr\left[\hat{G}(d)^{-1}\hat{G}_{2ab}(d)\right].$$

Altogether this gives

$$\begin{aligned}
\frac{\partial^2 R^r(d)}{\partial d_a \partial d_b} &= tr\Bigg[-\hat{G}(d)^{-1}\left(\frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)\hat{G}_{1a}(d)+o_P\left((logT)^{-1}\right)\right)\\
&\quad\times\hat{G}(d)^{-1}\left(\frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)\hat{G}_{1b}(d)+o_P\left((logT)^{-1}\right)\right)\\
&\quad+\hat{G}(d)^{-1}\left(\frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)^2\hat{G}_{2ab}(d)+\frac{\pi^2}{4}\hat{G}_{3ab}(d)+o_P(1)\right)\Bigg]\\
&= tr\left[\left(\frac{1}{m}\sum_{j=1}^{[mr]}v_j^2\right)\hat{G}(d)^{-1}\hat{G}_{2ab}(d)+\frac{\pi^2}{4}\hat{G}(d)^{-1}\hat{G}_{3ab}(d)\right]+o_P\left(log^2T\right)\\
&\to tr\left[\int_0^r(1+\log s)^2 ds\left(G^0\right)^{-1}G^0_{2ab}+\frac{\pi^2}{4}\left(G^0\right)^{-1}G^0_{3ab}\right]
\end{aligned}$$

so that

$$\begin{aligned}
\frac{\partial^2 R^r(d)}{\partial d\partial d'} &\to 2\int_0^r\left((1+\log s)^2(G^0\odot\left(G^0\right)^{-1}+I_q)+\frac{\pi^2}{4}(G^0\odot\left(G^0\right)^{-1}-I_q)\right)ds\\
&:= F(r). \qquad (2.14)
\end{aligned}$$

From the mean value theorem, we have

$$\sqrt{m}(\hat{d}-d^0)=\sqrt{m}\left(\frac{\partial^2 R(d)}{\partial d\partial d'}|_{\bar{d}}\right)^{-1}\frac{R(d)}{\partial d}|_{d^0}.$$

Since from Shimotsu (2007) $\frac{\partial^2 R(d)}{\partial d\partial d'}|_{\bar{d}}\to\Omega$, with $\Omega=2\left[G^0\odot(G^0)^{-1}+I_q+\frac{\pi^2}{4}(G^0\odot\left(G^0\right)^{-1}-I_q)\right]$, we have

$$\sqrt{m}(\hat{d}-d^0)\to\sqrt{m}\Omega^{-1}\frac{\partial R(d)}{\partial d}|_{d^0}$$

and finally using the result from (2.14)

$$\eta'\frac{\partial^2 R^r(d)}{\partial d\partial d'}\sqrt{m}(\hat{d}-d_0)\to\eta' F(r)\Omega^{-1}\sqrt{m}\frac{\partial R(d)}{\partial d}|_{d^0}. \qquad (2.15)$$

Now, $\sqrt{m}\frac{\partial R(d)}{\partial d}|_{d^0}$ can be treated as before. Thus, the right hand side of (2.15) has the covariance $\int_0^1(1+\log s)^2 2\eta' F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}+I_q)\Omega^{-1'}F(r)'\eta+\frac{\pi^2}{2}\eta' F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}-I_q)\Omega^{-1'}F(r)'\eta ds$. Al-

together, we obtain

$$\eta' \frac{\partial^2 R^r(d)}{\partial d \partial d'} \sqrt{m}(\hat{d} - d_0) \quad \Rightarrow \quad \int_0^1 \Big[(1+\log s)\Big(2\eta' F(r)\Omega^{-1}(G^0 \odot (G^0)^{-1} + I_q)\Omega^{-1'} F(r)' \eta\Big)^{1/2} + i\Big(\frac{\pi^2}{2}\eta' F(r)\Omega^{-1}(G^0 \odot (G^0)^{-1} - I_q)\Omega^{-1'} F(r)' \eta\Big)^{1/2}\Big] dB(s).$$

Like Qu (2011), we use Theorem 13.5 of Billingsley (2009) to prove tightness. Thus, we show that for every m and $r_1 \le r \le r_2$

$$E\left(\left|\sum_{t=1}^{T} z_{t,r} - \sum_{t=1}^{T} z_{t,r_1}\right|^2 \left|\sum_{t=1}^{T} z_{t,r_2} - \sum_{t=1}^{T} z_{t,r}\right|^2\right) \le K(\psi_m(r_2) - \psi_m(r_1))^2$$

where K is some constant and $\psi_m(\cdot)$ is a function on $[0,1]$ which is finite, nondecreasing and fulfills

$$\lim_{\delta \to 0} \limsup_{m \to \infty} |\psi_m(s+\delta) - \psi_m(s)| \to 0$$

uniformly in $s \in [0,1]$. Here we denote $z_t(s,r) = z_{t,r} - z_{t,s}$. Denote also $c_t(r,s) = c_{t,r} - c_{t,s}$ and $c_t = tr[\Theta_t + \tilde{\Theta}_t]$. Using this notation we can use Qu's (2011) Lemma B.8 to show that $E(|\sum_{t=1}^{T} z_{t,r} - \sum_{t=1}^{T} z_{t,r_1}|^2 |\sum_{t=1}^{T} z_{t,r_2} - \sum_{t=1}^{T} z_{t,r}|^2)$ is bounded from above by

$$K\left(\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2\right)\left(\sum_{t=1}^{T}\sum_{h=1}^{t-1} c_{t-h}(r,r_2)^2\right)$$

where K is some positive constant. By similar arguments as in Qu (2011) we obtain furthermore

$$\begin{aligned}\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2 &\le \left(\frac{1}{Tm}\sum_{j=[mr_1]+1}^{[mr]}\sum_{k\neq j}^{[mr]}\left(\nu_j^2+\nu_k^2\right) + \frac{1}{m}\sum_{j=[mr_1]+1}^{[mr]}\nu_j^2\right) \\ &\quad \times\left(2\sum_{a=1}^{q}\eta_a^2 + 2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}\right) \\ &\le \frac{3}{m}\sum_{j=[mr_1]+1}^{[mr]}\nu_j^2\left(2\sum_{a=1}^{q}\eta_a^2 + 2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}\right).\end{aligned}$$

As $(2\sum_{a=1}^{q}\eta_a^2 + 2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}(G^0)^{-1}_{ab}) \le K$ for some constant K we set $\psi_m(s) = 1/m \sum_{j=1}^{[ms]}\nu_j^2$. This satisfies the condition as

$$\lim_{\delta\to 0}\limsup_{m\to\infty}|\psi_m(s+\delta)-\psi_m(s)| = \lim_{\delta\to 0}\int_s^{s+\delta}(1+\log x)^2 dx \quad \to \quad 0.$$

This proves the theorem. □

Proof of Lemma 2.2:

To prove the lemma, we first need to show that $\eta'(G^0 \odot (G^0)^{-1})\eta = 1$, if $\eta = (1/\sqrt{q}, ..., 1/\sqrt{q})'$. For this denote

$$G^0 = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1q} \\ g_{21} & g_{22} & \cdots & g_{2q} \\ \vdots & \ddots & & \vdots \\ g_{q1} & g_{q2} & \cdots & g_{qq} \end{pmatrix}.$$

Thus, by using Cramer's rule we obtain for the inverse matrix

$$\left(G^0\right)^{-1} = \frac{1}{\det(G^0)} \begin{pmatrix} \det(G^0_{-11}) & -\det(G^0_{-21}) & \dots & (-1)^{1+q}\det(G^0_{-q1}) \\ -\det(G^0_{-12}) & \det(G^0_{-22}) & \dots & (-1)^{2+q}\det(G^0_{-q2}) \\ \vdots & \ddots & & \vdots \\ (-1)^{1+q}\det(G^0_{-1q}) & (-1)^{2+q}\det(G^0_{-2q}) & \dots & \det(G^0_{-qq}) \end{pmatrix},$$

where G_{-ab} denotes the matrix G with the a-th row and b-th column omitted. Therefore, by applying Laplace's formula and using that $g_{ij} = g_{ji}$ we have

$$G^0 \odot \left(G^0\right)^{-1} = \frac{1}{\det(G^0)} \begin{pmatrix} g_{11}\det(G^0_{-11}) & -g_{12}\det(G^0_{-21}) & \dots & (-1)^{1+q}g_{1q}\det(G^0_{-q1}) \\ -g_{21}\det(G^0_{-12}) & g_{22}\det(G^0_{-22}) & \dots & (-1)^{2+q}g_{2q}\det(G^0_{-q2}) \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{1+q}g_{q1}\det(G^0_{-1q}) & (-1)^{2+q}g_{q2}\det(G^0_{-2q}) & \dots & g_{qq}\det(G^0_{-qq}) \end{pmatrix}.$$

Therefore,

$$G^0 \odot \left(G^0\right)^{-1}\eta = \frac{1}{\det(G^0)} \begin{pmatrix} \frac{1}{\sqrt{q}}\sum_{a=1}^{q}(-1)^{1+a}g_{1a}\det(G^0_{-a1}) \\ \frac{1}{\sqrt{q}}\sum_{a=1}^{q}(-1)^{2+a}g_{2a}\det(G^0_{-a2}) \\ \vdots \\ \frac{1}{\sqrt{q}}\sum_{a=1}^{q}(-1)^{q+a}g_{qa}\det(G^0_{-aq}) \end{pmatrix} = \frac{1}{\det(G^0)} \begin{pmatrix} \frac{\det(G^0)}{\sqrt{q}} \\ \frac{\det(G^0)}{\sqrt{q}} \\ \vdots \\ \frac{\det(G^0)}{\sqrt{q}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{q}} \\ \frac{1}{\sqrt{q}} \\ \vdots \\ \frac{1}{\sqrt{q}} \end{pmatrix}$$

and thus finally

$$\eta'\left(G^0 \odot \left(G^0\right)^{-1}\right)\eta = 1.$$

From this we can conclude that $(2\eta'\eta + 2\eta'(G^0 \odot (G^0)^{-1})\eta)^{1/2} = 2$ and $(2\eta'\eta - 2\eta'(G^0 \odot (G^0)^{-1})\eta)^{1/2} = 0$, which shows that the first term in (2.8) has the desired form. The second term of (2.8) equals the second term of the limit distribution of Qu (2011) anyway, so it remains to consider the

third term.
We first show that $\left(2\eta^{'}F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}+I_q)\Omega^{-1'}F(r)^{'}\eta\right)^{1/2}=2\int_0^r(1+\log s)^2ds$. To see this note that

$$\begin{aligned}F(r)^{'}\eta &= 2\int_0^r(1+\log s)^2\left(G^0\odot(G^0)^{-1}+I_q\right)^{'}\eta ds\\ &= 4\eta\int_0^r(1+\log s)^2ds\end{aligned}$$

as $\left(G^0\odot(G^0)^{-1}+I_q\right)^{'}\eta=2\eta$ by the same arguments as before. By denoting with η^{-1} the pseudo inverse defined by the equality $A\eta^{-1}\eta=A$ for every matrix A, we have $\eta^{-1}=\eta'$. Consequently,

$$\Omega^{-1'}\eta=\left(\eta^{-1}\Omega'\right)^{-1}=(\Omega\eta)^{'-1}.$$

Now, $\Omega\eta=4\eta$, since again $(G^0\odot(G^0)^{-1})\eta=\eta$, so that $\Omega^{-1'}\eta=1/4\eta$.
Applying the same arguments to the term $\eta^{'}F(r)\Omega^{-1}$ on the left side gives us altogether

$$\begin{aligned}2\eta^{'}F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}+I_q)\Omega^{-1'}F(r)^{'}\eta &= 2\left(\int_0^r(1+\log s)^2ds\right)^2\eta^{'}(G^0\odot(G^0)^{-1}+I_q)\eta\\ &= 4\left(\int_0^r(1+\log s)^2ds\right)^2.\end{aligned}$$

Now applying the same arguments we can furthermore conclude that

$$\eta^{'}F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}-I_q)\Omega^{-1'}F(r)^{'}\eta=0$$

which proves the lemma. □

Proof of Theorem 2.2:

To prove the consistency of our test statistic we closely follow the arguments in the consistency proof of Qu (2011). We use the property that $I(\lambda_j)=O_P(1)$ when $jT^{-1/2}\to\infty$. Note that ν_j is monotonically increasing in j with $\nu_1<0$ and $\nu_m>0$. For our test statistic we write

$$\begin{aligned}MLWS &= \frac{1}{2}\sup_{r\in[\varepsilon,1]}\left\|\frac{2}{\sqrt{\sum_{j=1}^m\nu_j^2}}\sum_{a=1}^q\eta_a\sum_{j=1}^{[mr]}\nu_j\left(_ag\left[Re\left[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\right]\right]_a-1\right)\right.\\ &\quad\left.+tr\left[G(\hat{d})^{-1}\frac{1}{\sqrt{\sum_{j=1}^m\nu_j^2}}\sum_{j=1}^{[mr]}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j(\hat{d}))^{-1}(-i_aI(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(\hat{d}))^{-1}\right]\right]\right\|\\ &= \sup_{r\in[\varepsilon,1]}\|I+II\|.\end{aligned}$$

Let us consider the term I first. Define $j^* = \min\{j : \nu_j \geq 0\}$. From Qu's (2011) Lemma B.1 it follows that $j^* = Km$ for some constant K. Define now

$$A^I = \left\| \sum_{a=1}^{q} \eta_a \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=j^*}^{m} \nu_j \left({}_a g \left[Re \left[\Lambda_j(\hat{d})^{-1} I(\lambda_j) \Lambda_j^*(\hat{d})^{-1} \right] \right]_a - 1 \right) \right\|$$

and

$$B^I = \left\| \sum_{a=1}^{q} \eta_a \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=1}^{j^*-1} \nu_j \left({}_a g \left[Re \left[\Lambda_j(\hat{d})^{-1} I(\lambda_j) \Lambda_j^*(\hat{d})^{-1} \right] \right]_a - 1 \right) \right\|.$$

Applying the reverse triangle inequality to I gives

$$I \geq \max(A^I - B^I, B^I).$$

Now, if $A^I \geq 2B^I$ than we have

$$\begin{aligned} I &\geq A^I - B^I \\ &= \frac{A^I}{2} + \left(\frac{A^I}{2} - B^I \right) \\ &\geq \tfrac{A^I}{2}. \end{aligned}$$

On the other hand, if $A^I < 2B^I$ than $I \geq B^I > A^I/2$. Altogether, we have $I \geq A^I/2$. Thus, we have to show that $A^I \xrightarrow{P} \infty$ if $T \to \infty$. To do so, we write A^I in the form

$$A^I = \left\| \sum_{a=1}^{q} \eta_a \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=j^*}^{m} \nu_j \left({}_a g \left[Re \left[\Lambda_j(d)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j) \Lambda_j^*(d)^{-1} \right] \right]_a - 1 \right) \right\|.$$

Applying the reverse triangle inequality to A gives

$$A^I \geq \sum_{a=1}^{q} \eta_a \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=j^*}^{m} \nu_j - \sum_{a=1}^{q} \eta_a \sum_{j=j^*}^{m} \nu_j \left({}_a g \left[Re \left[\Lambda_j(d)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j) \Lambda_j^*(d)^{-1} \right] \right]_a \right).$$

For the first term of this inequality we have for each component

$$\left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=j^*}^{m} \nu_j = m^{1/2} \int_K^1 (1 + \log s) ds + o\left(m^{1/2} \right)$$

which is strictly positive and of exact order $m^{1/2}$.
For the second term we can conclude the following. As $m/T^{1/2} \to \infty$ it holds that $j^*/T^{1/2} \to \infty$. Thus, for every $j^* \leq j \leq m$ it follows $I(\lambda_j) = O_P(1)$. Furthermore, we have

$$\nu_j\left({}_a g\left[Re\left[\Lambda_j(d)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)\Lambda_j^*(d)^{-1}\right]\right]_a\right) = O_P\left(\lambda_j^{2\hat{d}}\right) = O_P\left(\lambda_j^{2\epsilon}\right) = o_P(1)$$

for every a. This holds because $\nu_j = O_P(1)$ for $j^* \leq j \leq m$, $\hat{G}(d) = 1/m\sum_{j=1}^m Re[\Lambda_j(d)^{-1}I(\lambda_j)$ $\Lambda_j^*(d)^{-1}]$, $P(\hat{d}_j > \epsilon) \to 1$ and $\lambda_j = o(1)$. Therefore, the second term is of lower order than $m^{1/2}$ and is dominated asymptotically by the first term. Thus, $A^I \xrightarrow{P} \infty$ if $T \to \infty$.
The treatment of term II is similar to that of term I. Write again

$$A^{II} = \left\|\left(\sum_{j=1}^m \nu_j^2\right)^{-1/2} \sum_{j=j^*}^m \frac{\lambda_j - \pi}{2} Im\left[(\Lambda_j(\hat{d}))^{-1}(-i_a I(\lambda_j) + I(\lambda_j)i_a)(\Lambda_j^*(\hat{d}))^{-1}\right]\right\| \tag{2.16}$$

and

$$B^{II} = \left\|\left(\sum_{j=1}^m \nu_j^2\right)^{-1/2} \sum_{j=1}^{j^*-1} \frac{\lambda_j - \pi}{2} Im\left[(\Lambda_j(\hat{d}))^{-1}(-i_a I(\lambda_j) + I(\lambda_j)i_a)(\Lambda_j^*(\hat{d}))^{-1}\right]\right\|. \tag{2.17}$$

As before we obtain from the reverse triangle inequality that $II \geq A^{II}/2$. Now, we write A^{II} in the form

$$A^{II} = \left\|\left(\sum_{j=1}^m \nu_j^2\right)^{-1/2} (-\frac{\pi}{2}) \sum_{j=j^*}^m Im\left[\Lambda_j(d)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)\Lambda_j^*(d)^{-1}\right]\right\| \tag{2.18}$$

By using exactly the same arguments as before for the second part of A^I the term II is of lower order than $m^{1/2}$, and therefore dominated asymptotically by term I. This proves the theorem. □

Proof of Lemma 2.3:

The test statistic is given by

$$\begin{aligned}\widetilde{MLWS} \quad = \quad & \frac{1}{2}\sup_{r\in[\varepsilon,1]}\left\| \frac{2}{\sqrt{\sum_{j=1}^m \nu_j^2}} \sum_{a=1}^q \eta_a \sum_{j=1}^{[mr]} \nu_j \left({}_a\tilde{G}^{-1}(\hat{d},\hat{B})\left[Re\left[\Lambda_j(\hat{d})^{-1}\tilde{I}(\lambda_j,\hat{B})\Lambda_j^*(\hat{d})^{-1}\right]\right]_a - 1\right)\right. \\ & \left. + \frac{1}{\sqrt{\sum_{j=1}^m \nu_j^2}} \sum_{a=1}^q \eta_a \left({}_a\tilde{G}^{-1}(\hat{d},\hat{B})\right) \sum_{j=1}^{[mr]} \frac{(\lambda_j - \pi)}{2} Im\left[\Lambda_j(\hat{d})^{-1}\tilde{I}(\lambda_j,\hat{B})\Lambda_j^*(\hat{d})^{-1}\right]_a\right\|.\end{aligned}$$

Now, from $\hat{B} - B^0 = o_p(m^{-1/2}\Delta_m^{-1})$, we have $\tilde{I}(\lambda_j, \hat{B}) = \tilde{I}(\lambda_j, B^0) + o_p(m^{-1/2}\Delta_m^{-1})$ and $\tilde{G}(\hat{d}, \hat{B}) = \tilde{G}(\hat{d}, B^0) + o_p(m^{-1/2}\Delta_m^{-1})$. Therefore,

$$\widetilde{MLWS} = MLWS + \frac{\sum_{j=1}^{[mr]} o_p(m^{-1/2}\Delta_m^{-1})}{\sqrt{\sum_{j=1}^{m} \nu_j^2}},$$

and since $\sum_{j=1}^{m} \nu_j^2 \to m$, we have $\widetilde{MLWS} \to MLWS + o_p(\Delta_m^{-1}) = MLWS + o_p\left(T^{(\delta-1)\max\{b_1,\ldots,b_{p_G}\}}\right)$, for $m = \lfloor T^\delta \rfloor$. □

Proof of Lemma 2.4:

The proof of the lemma is similar to the proof of Corollary 2 in Qu (2011) after replacing his univariate coefficients of the short-memory model with our multivariate coefficients and using the respective norms. It is therefore omitted here. □

Supplementary Appendix

δ	DAX	NIKKEI	S&P 500	FTSE	DAX	NIKKEI	S&P 500	FTSE	DAX	NIKKEI	S&P 500	FTSE
	$\log(\lvert r_t \rvert + 0.001)$											
	d_{ELW}				d_{HP}				$d_{HP+Noise}$			
0.60	0.365	0.290	0.476	0.396	0.350	0.204	0.436	0.334	0.354	0.545	0.436	0.334
0.65	0.341	0.289	0.415	0.368	0.292	0.233	0.327	0.306	0.488	0.446	0.562	0.306
0.70	0.337	0.289	0.363	0.310	0.291	0.247	0.242	0.230	0.427	0.247	0.608	0.484
0.75	0.273	0.258	0.307	0.299	0.187	0.205	0.136	0.231	0.534	0.380	0.648	0.415
	$d_{LPWN}(0,0)$				$d_{LPWN}(1,0)$				$d_{LPWN}(0,1)$			
0.60	0.436	0.545	0.472	0.446	0.452	0.540	0.477	0.461	0.515	0.553	0.472	0.631
0.65	0.488	0.446	0.582	0.462	0.431	0.459	0.521	0.477	0.440	0.575	0.481	0.538
0.70	0.452	0.373	0.608	0.531	0.476	0.397	0.551	0.416	0.448	0.535	0.558	0.387
0.75	0.534	0.408	0.648	0.489	0.412	0.428	0.600	0.493	0.413	0.399	0.602	0.490
	$\log RV_t$											
	d_{ELW}				d_{HP}				$d_{HP+Noise}$			
0.60	0.625	0.640	0.642	0.622	0.639	0.546	0.637	0.629	0.639	0.546	0.637	0.661
0.65	0.593	0.646	0.577	0.608	0.567	0.597	0.558	0.588	0.648	0.597	0.660	0.668
0.70	0.598	0.640	0.577	0.612	0.558	0.595	0.546	0.583	0.620	0.595	0.621	0.639
0.75	0.580	0.601	0.554	0.583	0.515	0.531	0.498	0.538	0.642	0.673	0.630	0.662
	$d_{LPWN}(0,0)$				$d_{LPWN}(1,0)$				$d_{LPWN}(0,1)$			
0.60	0.642	0.679	0.637	0.661	0.619	0.683	0.550	0.664	0.642	0.802	0.637	0.681
0.65	0.653	0.652	0.660	0.670	0.641	0.654	0.655	0.655	0.644	0.652	0.644	0.640
0.70	0.643	0.652	0.623	0.643	0.637	0.661	0.616	0.652	0.643	0.665	0.623	0.661
0.75	0.646	0.677	0.632	0.662	0.650	0.672	0.615	0.659	0.652	0.675	0.621	0.663

Table 2.12: Estimated memory parameters for different bandwidths $m = \lfloor T^\delta \rfloor$ using the estimators of Shimotsu and Phillips (2005) (d_{ELW}), Hou and Perron (2014) (d_{HP} and $d_{HP+Noise}$) and Frederiksen et al. (2012) (d_{LPWN}).

This supplementary appendix contains additional empirical results and simulation studies on the performance of the MLWS test. Table 2.12 shows estimates of the memory parameters in our empirical application in Section 2.5. The exact local Whittle estimates (d_{ELW}) are very close to the standard local Whittle estimates presented in Tables 2.9 and 2.10. For the log-absolute returns the estimates using the LPWN estimator are considerably higher, as reported by Frederiksen et al. (2012). Estimates using the estimator of Hou and Perron (2014), on the other hand, are slightly reduced. When considering the log-realized volatility series, the estimates are extremely robust - irrespective of the bandwidth and the choice of the estimator. The following tables contain additional Monte Carlo simulations. All DGPs are given in the captions. To analyze the power of the MLWS test against non-causal alternatives, we build on the work of Kechagias and Pipiras (2015) who consider non-causal multivariate fractionally integrated processes. Table 2.13 shows the results. The scaling factor ζ determines to which degree the behaviour of the process is determined by the non-causal part. It can be seen that the test is correctly sized, but only develops power very slowly.

	ζ	0		1		2	
T	δ/ϵ	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.009	0.015	0.005	0.008	0.006	0.013
	0.65	0.014	0.018	0.008	0.012	0.011	0.024
	0.70	0.015	0.024	0.009	0.019	0.021	0.040
	0.75	0.018	0.027	0.011	0.020	0.030	0.058
500	**0.60**	0.012	0.016	0.010	0.017	0.011	0.024
	0.65	0.018	0.030	0.013	0.024	0.022	0.043
	0.70	0.022	0.033	0.014	0.032	0.049	0.076
	0.75	0.030	0.033	0.025	0.031	0.085	0.139
1000	**0.60**	0.016	0.029	0.014	0.021	0.020	0.038
	0.65	0.025	0.029	0.023	0.035	0.044	0.079
	0.70	0.029	0.038	0.027	0.048	0.103	0.136
	0.75	0.037	0.045	0.042	0.057	0.201	0.262
2000	**0.60**	0.025	0.032	0.019	0.040	0.029	0.060
	0.65	0.031	0.038	0.031	0.055	0.074	0.102
	0.70	0.031	0.044	0.051	0.074	0.174	0.222
	0.75	0.046	0.055	0.074	0.097	0.403	0.458

Table 2.13: Power of the MLWS test against non-causal alternatives of the form $X_t = D(0.2,0.3)^{-1}Q_-v_t + \tilde{D}(0.2,0.3)^{-1}\zeta Q_-v_t$, where $\tilde{D}(d_1,d_2) = diag((1-L^{-1})^{d_1},(1-L^{-1})^{d_2})$, $v_t \sim N(0,\Sigma_v)$, Σ_v is an identity matrix and $Q_- = ((1,0.7),(-0.5,1))$.

Table 2.14 shows the effect of breaks in the variance-covariance matrix of a bivariate fractionally integrated process. Initially the innovations have unit variance and a correlation of $\rho_1 = 0.45$. After 100ξ percent of the sample, the variance and correlation switch. It can be seen that the size of the test is robust to unconditional breaks in variance as well as breaks in the correlation between the innovations. Only for increases of the correlation after 20 percent of the sample, the size increases to about 7 percent. In all other combinations it remains close to or below the nominal level of 5 percent.
Table 2.15 shows the effect of conditional heteroscedasticity on the size and power of the test. We find that the size increases slightly, but only for very persistent GARCH effects. The power remains largely unchanged. The MLWS test is therefore robust to conditional and unconditional heteroscedasticity.

Random level shifts are not the only data generating processes that cause spurious long memory. Similar effects are caused (among others) by linear or monotonous trends, non-monotonous

		σ_2^2	1						2					
		ρ_2	0		0.45		0.9		0		0.45		0.9	
ξ	T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
	250	**0.60**	0.006	0.011	0.007	0.007	0.012	0.015	0.004	0.008	0.006	0.010	0.008	0.010
		0.65	0.010	0.014	0.008	0.014	0.016	0.025	0.009	0.013	0.007	0.015	0.012	0.018
		0.70	0.014	0.018	0.012	0.015	0.020	0.035	0.011	0.021	0.015	0.019	0.018	0.023
		0.75	0.012	0.020	0.015	0.019	0.031	0.039	0.015	0.024	0.017	0.019	0.020	0.026
0.2														
	1000	**0.60**	0.015	0.021	0.011	0.020	0.030	0.046	0.015	0.020	0.017	0.022	0.024	0.029
		0.65	0.021	0.030	0.017	0.027	0.039	0.057	0.021	0.024	0.019	0.031	0.032	0.041
		0.70	0.020	0.031	0.024	0.032	0.058	0.074	0.022	0.034	0.025	0.033	0.036	0.049
		0.75	0.031	0.040	0.031	0.037	0.064	0.077	0.032	0.039	0.033	0.043	0.045	0.051
	250	**0.60**	0.005	0.011	0.006	0.008	0.010	0.016	0.008	0.011	0.009	0.011	0.011	0.015
		0.65	0.011	0.018	0.009	0.017	0.013	0.025	0.013	0.020	0.014	0.016	0.014	0.025
		0.70	0.016	0.018	0.012	0.018	0.019	0.026	0.016	0.024	0.014	0.022	0.017	0.028
		0.75	0.020	0.021	0.014	0.021	0.026	0.035	0.018	0.029	0.020	0.026	0.024	0.030
0.5														
	1000	**0.60**	0.016	0.022	0.012	0.016	0.025	0.03	0.017	0.028	0.016	0.025	0.023	0.035
		0.65	0.023	0.024	0.022	0.022	0.031	0.044	0.028	0.035	0.024	0.030	0.031	0.039
		0.70	0.024	0.035	0.021	0.027	0.039	0.049	0.030	0.045	0.028	0.039	0.039	0.048
		0.75	0.033	0.043	0.025	0.034	0.046	0.053	0.040	0.050	0.033	0.049	0.044	0.055
	250	**0.60**	0.006	0.01	0.007	0.007	0.008	0.011	0.008	0.01	0.005	0.011	0.005	0.012
		0.65	0.010	0.015	0.009	0.013	0.011	0.019	0.012	0.019	0.010	0.014	0.012	0.016
		0.70	0.014	0.017	0.016	0.018	0.016	0.023	0.017	0.024	0.012	0.020	0.015	0.023
		0.75	0.017	0.025	0.015	0.019	0.019	0.025	0.021	0.029	0.019	0.025	0.020	0.027
0.8														
	1000	**0.60**	0.016	0.024	0.014	0.021	0.018	0.025	0.017	0.033	0.015	0.025	0.020	0.030
		0.65	0.019	0.029	0.017	0.026	0.025	0.034	0.029	0.040	0.020	0.036	0.023	0.034
		0.70	0.029	0.034	0.023	0.031	0.031	0.033	0.032	0.052	0.029	0.040	0.034	0.041
		0.75	0.034	0.038	0.027	0.037	0.032	0.043	0.046	0.060	0.038	0.043	0.035	0.050

Table 2.14: Size of the MLWS test in presence of breaks in the variance-covariance matrix. $D(0.2,0.3)X_t = v_t$, with $v_t \sim N(0,\Sigma_{v,t})$ and $\Sigma_{v,t} = ((1,0.45),(0.45,1))'\mathbb{I}_{(t/T<\xi)} + \sigma_2^2((1,\rho_2),(\rho_2,1))'(1-\mathbb{I}_{(t/T<\xi)})$.

trends or Markov switching models. Theoretically, it has been shown by McCloskey and Perron (2013) that the effect of these processes on the periodogram is approximately of the same order as that of random level shifts. Qu (2011) also demonstrates the power of his test against these alternatives. A similar analysis for the MLWS test is provided in Table 2.16. As can be seen, the test has good power against all of these alternatives.

In our empirical application we opt to conduct the test after using a univariate pre-whitening. This is to avoid convergence issues of the numerical optimization involved in the maximum likelihood estimation of the VARFIMA model, that could arise due to the large number of free parameters. As one can see from Table 2.17, the size of the test is controlled using univariate pre-whitening as it is with multivariate pre-whitening. With regard to the power, in smaller samples with up to 1000 observations univariate pre-whitening even leads to improvements. For $T = 2000$, on the other hand, we observe a considerable power loss compared to the multivariate

		Size						Power					
	β_1	0.4		0.6		0.8		0.4		0.6		0.8	
T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.007	0.011	0.005	0.010	0.007	0.015	0.104	0.117	0.108	0.122	0.101	0.134
	0.65	0.011	0.015	0.010	0.015	0.015	0.026	0.247	0.313	0.251	0.296	0.257	0.295
	0.70	0.015	0.017	0.018	0.019	0.022	0.028	0.450	0.476	0.448	0.461	0.453	0.471
	0.75	0.017	0.022	0.020	0.027	0.035	0.039	0.639	0.638	0.634	0.643	0.629	0.654
1000	**0.60**	0.014	0.023	0.014	0.024	0.020	0.031	0.687	0.719	0.679	0.713	0.669	0.695
	0.65	0.020	0.026	0.023	0.027	0.033	0.046	0.901	0.909	0.898	0.911	0.890	0.902
	0.70	0.025	0.030	0.025	0.032	0.051	0.062	0.972	0.968	0.968	0.967	0.966	0.967
	0.75	0.027	0.036	0.031	0.038	0.069	0.087	0.990	0.987	0.988	0.988	0.991	0.985

Table 2.15: Size and power of the MLWS test for processes with GARCH effects. $D(0.2,0.3)X_t = v_t$, with $\Sigma_{v,t} = \left((\sigma^2_{1,t}, 0.8\sigma_{1,t}\sigma_{2,t}), (0.8\sigma_{1,t}\sigma_{2,t}, \sigma^2_{2,t})\right)'$ and $\sigma^2_{a,t} = 1 + 0.15v^2_{a,t-1} + \beta_1\sigma^2_{a,t-1}$.

pre-whitening. This effect is particularly pronounced for DGPs 1 and 4, where we observe a non-monotonicity of the power in T. It is therefore advantageous to use the computationally more demanding multivariate pre-whitening in larger samples in terms of power. Nevertheless, if the computational costs become very large - as it is the case in our empirical example - the size is well controlled by the univariate pre-whitening procedure.

Since especially the log-absolute return series in the empirical application are often modelled as the underlying volatility process that is perturbed by a measurement error, Table 2.18 evaluates the effect of perturbations on the size of the MLWS test. One can observe that the size is largely robust. The only exception is the situation when the bandwidth and the sample are large and the series are strongly correlated. In this case, the size can reach up to 10 percent, if also the variance of the perturbation is large compared to that of the long memory process. Apart from this special case, the test remains robust.
The effect of perturbations on the performance of the pre-whitening procedure is explored in Table 2.19. It can be seen, that the size remains conservative for all DGPs. Interestingly, the size distortions that occur if no pre-whitening is applied are considerably reduced. One could therefore consider to omit the pre-whitening and to choose a smaller bandwidth if one has reason to assume that the series is subject to a sizeable perturbation - especially since we also observe a power loss after pre-whitening and an increase in power for the unfiltered series.

		DGP	mon		lin		sin		ms	
T	ζ	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	0	**0.60**	0.004	0.010	0.006	0.010	0.006	0.008	0.008	0.008
		0.65	0.008	0.011	0.009	0.015	0.009	0.015	0.008	0.013
		0.70	0.009	0.017	0.012	0.017	0.013	0.017	0.014	0.020
		0.75	0.013	0.020	0.012	0.021	0.015	0.018	0.012	0.020
	1	**0.60**	0.068	0.070	0.125	0.123	0.973	0.987	0.103	0.106
		0.65	0.114	0.120	0.227	0.246	0.999	1.000	0.182	0.197
		0.70	0.167	0.131	0.342	0.279	1.000	1.000	0.296	0.286
		0.75	0.193	0.135	0.427	0.330	1.000	1.000	0.427	0.406
	2	**0.60**	0.216	0.232	0.225	0.282	1.000	1.000	0.116	0.129
		0.65	0.409	0.444	0.505	0.582	1.000	1.000	0.226	0.267
		0.70	0.609	0.545	0.772	0.756	1.000	1.000	0.395	0.403
		0.75	0.751	0.663	0.912	0.870	1.000	1.000	0.596	0.618
1000	0	**0.60**	0.015	0.020	0.014	0.024	0.015	0.017	0.011	0.022
		0.65	0.016	0.029	0.019	0.027	0.016	0.021	0.018	0.026
		0.70	0.023	0.029	0.024	0.030	0.023	0.029	0.021	0.033
		0.75	0.027	0.036	0.027	0.037	0.025	0.036	0.024	0.034
	1	**0.60**	0.819	0.751	0.919	0.882	1.000	1.000	0.612	0.640
		0.65	0.938	0.872	0.990	0.971	1.000	1.000	0.841	0.855
		0.70	0.975	0.912	0.998	0.988	1.000	1.000	0.938	0.939
		0.75	0.991	0.948	1.000	0.996	1.000	1.000	0.962	0.973
	2	**0.60**	0.988	0.987	0.992	0.992	1.000	1.000	0.578	0.615
		0.65	1.000	0.999	1.000	1.000	1.000	1.000	0.838	0.873
		0.70	1.000	1.000	1.000	1.000	1.000	1.000	0.959	0.966
		0.75	1.000	1.000	1.000	1.000	1.000	1.000	0.978	0.981

Table 2.16: Power against alternative DGPs, for $Y_t = \mu_t + X_t$, with $D(0.2, 0.3)X_t = v_t$, $\Sigma_v = ((1,0),(0,1))'$ and $\mu_t = \zeta\left(t^{-\frac{t}{5T}} - T^{-1}\sum_{t=1}^{T} t^{-\frac{t}{5T}}\right)$ (mon), $\mu_t = \zeta\,(t/T - 1/2)$ (lin), $\mu_t = \zeta \sin\left(\frac{4\pi t}{T}\right)$ (sin) or $\mu_t = \zeta_t - T^{-1}\sum_{t=1}^{T}\zeta_t$ (ms), where ζ_t is a markov sequence that takes the values ζ or 0.

Size

		Pre-whitened: $\tilde{X}_t$								**Unfiltered: X_t**			
	DGP	1		2		3		4		1		2	
T	**δ/ϵ**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.007	0.010	0.011	0.011	0.009	0.008	0.008	0.011	0.005	0.011	0.005	0.01
	0.65	0.007	0.013	0.009	0.016	0.009	0.012	0.011	0.016	0.010	0.016	0.014	0.028
	0.70	0.010	0.012	0.010	0.012	0.008	0.007	0.016	0.015	0.015	0.018	0.040	0.070
	0.75	0.008	0.010	0.008	0.007	0.006	0.007	0.011	0.013	0.016	0.019	0.127	0.197
500	**0.60**	0.012	0.014	0.017	0.016	0.015	0.015	0.013	0.015	0.012	0.014	0.011	0.013
	0.65	0.013	0.022	0.015	0.020	0.012	0.020	0.019	0.024	0.014	0.022	0.019	0.037
	0.70	0.018	0.025	0.021	0.021	0.014	0.015	0.025	0.026	0.018	0.028	0.064	0.104
	0.75	0.023	0.019	0.015	0.018	0.011	0.010	0.028	0.028	0.025	0.028	0.240	0.317
1000	**0.60**	0.018	0.021	0.018	0.026	0.018	0.022	0.021	0.022	0.015	0.022	0.012	0.022
	0.65	0.023	0.027	0.026	0.026	0.027	0.023	0.028	0.030	0.026	0.030	0.022	0.037
	0.70	0.026	0.027	0.026	0.023	0.022	0.019	0.039	0.033	0.033	0.029	0.069	0.111
	0.75	0.027	0.025	0.025	0.027	0.018	0.015	0.038	0.035	0.040	0.043	0.356	0.471
2000	**0.60**	0.019	0.026	0.019	0.024	0.016	0.031	0.018	0.024	0.018	0.025	0.013	0.028
	0.65	0.025	0.034	0.023	0.033	0.026	0.031	0.024	0.034	0.026	0.034	0.022	0.038
	0.70	0.028	0.031	0.026	0.033	0.030	0.023	0.036	0.036	0.030	0.034	0.072	0.109
	0.75	0.033	0.036	0.033	0.029	0.022	0.016	0.049	0.043	0.041	0.046	0.442	0.525

Power

		Pre-whitened: $\tilde{X}_t$								**Unfiltered: X_t**			
	DGP	1		2		3		4		1		2	
T	**δ/ϵ**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.038	0.044	0.053	0.056	0.038	0.039	0.040	0.046	0.041	0.042	0.013	0.015
	0.65	0.077	0.091	0.089	0.105	0.051	0.059	0.090	0.105	0.090	0.111	0.020	0.027
	0.70	0.127	0.126	0.119	0.111	0.056	0.044	0.150	0.143	0.175	0.180	0.024	0.023
	0.75	0.170	0.170	0.118	0.106	0.047	0.033	0.198	0.179	0.325	0.329	0.024	0.037
500	**0.60**	0.109	0.100	0.118	0.102	0.071	0.066	0.101	0.107	0.120	0.120	0.066	0.059
	0.65	0.197	0.217	0.176	0.195	0.093	0.105	0.189	0.221	0.249	0.281	0.083	0.100
	0.70	0.319	0.330	0.263	0.255	0.115	0.116	0.325	0.343	0.470	0.506	0.092	0.083
	0.75	0.411	0.410	0.272	0.233	0.118	0.089	0.423	0.416	0.672	0.673	0.068	0.063
1000	**0.60**	0.201	0.202	0.261	0.261	0.143	0.150	0.184	0.189	0.311	0.337	0.227	0.219
	0.65	0.378	0.362	0.400	0.391	0.201	0.167	0.334	0.325	0.577	0.589	0.325	0.298
	0.70	0.508	0.468	0.484	0.425	0.198	0.129	0.463	0.438	0.779	0.773	0.323	0.275
	0.75	0.577	0.552	0.486	0.407	0.156	0.097	0.537	0.497	0.895	0.887	0.246	0.190
2000	**0.60**	0.239	0.234	0.534	0.553	0.349	0.345	0.267	0.268	0.611	0.643	0.523	0.516
	0.65	0.329	0.320	0.699	0.697	0.415	0.380	0.389	0.352	0.838	0.842	0.656	0.643
	0.70	0.389	0.366	0.751	0.707	0.377	0.269	0.430	0.380	0.939	0.928	0.678	0.603
	0.75	0.448	0.404	0.765	0.631	0.287	0.121	0.469	0.380	0.965	0.959	0.586	0.432

Table 2.17: Size and power of the MLWS test after univariate pre-whitening for DGPs 1 to 4 described in Section 2.4.3 with and without pre-whitening. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

		σ^2_ϑ	0								1								2							
		ρ_v	-0.8		0		0.4		0.8		-0.8		0		0.4		0.8		-0.8		0		0.4		0.8	
T	ρ_ϑ	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	-0.8	**0.60**	0.006	0.010	0.007	0.009	0.005	0.008	0.005	0.010	0.009	0.009	0.006	0.013	0.009	0.011	0.013	0.014	0.009	0.011	0.007	0.011	0.010	0.011	0.012	0.014
		0.65	0.009	0.015	0.008	0.016	0.007	0.014	0.008	0.018	0.015	0.018	0.016	0.021	0.013	0.019	0.017	0.030	0.015	0.020	0.014	0.019	0.019	0.025	0.019	0.025
		0.70	0.013	0.016	0.013	0.018	0.011	0.019	0.011	0.013	0.021	0.023	0.023	0.025	0.022	0.025	0.025	0.028	0.018	0.024	0.022	0.024	0.023	0.025	0.029	0.031
		0.75	0.016	0.024	0.016	0.024	0.017	0.020	0.015	0.021	0.024	0.029	0.025	0.027	0.030	0.028	0.037	0.036	0.025	0.031	0.030	0.029	0.028	0.030	0.034	0.041
	0	**0.60**	0.005	0.009	0.005	0.008	0.010	0.008	0.007	0.008	0.012	0.011	0.008	0.008	0.008	0.013	0.012	0.010	0.009	0.010	0.009	0.011	0.008	0.011	0.010	0.012
		0.65	0.008	0.015	0.007	0.015	0.010	0.015	0.009	0.012	0.016	0.021	0.012	0.017	0.014	0.017	0.019	0.022	0.016	0.021	0.013	0.018	0.013	0.018	0.016	0.019
		0.70	0.012	0.018	0.013	0.018	0.013	0.016	0.012	0.017	0.022	0.026	0.021	0.020	0.018	0.020	0.024	0.026	0.024	0.028	0.019	0.022	0.020	0.024	0.022	0.025
		0.75	0.017	0.023	0.016	0.019	0.018	0.021	0.014	0.019	0.026	0.033	0.025	0.029	0.024	0.027	0.032	0.033	0.025	0.027	0.025	0.027	0.028	0.031	0.027	0.030
	0.4	**0.60**	0.005	0.011	0.008	0.007	0.004	0.011	0.005	0.009	0.009	0.011	0.006	0.010	0.005	0.013	0.008	0.011	0.010	0.012	0.009	0.012	0.009	0.010	0.009	0.010
		0.65	0.010	0.013	0.009	0.015	0.009	0.014	0.011	0.015	0.014	0.023	0.011	0.014	0.010	0.015	0.013	0.022	0.022	0.023	0.015	0.022	0.017	0.018	0.015	0.020
		0.70	0.014	0.016	0.014	0.018	0.013	0.018	0.015	0.017	0.026	0.024	0.018	0.020	0.019	0.021	0.020	0.023	0.027	0.026	0.018	0.024	0.023	0.020	0.022	0.024
		0.75	0.013	0.023	0.012	0.020	0.016	0.021	0.016	0.023	0.034	0.035	0.024	0.023	0.023	0.027	0.026	0.028	0.033	0.033	0.030	0.031	0.022	0.026	0.030	0.027
	0.8	**0.60**	0.007	0.012	0.005	0.011	0.006	0.010	0.006	0.008	0.010	0.012	0.010	0.013	0.010	0.009	0.010	0.013	0.014	0.013	0.011	0.015	0.008	0.010	0.009	0.009
		0.65	0.010	0.015	0.009	0.015	0.008	0.014	0.011	0.015	0.021	0.024	0.013	0.022	0.015	0.016	0.016	0.017	0.018	0.025	0.015	0.020	0.013	0.019	0.016	0.019
		0.70	0.014	0.019	0.012	0.016	0.013	0.017	0.012	0.016	0.028	0.029	0.021	0.021	0.016	0.024	0.018	0.019	0.030	0.026	0.021	0.023	0.021	0.025	0.020	0.024
		0.75	0.014	0.024	0.014	0.026	0.016	0.022	0.017	0.020	0.037	0.038	0.029	0.027	0.024	0.026	0.028	0.027	0.035	0.038	0.027	0.033	0.026	0.029	0.027	0.025
1000	-0.8	**0.60**	0.014	0.019	0.012	0.021	0.015	0.021	0.013	0.017	0.020	0.023	0.016	0.021	0.022	0.027	0.028	0.031	0.026	0.027	0.023	0.027	0.023	0.031	0.031	0.035
		0.65	0.018	0.026	0.018	0.026	0.019	0.025	0.019	0.025	0.036	0.035	0.024	0.035	0.032	0.037	0.046	0.046	0.045	0.045	0.039	0.042	0.044	0.042	0.050	0.049
		0.70	0.020	0.030	0.025	0.027	0.023	0.031	0.023	0.028	0.044	0.050	0.045	0.043	0.046	0.049	0.064	0.069	0.065	0.057	0.056	0.053	0.067	0.057	0.077	0.067
		0.75	0.029	0.035	0.026	0.036	0.030	0.034	0.029	0.037	0.067	0.074	0.057	0.066	0.068	0.073	0.100	0.093	0.093	0.097	0.077	0.086	0.095	0.093	0.100	0.102
	0	**0.65**	0.014	0.021	0.013	0.019	0.014	0.018	0.013	0.020	0.027	0.031	0.020	0.026	0.020	0.024	0.029	0.032	0.027	0.029	0.021	0.024	0.021	0.024	0.032	0.029
		0.65	0.015	0.026	0.019	0.025	0.017	0.023	0.021	0.026	0.039	0.041	0.025	0.033	0.031	0.034	0.041	0.046	0.043	0.045	0.035	0.036	0.039	0.041	0.041	0.043
		0.70	0.023	0.030	0.022	0.031	0.021	0.027	0.026	0.029	0.062	0.059	0.040	0.048	0.043	0.045	0.060	0.056	0.068	0.063	0.052	0.05	0.053	0.058	0.064	0.059
		0.75	0.028	0.036	0.029	0.038	0.029	0.034	0.029	0.037	0.085	0.086	0.063	0.071	0.070	0.075	0.092	0.094	0.100	0.095	0.078	0.088	0.083	0.082	0.093	0.093
	0.4	**0.60**	0.014	0.018	0.014	0.021	0.013	0.019	0.013	0.017	0.029	0.033	0.018	0.023	0.018	0.024	0.026	0.027	0.031	0.033	0.024	0.026	0.021	0.028	0.026	0.030
		0.65	0.023	0.025	0.020	0.023	0.017	0.026	0.019	0.026	0.043	0.042	0.032	0.032	0.029	0.035	0.037	0.043	0.047	0.052	0.036	0.035	0.039	0.035	0.045	0.046
		0.70	0.023	0.027	0.023	0.031	0.025	0.031	0.023	0.032	0.057	0.058	0.041	0.043	0.044	0.043	0.050	0.057	0.063	0.068	0.052	0.050	0.052	0.057	0.059	0.065
		0.75	0.027	0.034	0.029	0.033	0.025	0.038	0.029	0.035	0.090	0.087	0.059	0.067	0.065	0.069	0.082	0.080	0.097	0.093	0.078	0.085	0.085	0.090	0.092	0.094
	0.8	**0.60**	0.017	0.021	0.017	0.022	0.012	0.020	0.013	0.021	0.030	0.031	0.018	0.023	0.018	0.020	0.019	0.025	0.030	0.036	0.022	0.026	0.022	0.028	0.026	0.031
		0.65	0.023	0.025	0.016	0.025	0.020	0.026	0.019	0.025	0.051	0.048	0.026	0.036	0.029	0.035	0.029	0.039	0.044	0.050	0.035	0.041	0.036	0.038	0.044	0.048
		0.70	0.022	0.029	0.022	0.033	0.024	0.025	0.023	0.029	0.062	0.058	0.040	0.039	0.044	0.044	0.051	0.046	0.071	0.062	0.050	0.050	0.051	0.053	0.058	0.058
		0.75	0.030	0.037	0.025	0.035	0.028	0.037	0.026	0.037	0.100	0.095	0.065	0.068	0.062	0.062	0.069	0.072	0.103	0.103	0.087	0.085	0.078	0.083	0.095	0.094

Table 2.18: Size of the MLWS test for perturbed processes, where $Y_t = X_t + \vartheta_t$, $\vartheta_t \sim N(0,\Sigma_\vartheta)$, with $\Sigma_\vartheta = \sigma^2_\vartheta((1,\rho_\vartheta),(\rho_\vartheta,1))'$ and $D(0.2,0.3)X_t = v_t$ with $v_t \sim N(0,\Sigma_v)$ and $\Sigma_v = ((1,\rho_v),(\rho_v,1))'$.

Size

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
	DGP	1		2		3		4		1		2	
T	δ/ϵ	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.007	0.007	0.006	0.007	0.005	0.009	0.008	0.011	0.008	0.011	0.005	0.011
	0.65	0.007	0.014	0.006	0.010	0.006	0.010	0.008	0.010	0.015	0.024	0.007	0.016
	0.70	0.007	0.012	0.008	0.009	0.006	0.006	0.009	0.009	0.019	0.027	0.015	0.020
	0.75	0.011	0.008	0.004	0.004	0.003	0.005	0.013	0.013	0.024	0.033	0.020	0.043
500	**0.60**	0.012	0.013	0.016	0.015	0.016	0.011	0.010	0.013	0.019	0.020	0.011	0.013
	0.65	0.014	0.016	0.014	0.022	0.009	0.016	0.012	0.019	0.027	0.032	0.012	0.023
	0.70	0.014	0.015	0.014	0.011	0.006	0.007	0.014	0.017	0.038	0.045	0.020	0.039
	0.75	0.012	0.014	0.005	0.005	0.004	0.003	0.015	0.012	0.054	0.056	0.039	0.069
1000	**0.60**	0.015	0.021	0.024	0.029	0.025	0.023	0.019	0.021	0.024	0.032	0.019	0.024
	0.65	0.022	0.024	0.035	0.027	0.027	0.022	0.019	0.024	0.049	0.042	0.021	0.028
	0.70	0.024	0.015	0.025	0.019	0.014	0.015	0.017	0.019	0.064	0.060	0.025	0.049
	0.75	0.020	0.020	0.017	0.015	0.005	0.004	0.018	0.017	0.088	0.086	0.068	0.095
2000	**0.60**	0.016	0.031	0.021	0.033	0.022	0.037	0.021	0.030	0.027	0.039	0.019	0.028
	0.65	0.037	0.033	0.034	0.041	0.033	0.046	0.030	0.033	0.049	0.066	0.020	0.032
	0.70	0.027	0.026	0.041	0.035	0.037	0.031	0.029	0.027	0.079	0.081	0.031	0.044
	0.75	0.025	0.019	0.029	0.024	0.019	0.009	0.024	0.016	0.139	0.132	0.083	0.118

Power

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
	DGP	1		2		3		4		1		2	
T	δ/ϵ	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.034	0.033	0.028	0.030	0.018	0.017	0.038	0.037	0.074	0.081	0.030	0.033
	0.65	0.061	0.079	0.041	0.050	0.019	0.027	0.062	0.074	0.171	0.196	0.050	0.066
	0.70	0.114	0.117	0.061	0.055	0.020	0.012	0.114	0.117	0.323	0.332	0.078	0.068
	0.75	0.181	0.186	0.066	0.057	0.017	0.011	0.176	0.172	0.494	0.496	0.113	0.097
500	**0.60**	0.110	0.107	0.081	0.078	0.071	0.052	0.106	0.104	0.227	0.209	0.115	0.091
	0.65	0.172	0.207	0.119	0.140	0.061	0.076	0.170	0.209	0.408	0.460	0.164	0.189
	0.70	0.282	0.296	0.159	0.160	0.062	0.049	0.290	0.305	0.650	0.677	0.237	0.243
	0.75	0.358	0.337	0.169	0.143	0.047	0.022	0.367	0.346	0.818	0.811	0.286	0.254
1000	**0.60**	0.297	0.315	0.236	0.213	0.165	0.162	0.304	0.306	0.477	0.491	0.304	0.298
	0.65	0.443	0.423	0.331	0.313	0.202	0.184	0.457	0.460	0.736	0.746	0.450	0.440
	0.70	0.538	0.484	0.382	0.327	0.181	0.115	0.556	0.513	0.888	0.888	0.544	0.483
	0.75	0.581	0.488	0.405	0.294	0.115	0.054	0.586	0.537	0.947	0.941	0.573	0.514
2000	**0.60**	0.620	0.638	0.481	0.497	0.382	0.390	0.615	0.640	0.763	0.784	0.597	0.617
	0.65	0.776	0.760	0.626	0.602	0.469	0.439	0.781	0.768	0.917	0.922	0.745	0.740
	0.70	0.780	0.728	0.653	0.568	0.448	0.298	0.802	0.755	0.967	0.963	0.802	0.752
	0.75	0.733	0.623	0.641	0.448	0.323	0.109	0.781	0.670	0.980	0.979	0.810	0.728

Table 2.19: Size and power of the MLWS test after pre-whitening as in Tables 2.4 and 2.5, but for perturbed series, where $Y_t = X_t + \vartheta_t$, $\vartheta_t \sim N(0, \Sigma_\vartheta)$, with $\Sigma_\vartheta = ((1,0),(0,1))'$ and X_t is given by DGPs 1 to 4 as described in Section 2.4.3.

Chapter 3

Model Order Selection for Seasonal/Cyclical Long Memory Processes

Model Order Selection for Seasonal/Cyclical Long Memory Processes

Co-authored with Philipp Sibbertsen.

Under revision for the Journal of Time Series Analysis.

3.1 Introduction

The increasing availability of high frequency data poses new challenges to the analysis of cyclical time series, since intraday data often exhibits periodic behavior and potentially contains multiple cycles such as daily and weekly seasonalities. Important examples of those series include intraday volatilities as discussed by Andersen and Bollerslev (1997), Bisaglia et al. (2003), Bordignon et al. (2008) and Rossi and Fantazzini (2015) as well as trading volumes and electricity data, where the aforementioned features are especially pronounced. Recent contributions such as Haldrup and Nielsen (2006), Soares and Souza (2006) and Diongue et al. (2009) stress the long memory properties of electricity time series and suggest that Gegenbauer models are useful to analyze these datasets, because they allow for different degrees of long memory at arbitrary periodic frequencies. An unresolved issue however, is how to select the number of cyclical components that have to be modeled. This is why we propose a model selection procedure that consistently estimates the required model order and demonstrate how it can be applied to the analysis of electricity load data.

Intuitively speaking, seasonal long memory is an intermediate case between short memory seasonal processes such as seasonal ARMA models and the seasonally integrated model. While a time series exhibits long memory if it has a hyperbolically decaying autocorrelation function - the autocorrelation functions of seasonal long memory processes show sinusoidal patterns with hyperbolically decaying amplitude, so that the dependence between observations at distant periodic lags is relatively strong.

In general, the term cyclicality refers to any kind of periodic behaviour in a time series that will cause a peak in its spectrum at the cyclical frequency γ and possibly also at its integer multiples - the harmonic frequencies. Seasonality on the other hand, is a special case of cyclicalities which refers to cycles with period lengths $S = 2\pi/\gamma$ that coincide with "natural" time intervals such as years, weeks or days.

The model class considered here is the $k-$factor Gegenbauer model given by

$$\prod_{a=1}^{k}(1-2\cos\gamma_a L+L^2)^{d_a}X_t = u_t, \tag{3.1}$$

where u_t is a linear short memory process with continuous, bounded and positive spectral density. The filter $(1-2xL+L^2)^{-d}$ is the generating function of the orthogonal Gegenbauer polynomials

denoted by $C_s^d(x)$:

$$(1-2xL+L^2)^{-d} = \sum_{s=0}^{\infty} C_s^d(x)L^s,$$

where the Gegenbauer polynomials are given by

$$C_s^d(x) = \sum_{g=0}^{\lfloor s/2 \rfloor} \frac{(-1)^g (2x)^{s-2g}\Gamma(d-g+s)}{g!(s-2g)!\Gamma(d)}.$$

In this representation the operator $\lfloor . \rfloor$ returns the integer part of its argument and $\Gamma(x)$ denotes the gamma function. The process defined by applying this filter to a white noise sequence v_t is a general linear process with the MA(∞)-representation

$$Y_t = (1-2\cos\gamma L+L^2)^{-d} v_t = \sum_{s=0}^{\infty} \varphi_s v_{t-s}, \tag{3.2}$$

where the coefficients φ_s are the Gegenbauer polynomials $C_s^d(\cos\gamma_a)$.
The spectral density of (3.1) is given by

$$f_X(\lambda) = f_u(\lambda) \prod_{a=1}^{k} |2(\cos\lambda - \cos\gamma_a)|^{-2d_a}. \tag{3.3}$$

As can be seen from (3.3) for $d_a > 0$ the spectral density has poles due to the long memory behavior at the cyclical frequencies γ_a with $a = 1,...,k$. Under appropriate parameter restrictions, the k-factor Gegenbauer model is nested in the seasonal/cyclical long memory (SCLM) model of Robinson (1994).
If u_t takes the form of a finite order ARMA process the model coincides with the k-factor GARMA model (GARMA-k) proposed by Gray et al. (1989) and generalized by Giraitis and Leipus (1995) and Woodward et al. (1998). Giraitis and Leipus (1995) show that the GARMA-k model is causal, invertible and has long memory, if $|d_a| < 1/2$, $\forall\ \gamma_a \in (0,\pi)$ and if $|d_a| < 1/4$, $\forall\ \gamma_a \in \{0,\pi\}$. The dependence of the stationarity conditions on the cyclical frequency γ_a is due to a non-uniform power law of the spectral density that will be discussed in more detail in Section 3.4. GARMA models generalize the ARFIMA class by allowing for poles in the spectral density at arbitrary frequencies γ_a and they nest most of the seasonal long memory models proposed in the literature such as the (rigid) SARFIMA model of Porter-Hudak (1990) or the flexible SARFIMA of Hassler (1994).
It should be noted, that even though the Gegenbauer model assumes cyclicality of a sinusoidal form, it can also be used to model non-sinusoidal cycles if the model order is increased. The reason for this is, that non-sinusoidal cycles affect the periodogram not only at the cyclical frequency γ_a, but also at its harmonics. Therefore, they can be accounted for by adding additional

Gegenbauer filters at the harmonics of γ_a.
One method to choose the model order k and the locations γ_a of the poles is based on the LM tests of Robinson (1994) and Hassler et al. (2009) who test whether the sample supports a given specification of (3.1). The null hypotheses are of the form $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$, where $\theta = \{\gamma_1, d_1, ..., \gamma_k, d_k\}'$. These procedures are useful in two situations. First, if the theory suggests a model for the process considered, or the researcher wants to test whether one of the nested special cases such as a (rigid) SARFIMA fits the data. Second, if the tests are applied on a grid of values for θ to obtain "confidence sets" that contain the true model with a certain coverage probability as suggested by Hassler et al. (2009). This second application of the LM-tests allows to specify the location of periodic frequencies as well as their number, because θ implicitly contains the model order k. The set of models that is not rejected should thus contain models of the true model order k_0. This grid search procedure has become the most common specification method. Examples of its application include Gil-Alana (2002) and Gil-Alana (2007).
For larger k however, this model selection procedure suffers from severe dimensionality problems. Consider a sample of T observations. Assume that the grid for γ_a has as many points as there are periodogram ordinates and let n_d denote the number of values on the grid for the respective d_a. Then the number of points on the combined grid for a k-factor model is $n_d^k \prod_{a=1}^{k} \{\lfloor T/2 \rfloor - (a-1)\}$, which is $O\left(\{n_d \lfloor T/2 \rfloor\}^k\right)$, so that the procedure quickly becomes unapplicable for models with a larger number of relevant cyclical frequencies. In these situations the model order k is usually selected discretionary based on a visual inspection of the periodogram. Unfortunately, this often leads to misspecifications as demonstrated below in our empirical application.

To overcome this problem, we suggest a model order selection procedure based on iterative filtering and periodogram based tests for persistent cyclical behavior in a time series. The procedure is based on the observation that the residual series from a correctly specified model for X_t given in (3.1) with $k = k_0$ Gegenbauer filters has no poles in the periodogram. If the selected model order $k < k_0$ is too low, on the other hand, the periodogram of the filtered process still exhibits poles. Consequently, we can apply k-factor Gegenbauer filters of increasing order k, until no significant periodicity can be detected anymore. A similar sequential testing procedure has been mentioned by Hidalgo and Soulier (2004). However, the performance of this procedure has not been thoroughly studied, yet. We therefore include it in our Monte Carlo analysis as a benchmark.
Section 3.2 discusses our model order selection procedure in more detail. To make the procedure feasible, we need a test that can be used after each filtering step to determine whether the residual process still contains significant persistent periodicity. Such a test for periodicity of unknown frequency is suggested in Section 3.3. We also need estimators for $f(\lambda)$, γ_a and d_a that will be discussed in Section 3.4. Subsequently, we provide a Monte Carlo analysis of the finite sample properties of our model order selection procedure and apply it to a Californian electricity load series.

3.2 Infeasible Automatic Model Selection by Sequential Filtering

To introduce the main idea of our procedure let the partial parameter vector $\theta_i = (\gamma_1, ..., \gamma_{k^{(i)}}, d_1, ..., d_{k^{(i)}})'$ contain the true cyclical frequencies γ_a and the respective memory orders d_a at these frequencies for some non-negative integer $k^{(i)} \leq k_0$, where k_0 denotes the true model order. In addition to that, denote the Gegenbauer filter by $GG(\gamma_a, d_a) = (1 - 2\cos\gamma_a L + L^2)^{d_a}$ and define $\Delta^{k^{(i)}}(\theta_i) = \prod_{a=1}^{k^{(i)}} GG(\gamma_a, d_a)$. Then

$$\Delta^{k^{(i)}}(\theta_i)X_t = \begin{cases} \prod_{a=k^{(i)}+1}^{k_0} GG(\gamma_a, -d_a)u_t, & \text{if } k^{(i)} < k_0 \\ u_t, & \text{if } k^{(i)} = k_0 \\ \prod_{a=k^0+1}^{k^{(i)}} GG(\gamma_a, d_a)u_t, & \text{if } k^{(i)} > k_0. \end{cases} \tag{3.4}$$

As one can see, if the $k^{(i)}$-factor Gegenbauer filter $\Delta^{k^{(i)}}(\theta_i)$ is applied to the observed series X_t and $k^{(i)} < k_0$, than the filtered process is a Gegenbauer process of order $k_0 - k^{(i)}$. If $k^{(i)} = k_0$ on the other hand, the filtered series is simply the short memory process u_t. Finally, if a filter of an order $k^{(i)} > k_0$ is applied to the series, the residual process displays cyclical antipersistence with $k^{(i)} - k_0$ zeros in its spectrum at the frequencies corresponding to the redundant filters.

This is the observation that our model selection procedure is based upon. Since the spectrum of the short memory process and that of the residuals from an overspecified model are bounded, a test for the null hypothesis that the spectrum is bounded will reject as long as the model is underspecified, because the $k_0 - k^{(i)}$-factor Gegenbauer process still exhibits poles in the spectrum. Assume for now that such a test statistic exists and denote it by G^*. We can than use this test in a sequential procedure to determine the model order of the Gegenbauer process. Starting with the observed series X_t, we test the null hypothesis, that X_t is short memory with $k^{(1)} = 0$. If this is the case, a test against poles in the spectrum will not be rejected and the procedure terminates. But if the true model order k_0 is larger than $k^{(1)}$, the test will reject. We then remove the largest peak from the spectrum by applying the corresponding Gegenbauer filter $GG(\gamma_a, d_a)$. If X_t is of order $k^{(2)} = 1$, the residuals obtained will be short memory, so that the G^*-statistic will no longer reject. Otherwise we proceed with an additional filtering step. Formally, we set $k^{(1)} = 0$ and test the null hypotheses

$$H_0 : \ k_0 = k^{(i)} \text{ vs. } H_1 : \ k_0 > k^{(i)} \tag{3.5}$$

repeatedly for $k^{(i)} = i - 1$ until the null hypothesis cannot be rejected anymore. The smallest $k^{(i)}$ for which the null hypothesis of no significant periodicity cannot be rejected anymore is than selected as the model order:

$$\hat{k} = \min \left\{ k^{(i)} \text{ such that } G^* < g_{crit} \right\}, \tag{3.6}$$

where g_{crit} is the critical value for G^*.

To make this procedure feasible, we need estimators for γ_a and d_a that are consistent under the null hypothesis as well as under the alternative - otherwise the procedure could not be applied sequentially. The latter can be achieved using semiparametric estimators for the fractional exponent d_a as in Arteche and Robinson (2000) and semiparametric estimators of the location of the pole γ_a as in Yajima (1996) and Hidalgo and Soulier (2004). A test G^* that allows to test whether there is unmodeled cyclical long memory in the process is presented in the next section.

3.3 Testing for Periodicity of Unknown Frequency

Note that under the null hypothesis in equation (3.5) the filtered process $\Delta^{k^{(i)}}(\theta_i)X_t$ is a short memory process. Define the periodogram of the weakly dependent linear process $Z_t = \sum_{j=0}^{\infty} a_j z_{t-j}$ with $z_t \overset{iid}{\sim} (0, \sigma_z^2)$ and $\sum_{j=0}^{\infty} |a_j| < \infty$ as

$$I(\lambda) = (2\pi T)^{-1} \left| \sum_{t=1}^{T} Z_t e^{-\mathbf{i}\lambda t} \right|^2, \quad \text{with} \quad \lambda \in [-\pi, \pi].$$

Here [.] denotes the closed interval. Further denote by I_j the periodogram evaluated at the Fourier frequency $\lambda_j = \frac{2\pi j}{T}$ with $j = 1, ..., n$ and $n = \lfloor (T-1)/2 \rfloor$.
Tests for periodicity at an unknown frequency are based on the well-known result that the periodogram ordinates of weakly dependent linear processes are approximately equal to $f(\lambda_j)$ times an exponentially distributed random variable with mean one:

$$\frac{I_j}{f(\lambda_j)} \overset{appr.}{\sim} Exp(1). \tag{3.7}$$

A detailed discussion of traditional periodogram based tests for periodicity can be found in Priestley (1981), pp. 406-415. We will focus on the large sample g-test of Walker (1914) for the null hypothesis of a Gaussian *iid*-sequence ($H_0 : Z_t = z_t$). In this case the relationship in (3.7) holds without an approximation error and we have $f(\lambda) = \sigma_z^2/(2\pi)$. Walker's large sample g-statistic is based on the maximum of the periodogram and it is defined as

$$g_Z^* = \frac{2\pi \max(I_j)}{\hat{\sigma}_z^2}, \tag{3.8}$$

where $\hat{\sigma}_z^2$ is a consistent estimator of σ_z^2. Due to the consistency of $\hat{\sigma}_z^2$, the distribution of g_Z^* is asymptotically the same as if σ_z^2 was known. Since the periodogram ordinates I_j are independent for *iid*-sequences, the distribution of g_Z^* for every fixed n is given by

$$p(g_Z^* > \tilde{z}) = 1 - (1 - \exp(-\tilde{z}))^n,$$

which converges suitably standardized to a Gumbel distribution (cf. Johnson et al. (1995), pp.

4-11).

As shown in Proposition 3.1 below, Walker's g-test can be extended to test the null hypothesis of a stationary weakly dependent process against the alternative of cyclical long memory, if $\hat{\sigma}_z^2$ in (3.8) is replaced by a consistent estimate of the spectral density $f(\lambda)$. In finite samples the distribution of the periodogram ordinates normalized by the spectral density is unknown for non-iid series, so that we can no longer determine the exact distribution of the test statistic. Asymptotically however, the periodogram ordinates are independent and exponential distributed (cf. Giraitis et al. (2012), Theorem 5.3.1). We therefore have to work with the limiting Gumbel distribution and normalize the test statistic by subtracting $\log(n)$. Then the modified G^*-statistic is defined as

$$G_Z^* = \max\left\{\frac{I_j}{\hat{f}_Z(\lambda_j)}\right\} - \log n. \tag{3.9}$$

To obtain consistency of the test against cyclical long memory effects, we require the following assumptions.

Assumption 3.1. *The fractional exponents are restricted to* $0 < d_a < 1/2$ *for all* $\gamma_a \in (0,\pi)$ *and* $0 < d_a < 1/4$ *for* $\gamma_a \in \{0,\pi\}$.

Assumption 3.2. *The fractional exponents* d_a *in (3.1) are bounded away from zero:* $d_a > c_a > 0$, *where* c_a *is a small constant* $\forall\ a = 1,...,k$.

Assumption 3.3. *For* $k^{(i)} = k_0$, $\hat{f}_Z(\lambda) \xrightarrow{p} f_Z(\lambda)$ *and for* $k^{(i)} < k_0$ *the estimator* $\hat{f}_Z(\lambda)$ *is bounded for all* $\lambda \in [0,\pi]$.

First of all, under Assumption 3.1 the process is stationary, so that the spectral density is well defined. Assumption 3.2 guarantees the identifiability of the poles and is the same as in Hidalgo and Soulier (2004). Finally, we require the spectral density estimate to be consistent under the null hypothesis and bounded under the null and the alternative. For the modified G^*-test we then obtain the following result.

Proposition 3.1. *For* $\Delta^{k^{(i)}}(\theta_i)X_t$ *characterized by (3.4):*

1. *If* $k^{(i)} = k_0$ *and under Assumptions 3.1 and 3.3, we have* $\lim_{T\to\infty} P(G^* > \tilde{z}) = 1 - \exp(-\exp(-\tilde{z}))$.
2. *If* $k^{(i)} < k_0$ *and under Assumptions 3.1, 3.2 and 3.3, we have* $\lim_{T\to\infty} P(G^* < \tilde{c}) = 0$, *for all* $\tilde{c} < \infty$.

Proofs of the main results are given in the appendix. Proposition 3.1 establishes the limiting Gumbel distribution of the G^*-statistic under the null hypothesis and the consistency under the alternative. We can therefore use this test to determine whether all significant periodicities have been removed after the i-th iteration step.

Note that the common g^* statistic has power against weakly dependent processes as well as against cyclical long memory processes. The G^*-test on the other hand is constructed so that it is robust to weak dependence which makes it suitable for the selection of the order in model (3.1).

3.4 Local Semiparametric Estimators of Cyclical Frequencies and Memory Parameters

We derived the iterative model selection procedure based on the modified G^*-test under the assumption that consistent estimators of the spectral density $f(\lambda)$, the cyclical frequencies γ_a and the memory parameters d_a are available. In this section, we provide a discussion of the relevant methods available from the literature and the necessary modifications. Details on the selection of bandwidth parameters in practical applications will be discussed in Section 3.5 and are therefore omitted here.

First consider the estimation of the spectral density. Usually $f(\lambda)$ is estimated through kernel-smoothed versions of the periodogram. For our purpose, however, it is important that the estimate $\hat{f}(\lambda_j)$ is very smooth in small samples and a single large spike in the periodogram has little impact on the estimated spectrum in its immediate neighborhood. This is why we use a logspline spectral density estimate as proposed by Cogburn and Davis (1974), who showed that this estimator is asymptotically equivalent to a kernel spectral density estimate. In particular, we use the maximum likelihood logspline spectral density estimator of Kooperberg et al. (1995) and regression splines.

Following Kooperberg et al. (1995), define $A_h = [(h-1)\pi/H_T, h\pi/H_T)$ as a subinterval of $[0,\pi]$, for $1 \leq h < H_T$ and set $A_{H_T} = \pi$. The function g defined on $[0,\pi]$ is a spline function if it is a polynomial of degree ν on each subinterval A_h and if it is $q_\nu = max(0,\nu-1)$ times continuously differentiable on $[0,\pi]$.[1] A spline function can be expressed in terms of Basis-splines denoted by B_i, $1 \leq i \leq I$ where $I = (\nu+1)H_T - (q_\nu+1)(H_T-1)$ such that

$$g(\lambda,\beta) = \beta_1 B_1(\lambda) + ... + \beta_I B_I(\lambda).$$

For further details on B-splines cf. De Boor (1978). The basis for the application of splines in spectral density estimation is the result in (3.7). For weakly dependent linear processes Z_t, the periodogram ordinates are asymptotically uncorrelated and $I_j = f(\lambda_j)Q_j$, where Q_j is exponential distributed with mean one. For the logarithm of I_j follows that $\log I_j = \varphi(\lambda_j) + q_j$, where q_j is the log of the exponential variable Q_j and $\varphi(\lambda_j)$ is the log-spectral density. This linearization allows to apply a spline function $g(\lambda,\beta)$ to estimate $\varphi(\lambda)$. The spectral density estimate $\hat{f}(\lambda) = \exp(g(\lambda,\hat{\beta}))$ is then obtained after reversing the log-transformation.

Kooperberg et al. (1995) suggest a maximum likelihood estimator for $f(\lambda)$ that is directly based

[1] It is common to use cubic splines, where the degree of the local polynomials is $\nu = 3$. This is also the case which we consider in our simulation studies and empirical application.

on the periodogram. They show that from (3.7) the observed periodogram ordinate I_j has the log-density

$$\begin{aligned}\psi(I_j,\beta) &\propto \left[\varphi(\lambda_j) + I_j \exp\left(-\varphi(\lambda_j)\right)\right] \\ &\approx \left[\sum_{i=1}^{I}\beta_i B_i(\lambda_j) + I_j \exp\left(-\sum_{i=1}^{I}\beta_i B_i(\lambda_j)\right)\right].\end{aligned}$$

Therefore, the log-likelihood of the observed periodogram ordinates is given by $l(\beta|I_1,...,I_n) = \sum_{j=1}^{n}\psi(I_j,\beta)$. Since the spectral density is symmetric around zero and periodic, β is restricted to the subspace Ω of $\mathbb{R}^I$ such that $g'(0) = g''(0) = g'(\pi) = g''(\pi) = 0$. The number of segments in the partition $(A_1,...,A_{H_T})$ is determined according to $H_T = \lfloor 1 + T^{\zeta} \rfloor$, with $0 < \zeta < 1/2$. The maximum likelihood estimator is given by

$$\hat{\beta}_{ML} = \arg\max_{\beta\in\Omega} l(\beta). \tag{3.10}$$

For a linear process $Z_t = \sum_{j=1}^{\infty} a_j z_{t-j}$, with $z_t \sim N(0,\sigma_z^2)$ where $\sum_{j=1}^{\infty}|a_j||j|^p < \infty$ for some $p > 1/2$, Kooperberg et al. (1995) show that $\|\hat{\varphi} - \varphi\| = O_p(\sqrt{I/T} + I^{-p})$. So that the consistent spectral density estimate required in Proposition 3.1 is available. Since the spline function $g(\lambda,\beta)$ is continuous and differentiable on $[-\pi,\pi]$ by definition, it is also bounded as required for Proposition 3.1. To enforce the symmetry condition $g'(0) = g''(0) = 0$ at the zero frequency, we maximize the likelihood $l(\beta|I_{-n},...,I_n) = \sum_{j=-n}^{n}\psi(I_j,\beta)$ over all Fourier frequencies in $[-\pi,\pi]$ for a periodic spline, so that also $g'(\pi) = g''(\pi) = 0$. Accordingly, the number of B-splines I is determined using $2H_T$ segments. Due to the symmetry of the spectrum this does not affect the consistency or limit distribution.

Even though this maximum likelihood estimator is asymptotically the most efficient, the estimation can be time-consuming for large datasets. Alternatively a regression spline can be used that is defined by

$$\hat{\beta} = \arg\min_{\beta\in\Omega}\sum_{j=-n}^{n}\left[\log(I_j) + \eta - g(\lambda_j,\beta)\right]^2, \tag{3.11}$$

where the Euler-Mascheroni constant η enters the loss function because $E(\log(I_j)) = \log(f(\lambda_j)) - \eta$. Since this OLS estimator has a closed form solution, this approach does not require numerical optimization. Here, we use the OLS estimator to determine starting values for the numerical optimization procedure used by the ML estimator.

With regard to the estimators for the location parameters γ_a and the memory parameters d_a the following assumptions are required in addition to Assumption 3.2 to ensure consistency.

Assumption 3.4. *For every frequency $\gamma_a \in [0,\pi]$, there exists an $\alpha^* \in (0,2]$, such that as $\lambda \to 0$, $f(\gamma_a + \lambda) = G_a\lambda^{-2\delta d_a}(1 + O(\lambda^{\alpha^*}))$, where $G_a \in (0,\infty)$ and $\delta = 2$ if $\gamma_a \in \{0,\pi\}$ and $\delta = 1$, otherwise. Additionally, in a neighbourhood $(-\delta,0)\cup(0,\delta)$ of γ_a, $f(\lambda)$ is differentiable and $|\frac{d}{d\gamma}f(\gamma_a + \lambda)| =$*

$O(\lambda^{-1-2d_a})$, *as* $\lambda \to 0$.

Assumption 3.5. *The bandwidth* m *satisfies* $\frac{1}{m} + \frac{m^{1+2\alpha^*}(\log(m))^2}{n^{2\alpha^*}} \to 0$, *as* $T \to \infty$.

Assumption 3.4 is required to ensure that the spectral density and its derivative are sufficiently smooth and neighbouring poles are sufficiently far away from each other so that their effects on the periodogram do not interfere. This is required in addition to Assumption 3.2 for an estimation of the location of the cyclical frequencies with the method of Hidalgo and Soulier (2004). Assumption 3.5 is a modification of Assumption B.4 in Arteche and Robinson (2000) that they discuss for the case of symmetric poles. The parameter m is used by the generalized local Whittle estimator that will be introduced below. This generalized local Whittle estimator of Arteche and Robinson (2000) is consistent for the k-factor Gegenbauer process (3.1) under Assumptions 3.4 and 3.5.

If a significant periodicity is detected, we have to estimate the frequency at which it occurs. The maximum of the periodogram is shown to be a consistent semiparametric estimator for the cyclical frequencies γ_a in (3.1) by Yajima (1996) under Gaussianity and by Hidalgo and Soulier (2004) if Assumptions 3.1 and 3.2 are fulfilled and the innovation process u_t has finite 8-th moments. To estimate γ_a we therefore use

$$\hat{\gamma}_a = \arg\max_{\lambda_j} I_{k^{(i)}}(\lambda_j), \tag{3.12}$$

where $I_{k^{(i)}}(\lambda_j)$ denotes the periodogram of the residual process $\Delta^{k^{(i)}} X_t$. Note that due to the sequential nature of our model selection procedure, we only estimate the location of one of the remaining $(k^{(i)} - k_0)$ poles in every iteration of the procedure. However, the order in which the poles are identified has no effect on the consistency of the procedure since the Gegenbauer filters are commutative.

To estimate the fractional exponents d_a, we use a generalized local Whittle approach similar to that suggested by Arteche and Robinson (2000). Since they consider an asymmetric model that allows for different fractional exponents on each side of the pole, they estimate the exponents separately with a generalized local Whittle estimator using m frequencies on the respective side of the pole. For our filtering procedure we determine the bandwidth $m = \lfloor 1 + T^{\xi} \rfloor$ by changing the bandwidth parameter ξ that is subject to $0 < \zeta < \xi < 1$, where ζ is the bandwidth parameter that determines the number of knots in the logspline spectral density estimation.

For γ_a close to 0 or π, there can be less than m Fourier frequencies on the respective side of γ_a that is close to the boundary. Therefore, we conduct the estimations using $m_q \leq m$ periodogram ordinates on the respective side of γ_a, where m_q is the maximal number of Fourier frequencies less than or equal m that is available at that side of the pole. The estimator is defined as

$$\tilde{d}_a = \arg\min_{d_a}(R_1(d_a, m_1) + R_2(d_a, m_2)) \tag{3.13}$$

for $q = 1, 2$, where $R_q(d_{a,q}, m_q) = \log \tilde{C}_q(d_{a,q}) - \frac{2d_{a,q}}{m_q} \sum_{j=1}^{m_q} \log \lambda_j$, $\tilde{C}_1(d_{a,1}) = \frac{1}{m_1} \sum_{j=1}^{m_1} \lambda_j^{2d_{a,1}} I(\gamma_a + \lambda_j)$ and

$\tilde{C}_2(d_{a,2}) = \frac{1}{m_2} \sum_{j=1}^{m_2} \lambda_j^{2d_{a,2}} I(\gamma_a - \lambda_j)$.

Even though Arteche and Robinson (2000) assume that the locations of the poles are known, Hidalgo and Soulier (2004) show for the GPH estimator that the estimation of the cyclical frequencies does not affect the limit distribution of the estimator, which suggests that the same is the case for the local Whittle estimator.

Note that the estimator of Arteche and Robinson (2000) is restricted to $\gamma_a \in (0, \pi)$. As Hidalgo and Soulier (2004) point out, the power law of the spectral density is $|\lambda - \gamma_a|^{-2\delta d_a}$, where $\delta = 2$ if $\gamma_a \in \{0, \pi\}$ and $\delta = 1$ otherwise. For this reason let $j^*(T)$ denote a positive integer that is increasing in T but very slowly such that $j^*/T \to 0$ as $T \to \infty$. Define

$$\hat{d}_a^* = \hat{d}_a / \hat{\delta}, \tag{3.14}$$

where $\hat{\delta} = 1 + I_{(\hat{\gamma}_a < \lambda_{j^*} \text{ or } \hat{\gamma}_a > \pi - \lambda_{j^*})}$. The function $I_{(\hat{\gamma}_a < \lambda_{j^*} \text{ or } \hat{\gamma}_a > \pi - \lambda_{j^*})}$ is an indicator function that takes on the value 1, if $\hat{\gamma}_a$ is one of the j^* Fourier frequencies closest to 0 or closest to π. Since $\frac{j^*}{T} + \frac{1}{j^*} \to 0$, we have $\lambda_{j^*} \to 0$ and $\lambda_{n-j^*} \to \pi$, so that $\hat{\delta}$ is a consistent estimator for the power law coefficient δ.

From Theorem 2 in Arteche and Robinson (2000) and Assumptions 3.4 and 3.5 follows immediately that we have $\hat{d}_a^* \xrightarrow{p} d_a$ for $d_a \in (-1/2, 1/2)$. For the case of a pole at 0, Velasco (1999) shows, that the consistency extends to the interval $d_a \in (-1/2, 1)$ under conditions very similar to those in Robinson (1995a) and Arteche and Robinson (2000). This suggests that our procedure could also be applied within this interval. For poles at the origin, both Shimotsu and Phillips (2005) and Abadir et al. (2007) suggest modified local Whittle estimators that are consistent and asymptotically normal for non-stationary linear processes, so that our approach could be extended along these lines as well.

3.5 Feasible Automatic Model Order Selection

With the estimators discussed in the previous sections the infeasible sequential filtering procedure becomes feasible. It is carried out in the following steps:

Step 0: Initialize the procedure with $i = 1$.

Step 1: Set $k^{(i)} = i - 1$ and apply the filter $\Delta^{k^{(i)}}$ to the time series X_t.

Step 2: Test whether there are any significant poles in the spectrum of $\Delta^{k^{(i)}} X_t$ using the modified G^*-test in (3.9). Proceed to Step 3 if the null hypothesis $H_0 : k_0 = k^{(i)}$ is rejected - otherwise go to Step 5.

Step 3: Estimate γ_a and d_a using the estimators defined in (3.12) and (3.14).

Step 4: Set $i = i + 1$ and go back to Step 1.

Step 5: Estimate k with the estimator given in (3.6).

Due to the consistency of the modified G^*-test established in Proposition 3.1, it asymptotically has a power of 1 if $k^{(i)} < k_0$, so that the test rejects until $k^{(i)} = k_0$. If $k^{(i)} = k_0$, then there is a probability of α that a type I error occurs and our procedure selects a model order of $\hat{k} > k_0$. To achieve consistency for $\hat{k}$, we follow Bai (1997) and make the size dependent on the sample size T so that $\alpha(T) \to 0$. Under these conditions, we can establish the results in Proposition 3.2 below.

Proposition 3.2. *Suppose that X_t is given in (3.1) and let $\hat{k}$, G^*, $\hat{f}(\lambda)$, $\hat{\gamma}_a$ and $\hat{d}_a$ be defined as in equations (3.6), (3.9), (3.10), (3.12) and (3.14) and denote by $g_{crit}(\alpha)$ the critical value of the G^* test at the significance level α. Then under Assumptions 3.1 to 3.5 we have:*

1. *For a fixed significance level α, $P(\hat{k} = k_0) = 1 - \alpha$.*
2. *For any non-increasing $\alpha(T)$ such that $\alpha(T) \to 0$, but $g_{crit}(\alpha(T)) = o(n^{2c_a} - \log(n))$ we have $P(\hat{k} = k_0) \to 1$, as $T \to \infty$.*

A proof for Proposition 3.2 can be found in the appendix. Of course $\alpha(T) = \alpha$ can be kept constant in empirical applications.

After the selection of appropriate long memory dynamics, the correct ARMA model orders p and q of a GARMA-model can be selected using an information criterion if the selected model is re-estimated parametrically using an approximate Whittle likelihood procedure. Giraitis and Leipus (1995) prove the consistency of this estimator. The semiparametric estimation results can be used as starting values for the numerical optimization.

As usual for semiparametric estimators there is the problem of an optimal bandwidth choice. We have to select ζ that determines the number of knots in the smoothing spline via $H_T = \lfloor 1 + T^{\zeta} \rfloor$, ξ that determines the usual bandwidth $m = \lfloor 1 + T^{\xi} \rfloor$ and j^* that is used for the estimation of the power law coefficient δ. In the case of only a single pole at the origin the MSE-optimal bandwidth m could be selected with the procedure of Henry and Robinson (1996) and Henry (2001). For multiple poles however, no such results are available. An additional complication in the k-factor Gegenbauer model is, that the parameter estimates might be negatively effected in small samples if the selected bandwidth is too large, so that Assumption 3.4 is not sufficient to inhibit that the effects from neighbouring poles interfere within the selected bandwidth. Hassler and Olivares (2013) show that a conservative deterministic bandwidth selection outperforms data-driven approaches in most situations.

In addition to that, simulation results not reported here show, that $\hat{k} > k_0$ mainly occurs in the presence of short memory dynamics if $\hat{f}(\lambda)$ is not flexible enough to remove peaks in the spectrum that originate from the short memory component. This typically leads to a selection of several cyclical frequencies $\hat{\gamma}_a$ within a very narrow frequency interval. We thus recommend to repeat the model selection procedure using a grid of different values for ξ and ζ and then to select the largest model with clearly distinct cyclical frequencies $\hat{\gamma}_a$. This is the strategy we will follow in our empirical application in Section 3.8. With regard to the choice of j^*, simulation

studies show, that $j^* = \sqrt{T}$ works well for the sample sizes considered in the Monte Carlo study, so that $\lambda_{j^*} = \frac{2\pi}{\sqrt{T}}$. Note that the distribution of the j, for which I_j is maximal if there is a pole at $\gamma = 0$ seems to be independent of the sample size T, but strongly dependent on the memory parameter d. Also, an overestimation of the d_a causes the appearance of zeros in the spectrum of the filtered process. Therefore, it will not cause an overestimation of the model order.
Situations with $\hat{k} < k_0$ can be identified by plotting the autocorrelation function of the filtered process $\Delta^{\hat{k}} X_t$. If the selected model is specified correctly, the hyperbolically decaying sinusoidal pattern should be removed.

3.6 Model Order Selection in Presence of Deterministic Seasonality

So far we have assumed that the cyclicality of the series is purely stochastic. In practice however, it is likely that there are deterministic as well as stochastic cyclical effects. In this case the mean $E[X_t]$ of the process in (3.1) is no longer constant. Instead, it is given by

$$\mu_t = \sum_{s=1}^{S} D_{st} \mu^{(s)}, \tag{3.15}$$

where S is the cycle length, the $\mu^{(s)}$ denote the cyclical means and D_{st} is an indicator variable that takes the value 1, for $t = s + S(q-1)$, where $q = 1, ..., r$ and $r = \lfloor T/S \rfloor$.
Similar to stochastic seasonality, deterministic cycles cause peaks in the (pseudo-) spectrum of the process. However, Arteche (2002) shows that these effects only appear at a single Fourier frequency at the cyclical frequency and each of its harmonics. Furthermore, since the generalized local Whittle estimator employs only periodogram ordinates to the left and to the right of these frequencies, it is robust to these effects. In a recent application of Gegenbauer models García-Enríquez et al. (2014), use this observation to estimate the stochastic component of the seasonality prior to the deterministic one. Here, we do not assume any prior knowledge about the order of the process and the period lengths of the cycles. We therefore employ the G^*-test to detect omitted cycles, which is likely to reject if there are peaks in the periodogram due to deterministic effects. This is why we proceed in the opposite order to García-Enríquez et al. (2014) and remove potential deterministic cycles by cyclical demeaning - prior to the application of the sequential-G^* procedure. In the context of seasonally integrated processes, the effect of seasonal demeaning has been studied by Abeysinghe (1991), Abeysinghe (1994) and Franses et al. (1995). Here, we treat stationary processes that are correctly specified and the demeaning can be conducted using simple OLS estimates of the cyclical means $\mu^{(s)}$. Afterwards, the model selection procedure is carried out on the residuals of this regression.
To see why this approach will not affect our model selection procedure asymptotically, consider the following argument of Murphy and Topel (2002) on step-wise estimation. In the first step we estimate the partial parameter vector θ_1 by minimization of some loss function $LF_1(X_1, ..., X_T|\theta_1)$ and in the second step we minimize the loss function $LF_2(X_1, ..., X_T|\theta_2, \hat{\theta}_1)$ with respect to the

partial parameter vector θ_2, conditional on $\hat{\theta}_1$. If $\hat{\theta}_1 \xrightarrow{p} \theta_1$, the estimation in the second step is asymptotically equivalent to the case where θ_1 is known. Therefore, Murphy and Topel (2002) show that - in contrast to the asymptotic distribution - the consistency of $\hat{\theta}_2$ is not affected by the step-wise estimation. Our procedure does not rely on the distributional properties of the estimators employed, but only on the distribution of the normalized periodogram ordinates of a weakly dependent process. Therefore, asymptotically our procedure remains unaffected, if it is applied to the residuals from a regression on seasonal dummies. The finite sample effects will be considered in the Monte Carlo study in Section 3.7.

Note that the same argument applies to the iterative nature of our procedure. In the first step the G^*-test is applied to the original series and d_1 as well as γ_1 are estimated from the original data. From the second step onwards, we proceed with the residual series $\Delta^{k^{(i)}}(\hat{\theta}_i)X_t$. Nevertheless, since every estimator is consistent given a consistent estimation of the parameters in the preceding steps, the succeeding estimator remains consistent, too and the G^*-test, that is based on the exponential distribution of the normalized periodogram ordinates, remains unaffected according to Slutsky's theorem.

3.7 Monte Carlo Study

In this section we conduct Monte Carlo experiments to evaluate the performance of our model order selection procedure. To separate the performance of the sequential filtering procedure from that of the modified G^*-test, we first use Walker's g-test to detect significant periodic behavior. We start by considering the case of a single cyclical frequency and investigate how the performance of the selection procedure depends on the location of the pole and the magnitude of the long memory parameter. Then, we analyze how the procedure performs if the number of poles is increased. Subsequently, we replace Walker's g-test with the modified G^*-test and analyze the robustness of the selection procedure to short memory dynamics. We also consider the effect of seasonal demeaning and we compare the performance of our G^*-test to the alternative approach suggested by Hidalgo and Soulier (2004).

All results shown are obtained with M=5000 Monte Carlo repetitions. Simulations from the Gegenbauer process are generated using its AR(∞)-representation with a truncation after 1000 lags and a burn-in period of 1000 observations. In all experiments the shorter series uses only the first $T_1 < T_2$ observations of the longer series. The DGP is always the k-factor Gegenbauer process from equation (3.1). Hereafter, we will use the term "power" to refer to the ability of the procedure to identify the true model order k_0.

3.7.1 Power Depending on the Location of the Pole and the Number of Cyclical Frequencies

First, we consider the 1-factor Gegenbauer process $(1-2\cos\gamma L+L^2)^d X_t = \varepsilon_t$ that does not contain any short memory dynamics. In this case the filtered process $\Delta^{k^{(i)}} X_t$ equals the white noise

sequence $\varepsilon_t \sim N(0,1)$ under the null hypothesis. Consequently, we can use Walker's original large sample statistic (3.8) instead of the modified G^*-statistic (3.9) to determine whether the filtered process contains significant periodicity. We allow the fractional exponent d that determines the shape of the pole to increase in steps of 0.05 from 0 to 0.45 and we shift the location of the pole in 14 steps from $\gamma = 0$ to $\gamma = \pi$. Because of the change of the power law parameter δ discussed above the results for $\gamma = 0$ and $\gamma = \pi$ are simulated with half of the d reported. The sample sizes considered are given by $T \in \{500, 1000, 2000, 5000\}$. Table 3.4 in the appendix gives a detailed overview of the results obtained in these simulations. A graphical summary of the results is depicted in Figure 3.1.

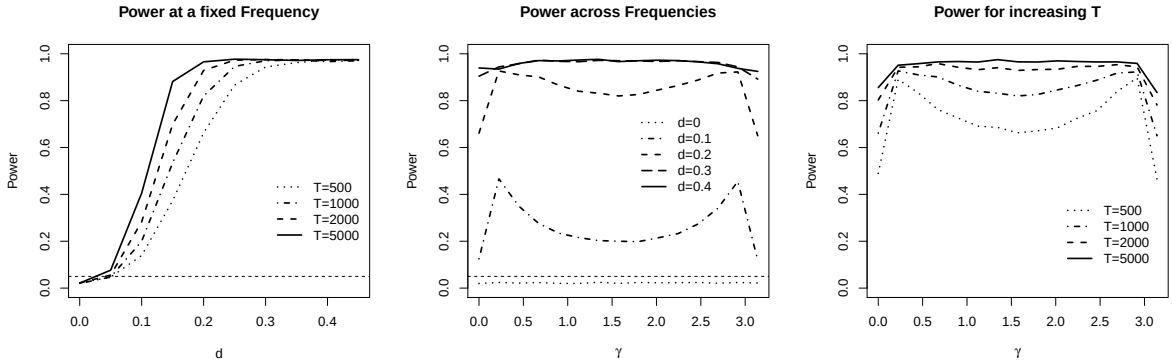

Figure 3.1: The DGP is $(1-2\cos\gamma L + L^2)^d X_t = \varepsilon_t$. On the left the cyclical frequency is fixed to $\gamma = \pi/2$ and T is increased. In the middle the sample size is fixed to $T = 1000$ and d is increased. In the graph on the right $d = 0.2$ is fixed and T is increased. We set $\xi = 0.7$.

Since we are interested in the effects of γ, T and d, we always keep one of the parameters fixed and show how the power depends on the other two. The graph on the left shows power curves if the pole is fixed to the frequency $\gamma = \pi/2$. It can be seen, that the power of the selection procedure is increasing in the magnitude of the fractional exponent d and the sample size T. For $d \leq 0.05$ however, it barely increases. This is because the location of the pole can only be identified consistently if the fractional exponent is bounded away from zero as stated in Assumption 3.2. Note that for the reasons discussed in Section 3.5, the maximal power that the selection procedure can achieve in practical applications is $1-\alpha$. It is clear to see that the procedure is slightly conservative in the size case where $d = 0$.

The graph in the middle shows the power for a fixed sample size of $T = 1000$ across the spectrum of possible periodic frequencies γ and for different values of d. One can see that the conservativeness of the procedure and the non-existing power for small d occur independent of the location of the periodic frequency. For smaller d the power is increasing with increasing distance of γ from $\pi/2$. Only for $\gamma \in \{0, \pi\}$ there is a drop in power again. This is caused by the fact that the fractional exponents at frequencies further away from the boundaries are estimated by using $2m$ Fourier frequencies - that is m frequencies on either side of the pole. At the boundaries however, we can only use $m_a < m$ Fourier frequencies on one side of the pole so that the estimator for the

fractional exponent has a higher variance at these frequencies.
In the graph on the right in Figure 3.1 we keep d fixed at 0.2 and the curves show the power across frequencies for increasing T, so that one can see that the increasing power in T remains intact across all frequencies and the power does not depend on the cyclical frequency γ asymptotically.

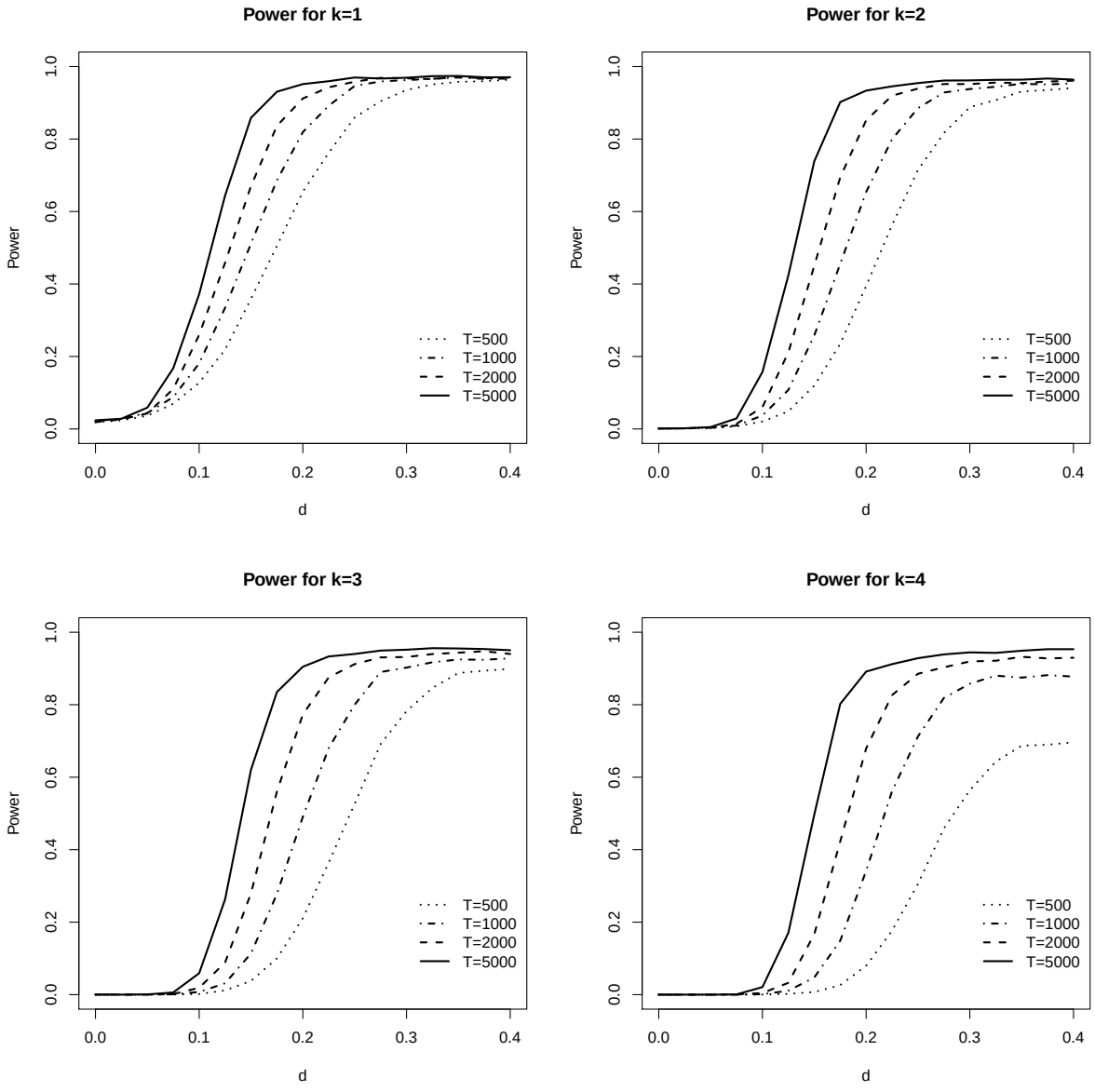

Figure 3.2: The DGP is $\prod_{a=1}^{k}(1-2\cos\gamma_a L+L^2)^{d_a}X_t=\varepsilon_t$ with $\gamma_a=(\pi a)/(k+1)$. Power curves are shown for $k\in\{1,2,3,4\}$ and increasing sample sizes for the case when $\xi=0.6$.

Since the purpose of our procedure is to select the true model order k_0 if there are multiple cyclical frequencies, we now allow the number of cyclical frequencies k to take values from 1 to 4. The DGP is given by $\prod_{a=1}^{k}(1-2\cos\gamma_a L+L^2)^{d_a}X_t=\varepsilon_t$, but the specification of the test, the sample sizes and the values of the fractional exponents considered are the same as before. We set $\gamma_a=(\pi a)/(k+1)$, so that the seasonal frequencies are equally spaced and the outer minimal and

maximal periodic frequencies have the distance $\pi/(k+1)$ to the boundaries. All long memory parameters are kept equal at $d_a = d$. The results of this experiment are shown in Figure 3.2 for $m = \lfloor T^{0.6} \rfloor$. Results for other bandwidths can be found in Table 3.5 in the appendix. From Figure 3.2, we can observe that the power is lower for higher k if d and T remain constant. Especially for small values of d repeated β-errors can lead to the selection of a lower model order. Nevertheless, it can easily be seen that the procedure remains consistent if k is increased. If one considers Table 3.5, one can observe that initially a larger bandwidth parameter ξ leads to a higher power, because the estimates of the d_a are more precise. For $\xi = 0.7$ however, the power starts to become non-monotonic in d. This is due to interference effects from neighboring poles. Nevertheless, asymptotically the procedure selects the right model order with probability $1-\alpha$. So the experiments show that the sequential-g procedure also works well for larger k. To avoid interference effects in empirical applications setting $\xi < \frac{\log((2\pi)/S)-\log(4\pi)}{\log(T)} + 1$ can be a useful rule of thumb, where S is the period of any suspected seasonality. This choice assures that the frequency band considered on each side of the poles has a width smaller than half of the distance between two neighboring harmonics.

3.7.2 The Influence of Short Memory Dynamics and Seasonal Demeaning

So far we have only considered the restricted version of the model selection procedure that is based on Walker's original large sample g-test given in (3.8). Under the null hypothesis, this test assumes that the process is *iid*. Since short memory dynamics are present in most applications, we now consider the case in which the innovations in (3.1) show stationary ARMA behavior and the modified G^*-test is used instead. The simulated DGP is $(1-\phi L)(1-2\cos\left(\frac{\pi}{2}\right)L+L^2)^d X_t = \varepsilon_t$. This means the short memory component of the process takes the form of an AR(1). We consider the case of one pole at $\gamma = \pi/2$ with increasing fractional exponents and increasing persistence of the short memory process. It is well known that the semiparametric estimators of the fractional exponents d_a are asymptotically unaffected by the presence of short memory dynamics. The critical part of our procedure is the performance of the g-test. Apart from the baseline case in which we use the g-test, the number of segments in the spline used for the G^*-test is determined according to $H_T = \lfloor T^{1/4} \rfloor$.

Figure 3.3 compares the performance of our sequential procedure using the g-test with the performance if the modified G^*-test from (3.9) is used. A complete overview of the results is given in Table 3.4 in the appendix, where we also consider the seasonal autoregressive model $(1-\phi L^4)(1-2\cos\left(\frac{\pi}{2}\right)L+L^2)^d X_t = \varepsilon_t$. In quarterly data a period of length 4 corresponds to an annual cycle. The spectral density of this seasonal AR-component has a peak at $\pi/2$, which coincides with the peak of the seasonal long memory process.

If we focus on the results for the g-test in a sample of $T = 500$ observations on the left hand side of Figure 3.3, we observe that the test maintains good size and power for small values of ϕ. For larger values of ϕ however, if $d = 0$ the size increases with ϕ and the procedure loses its ability to identify the correct model order k_0. As we noted in Section 3.5, the procedure tends to find

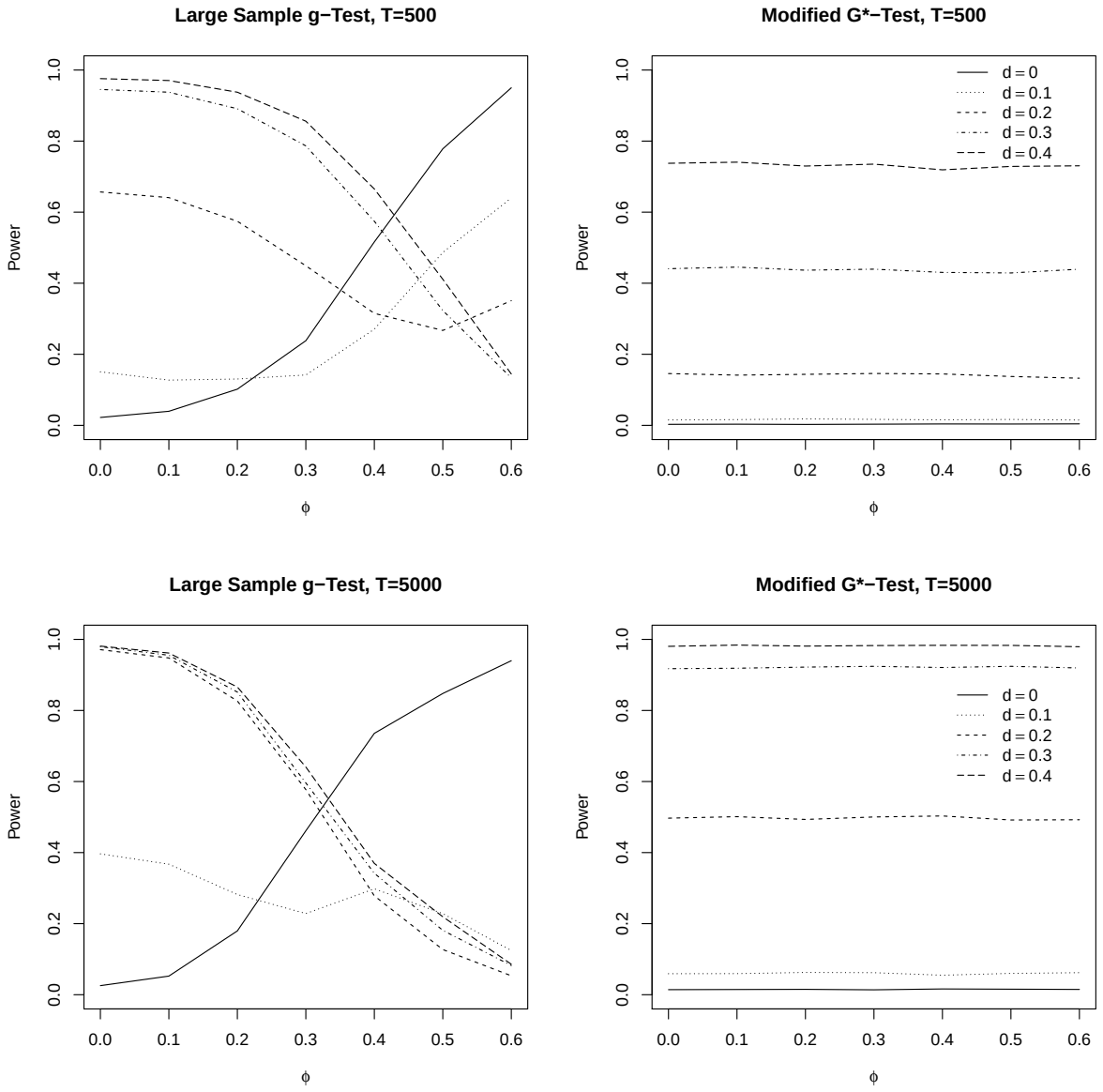

Figure 3.3: Power curves of the sequential G^*-procedure compared to the procedure using Walker's large sample g-test, for $(1-\phi L)(1-2\cos\left(\frac{\pi}{2}\right)L+L^2)^d X_t = \varepsilon_t$ with increasing autoregressive parameters ϕ. The number of knots in the spline is determined as $H_T = \lfloor T^{1/4} \rfloor$.

spurious seasonal peaks if short memory dynamics are present. If the G^*-test is used on the other hand, we see that the procedure maintains good size and power properties for all sample sizes. Only for larger values of ϕ these deteriorate slightly, but this affect can be mitigated easily if the bandwidth parameter ζ is increased further so that the logspline estimate is more flexible. As can be seen for the case when $\phi = 0$, this additional flexibility comes at the cost of a power loss, so that in empirical applications the bandwidth has to be chosen carefully. Overall however, we find that the sequential G^*-procedure performs well - especially if the short memory dynamics are only moderately persistent.

		HS-Test							G-Test						
T	m/ϕ	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**
500	**0.60**	0.11	0.19	0.41	0.66	0.85	0.94	0.98	0.02	0.02	0.02	0.02	0.03	0.05	0.10
	0.65	0.09	0.23	0.56	0.80	0.92	0.97	0.99	0.02	0.02	0.02	0.03	0.03	0.04	0.09
	0.70	0.07	0.23	0.58	0.81	0.92	0.97	0.99	0.02	0.03	0.02	0.03	0.03	0.05	0.09
	0.75	0.06	0.15	0.41	0.65	0.79	0.89	0.95	0.03	0.02	0.03	0.03	0.03	0.05	0.09
	0.80	0.07	0.08	0.14	0.22	0.30	0.37	0.43	0.02	0.03	0.02	0.03	0.03	0.05	0.09
5000	**0.60**	0.06	0.09	0.17	0.31	0.56	0.80	0.92	0.05	0.05	0.05	0.06	0.05	0.06	0.08
	0.65	0.05	0.12	0.38	0.69	0.87	0.96	1.00	0.05	0.05	0.05	0.05	0.05	0.05	0.08
	0.70	0.04	0.29	0.71	0.92	0.98	1.00	1.00	0.04	0.05	0.05	0.05	0.05	0.06	0.09
	0.75	0.05	0.52	0.86	0.97	1.00	1.00	1.00	0.05	0.04	0.04	0.05	0.05	0.06	0.10
	0.80	0.05	0.51	0.80	0.94	0.99	1.00	1.00	0.05	0.05	0.05	0.05	0.05	0.06	0.09

Table 3.1: Size comparison of the G^*-Test and the HS-Test for cyclical long memory in seasonal AR(1) process $(1-\phi L^4)X_t = \varepsilon_t$. The number of knots in the spline is determined as $H_T = \lfloor T^{1/4} \rfloor$.

From Table 3.4 it can easily be seen that these results carry over to the case of seasonal short memory dynamics since the modified G^*-test is designed to detect cyclical long memory irrespective of the short memory dynamics. Also, even though we only present results for AR(1) dynamics here, the results carry over to more general ARMA processes.

As discussed in Section 3.1, Hidalgo and Soulier (2004) outline another iterative procedure to determine the model order in (3.1). It consists of the following step: i.) Find the largest periodogram ordinate and test for $d_a = 0$ at this Fourier frequency. ii.) If the null hypothesis is rejected, add the respective Gegenbauer filter to the model - otherwise terminate the procedure. iii.) Exclude the neighbourhood of the last significant pole from the periodogram and repeat the procedure from i.) onwards.

Both, the HS-procedure and ours are sequential testing procedures. The main difference between them is that Hidalgo and Soulier (2004) test for $d_a = 0$ whereas we employ a non-parametric test for the significance of peaks in the periodogram. While their approach has the advantage that it does not require a filtering step, it has the disadvantage that its size will suffer from the slow convergence of semiparametric estimates for d_a to their asymptotic Normal distributions (cf. Hassler and Olivares (2013)). Therefore, it can be expected to overestimate the true model order. In addition to that, our approach is designed to account for short memory dynamics, that will also affect tests for $d_a = 0$ in finite samples. We therefore compare the performance of the G^*-test with that of a test for $d_a = 0$ on the basis of $\hat{d}_a$ estimated through (3.14). Table 3.1 displays the results of this procedure for increasing sample sizes and seasonal autoregressive effects.

As expected, the HS-test is oversized for $T = 500$, even if no autoregressive effects are present

	γ					$\pi/2$									$\pi/4$				
T	μ_4/**d**	**0**	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**	**0**	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**
	0*	0.02	0.04	0.14	0.37	0.66	0.86	0.95	0.97	0.98	0.02	0.04	0.17	0.45	0.74	0.90	0.95	0.97	0.97
500	0	0.02	0.03	0.12	0.31	0.57	0.80	0.93	0.97	0.98	0.02	0.04	0.17	0.45	0.75	0.90	0.95	0.97	0.97
	4	0.02	0.04	0.17	0.46	0.78	0.93	0.97	0.98	0.98	0.02	0.05	0.18	0.43	0.74	0.90	0.95	0.98	0.97
	0*	0.02	0.04	0.21	0.52	0.83	0.94	0.97	0.98	0.98	0.02	0.05	0.25	0.62	0.88	0.96	0.97	0.97	0.97
1000	0	0.02	0.04	0.18	0.46	0.78	0.93	0.97	0.98	0.98	0.02	0.05	0.25	0.62	0.88	0.96	0.97	0.97	0.97
	4	0.02	0.04	0.17	0.46	0.78	0.93	0.97	0.98	0.98	0.02	0.05	0.26	0.62	0.88	0.96	0.97	0.97	0.97
	0*	0.02	0.05	0.27	0.69	0.93	0.97	0.97	0.98	0.98	0.02	0.06	0.34	0.79	0.95	0.97	0.97	0.97	0.97
2000	0	0.02	0.05	0.24	0.64	0.91	0.97	0.97	0.98	0.98	0.02	0.06	0.35	0.79	0.95	0.97	0.97	0.98	0.97
	4	0.02	0.05	0.24	0.65	0.92	0.97	0.98	0.98	0.98	0.02	0.06	0.33	0.78	0.95	0.97	0.98	0.97	0.98
	0*	0.03	0.07	0.41	0.87	0.97	0.98	0.98	0.98	0.98	0.02	0.08	0.50	0.91	0.97	0.97	0.98	0.97	0.97
5000	0	0.03	0.07	0.39	0.85	0.97	0.98	0.98	0.98	0.98	0.02	0.08	0.50	0.91	0.97	0.97	0.98	0.97	0.97
	2	0.03	0.07	0.38	0.85	0.97	0.98	0.98	0.98	0.98	0.03	0.09	0.49	0.92	0.97	0.97	0.97	0.98	0.98

Table 3.2: Effect of seasonal demeaning: The DGP is $(1-2\cos\gamma L+L^2)^d(X_t-\mu_t)=\varepsilon_t$, with μ_t following (3.15), where $\mu^{(1)}=\mu^{(2)}=\mu^{(3)}=0$. For $\mu^{(4)}=0^*$ the seasonal mean is zero and no demeaning is conducted.

and if ϕ increases, there is no size control - not even for $T=5000$. The G^*-test on the other hand maintains a satisfactory size for all values of T and ϕ. Only for $\phi=0.6$ it becomes slightly liberal. However, as discussed above, this effect can easily be mitigated if ζ is increased.

In Section 3.6, we discussed that it might be necessary in empirical applications to conduct a seasonal demeaning prior to the application of our model selection procedure if one suspects that deterministic seasonal effects are present. We therefore consider the effect of this approach in Table 3.2. Again, we consider a 1-factor Gegenbauer process $(1-2\cos\gamma L+L^2)^d(X_t-\mu_t)=\varepsilon_t$ but now combined with a deterministic cycle with period 4. For the sake of simplicity we set $\mu^{(1)}=\mu^{(2)}=\mu^{(3)}=0$ in (3.15) and then consider different values of $\mu^{(4)}$. If $\gamma=\pi/2$ the stochastic and deterministic cycles coincide. If $\gamma=\pi/4$ on the other hand, the stochastic seasonality would have a period of 8, so that the peaks in the periodogram are at different frequencies. The baseline case when there are no deterministic seasonalities and no demeaning is conducted is denoted by $\mu^{(4)}=0^*$. As one can see, the procedure has good size properties, irrespective of the demeaning. With regard to the power, we observe that there is a power loss in small samples if the periods of the deterministic and the stochastic seasonality coincide. With increasing sample size however, the power loss vanishes. A demeaning prior to the application of the sequential-G^* procedure therefore does not affect the size and reduces the power only moderately in finite samples. In addition to that, it can also be seen that the demeaning has no effect on the power if the periods of the cycles do not coincide. For $\gamma=\pi/4$, we observe that the power is unaffected by a seasonal demeaning at frequency $\gamma=\pi/2$. Additionally, as in Figure 3.1, we see that the power is higher at $\gamma=\pi/4$ than for $\gamma=\pi/2$.

With respect to conditional heteroscedasticity Robinson and Henry (1999) show that the consistency, the asymptotic normality and even the asymptotic variance of the Gaussian semiparamet-

ric estimator for the fractional exponents remain unaffected by conventional ARCH/GARCH type conditional heteroscedasticity and also by long memory conditional heteroscedasticity. Additional simulation studies showed that these results carry over to our model selection procedure, even though this also depends on the semiparametric estimator of Hidalgo and Soulier (2004) and the modified G^*-test.

3.8 Modeling Californian Electricity Loads with k-factor GARMA Models

As discussed in the introduction, GARMA models are useful in high frequency datasets that potentially exhibit multiple seasonalities. One such example are electricity load series. In particular the forecasting of the latter has attracted continued attention in the literature, because electricity demand is very volatile and it has to be matched by supply in real time. Since different means of electricity generation have very different marginal costs, precise forecasts are of major importance for electricity producers to schedule production capacities accordingly. Recently, GARMA models have been used to generate such forecasts by Soares and Souza (2006) who apply a 1-factor GARMA model to forecast electricity demand in the area of Rio de Janeiro (Brazil) and Diongue et al. (2009) who use a 3-factor GIGARCH model to forecast German spot market electricity prices.

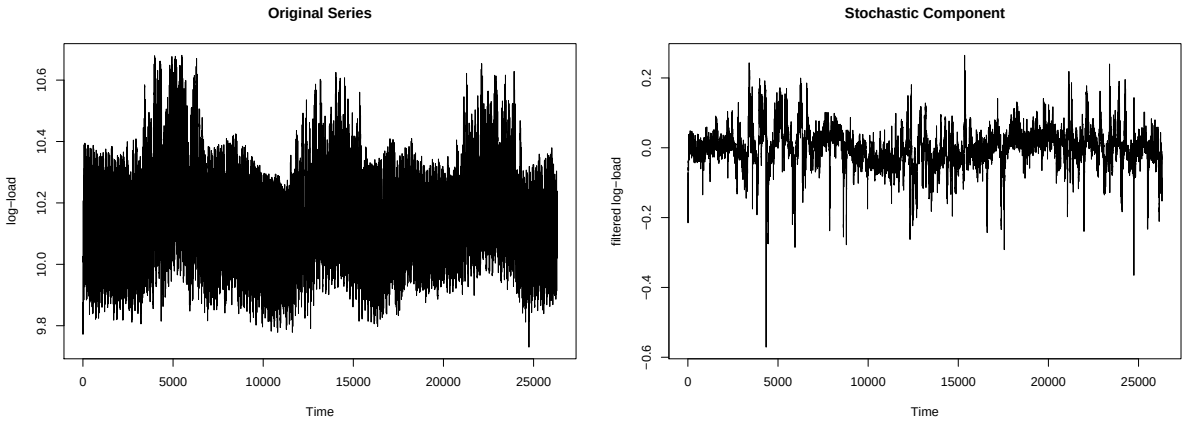

Figure 3.4: Californian system wide log-load series from 2000 to 2002 before and after deterministic seasonality is removed.

In general, two different approaches can be distinguished in the literature on electricity load forecasting. The global approach is to fit a relatively complex model to the hourly series, whereas local approaches forecast all 24 hourly series with separately estimated simpler models. By applying our model selection procedure to electricity load series from the Californian power system, we complement this literature with an in-sample perspective on model selection in the

global series. The data is downloaded from UCEI[2] and covers the period from 2000 to 2002, which gives us 26304 hourly observations. A similar dataset was considered in Weron and Misiorek (2008) who focus on prices in the period 1999-2000. As discussed in Section 3.6, we remove deterministic seasonality upfront by considering the residual series obtained from regressing the log of each hourly series separately on a trend and dummy variables for the day of the week and the month of the year. Insignificant dummies are discarded stepwise using a general-to-specific procedure. A similar approach was suggested in Haldrup and Nielsen (2006). This simple deterministic model already achieves an R^2 of approximately 90.49 percent, which implies that a large proportion of the seasonality in the series is deterministic. The original series as well as the residual series from this regression are plotted in Figure 3.4. We will hereafter refer to the residuals from this regression as the filtered electricity load series. It is obvious, that the variance of the series is strongly reduced. The left side of Figure 3.5 depicts the autocorrelation function of the filtered electricity load series. It shows clearly that the series exhibits long memory and periodicity.

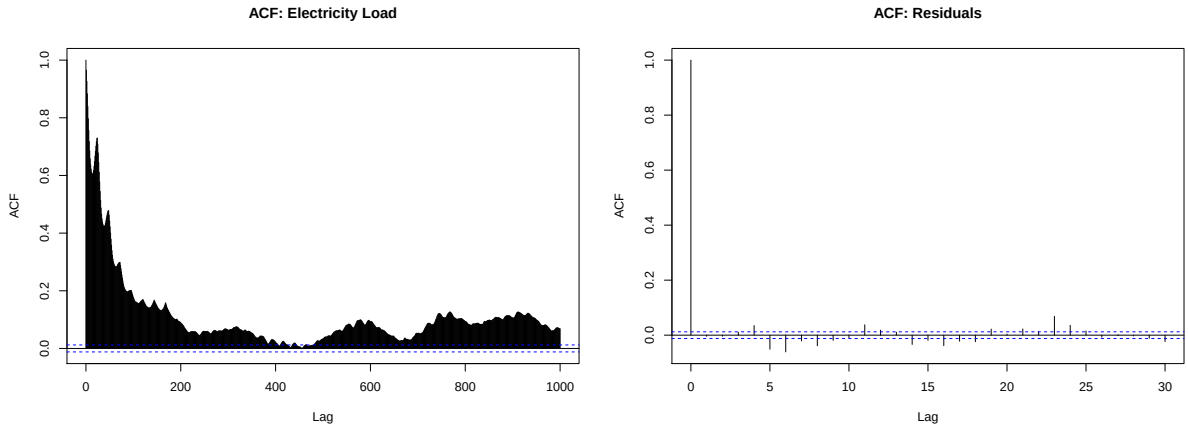

Figure 3.5: Autocorrelation function of the original series (left) and the residual series of 14-factor Gegenbauer process (right).

If one would discretionary fit a model based on a visual inspection of the periodogram shown on the left side of Figure 3.6, one would probably consider a two factor model with a pole at the origin and one in the neighborhood of $\lambda = 0.25$ which approximately corresponds to the daily frequency $2\pi/24$. Such a process could be approximated well by an ARFIMA process, since the effect at $\lambda = 0.25$ seems to be of a small magnitude. In sharp contrast to that, our model selection procedure indicates that a 14-factor Gegenbauer process should be used to model the series. The intuition for this unexpected result becomes obvious, if one considers the residual series $\Delta^{k^{(i)}} X_t$ from (3.4) after filtering out the effect of the pole at the origin. The periodogram of the filtered process is depicted on the right hand side of Figure 3.6. The poles that become visible here are not visible in the original periodogram on the left hand side, because of the

[2] www.ucei.berkeley.edu

magnitude of the Fourier frequency closest to the origin. To include these effects, it is necessary to increase the model order.

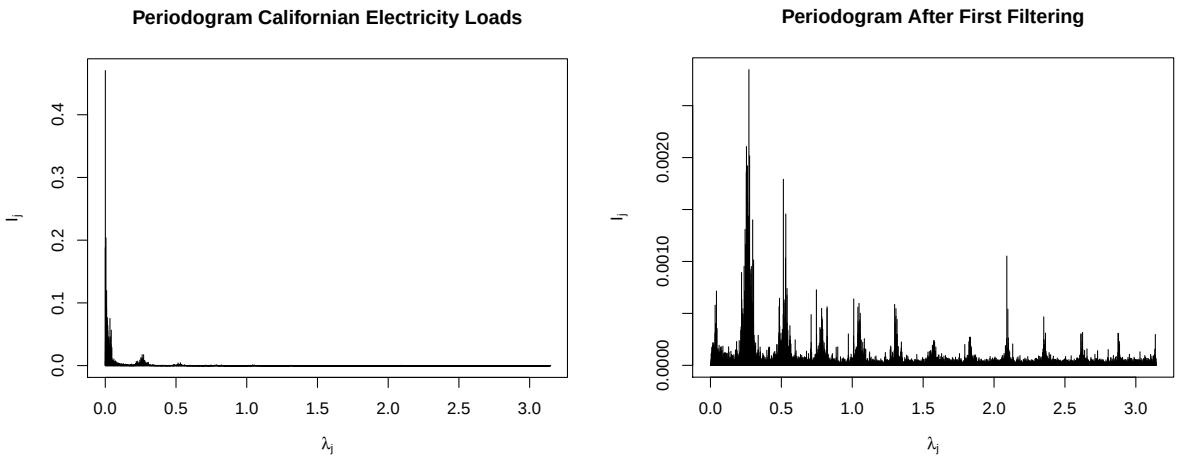

Figure 3.6: Periodogram the Californian load series (left) and the residual series $\Delta^{\hat{k}^{(i)}} X_t$ obtained after removing the non-cyclical long memory effects (right).

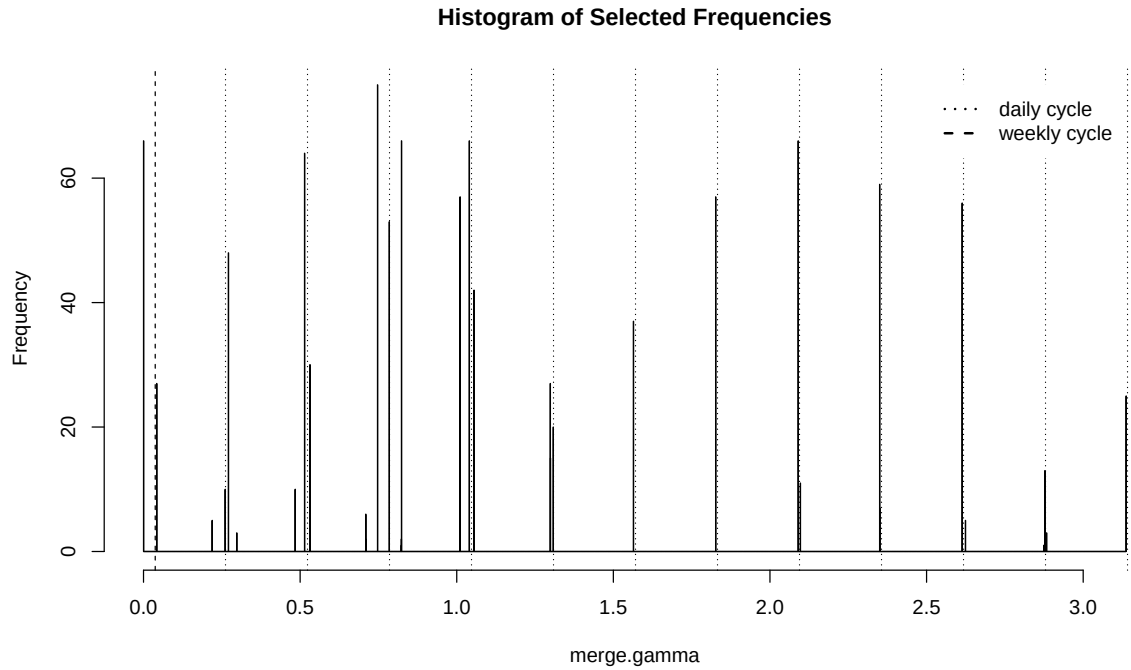

Figure 3.7: Histogram of the frequencies selected by the sequential-G^* procedure applied on a grid of bandwidth parameters.

As discussed in Section 3.5, the bandwidth choice has a critical influence on the results of the selection procedure. This is why we repeat the analysis using a grid of values for ξ and ζ, that

a	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\hat{d}_a$	0.3373	0.2401	0.5550	0.3432	0.2208	0.3237	0.2791	0.2111	0.1902	0.2042	0.1979	0.1803	0.1241	0.1155
$\hat{\gamma}_a$	0.0025	0.0425	0.2625	0.5125	0.7825	1.0425	1.3075	1.5625	1.8275	2.0925	2.3525	2.6125	2.8775	3.1375

Table 3.3: Fitted 14-factor Gegenbauer model. Standard errors for d_a are 0.0233 for γ_1 and γ_{14}, 0.0200 for γ_2, and 0.0167 for all other periodic frequencies.

determine the number of frequencies m used in the estimation of the long memory parameters d_a and the number of segments H_T used in the logspline estimate of the spectral density. We allow ξ to increase from 0.45 to 0.65 in steps of 0.01 and determine ζ so that the number of segments varies from 7 to 12. For every parameter constellation we store all estimated cyclical frequencies $\hat{\gamma}_a$.

The histogram in Figure 3.7 shows which frequencies are selected and how often they are found to be significant. It can be seen, that not every of the cyclical frequencies is selected with every bandwidth parameter. However, 14 frequencies are selected in the majority of parameter constellations. These are the zero frequency that corresponds to a non-seasonal long memory component, $2\pi/(24\times 7)$ - the frequency of a weekly cycle and $2\pi\upsilon/24$, with $\upsilon = 1,...,12$ which is the frequency of a daily cycle and its harmonics. For better comparison, these theoretical frequencies are superimposed as dashed and dotted vertical lines in Figure 3.7.

These findings indicate, that the data is best explained by the 14-factor Gegenbauer model with these respective frequencies, which confirms the initial expectation that there are multiple cycles in the data generating process. To fit the 14-factor Gegenbauer model we estimate the relevant cyclical frequencies by the local modes in the histogram. These estimates $\hat{\gamma}_a$ along with the estimated fractional exponents $\hat{d}_a$ are given in Table 3.3. If one considers the $\hat{d}_a$, one observes that the exponent at the origin is larger than 0.25 which implies non-stationary but mean reverting long memory. The memory parameter of the daily frequency is 0.5550 which is also in the non-stationary but mean reverting region.[3] The other exponents lie between 0.12 and 0.35 and tend to decrease for higher frequencies.

The R^2 obtained using the 14-factor model is as high as 81.44 percent, which means that the deterministic model and the Gegenbauer model together explain approximately 98.23 percent of the variation of the original series. Figure 3.5 shows the autocorrelation function of the residuals after the short memory dynamics are accounted for by an ARMA(2,2) model, as discussed in Section 3.5. As one can see, the dependence is completely removed.

Based on the visual inspection of the periodogram we would have chosen a 2-factor Gegenbauer model with non-seasonal long memory and a daily cycle. In contrast to that, our model order selection procedure finds a 14-factor model that also includes cycles at the weekly frequency and the harmonics of the daily frequency. These different model choices have very different implications. In the case of the 2-factor model no weekly cycle is detected and one would

[3]The non-stationary part of the parameter space for the d_a was excluded from our analysis in the preceding sections. However, we discussed in Section 3.4 that the results can be expected to carry over to the mean reverting region where $d_a \in (0.5, 1)$.

conclude that the weekly seasonality in the original series is best modeled by deterministic dummies. As discussed in the introduction, the occurrence of significant cycles at the harmonics of the daily frequency indicates that the daily cycle is not of a sinusoidal form.
A test for the equality of the fractional exponents at the daily frequency and its harmonics using a simple χ^2_{11}-statistic gives a test statistic of 535.75, which is strongly rejected at every common significance level. A more parsimonious rigid seasonal fractionally integrated model would therefore be misspecified.
Furthermore, the exceptional fit of the 14-factor model for the global series shows that this series can indeed be modeled directly, if the daily load profile is accounted for by a suitable seasonal model. Considering the local series for every hour of the day separately can be done with a relatively small loss of information, if the hourly series has low persistence. For a series with non-stationary long memory, as found in Table 3.3, on the other hand, the information loss is considerable. It is therefore desirable to find an appropriate global model.
Since the selected frequencies are seasonal frequencies and their harmonics, one might be inclined to conclude that the model order could be inferred from theoretical considerations and tested using the methods of Hassler et al. (2009) or Robinson (1994). Note however, that there are 12 daily harmonics and 84 weekly harmonics. Only four of the weekly harmonics are equal to a daily harmonic, so that even with a grid of just 10 values for each d_a there would be 10^{93} possible combinations in the grid, which makes this approach computationally infeasible.

3.9 Conclusion

We introduce an automatic model order selection procedure for k-factor Gegenbauer models that is based on iterative filtering and can be used for general model selection in cyclical long-memory processes. As a byproduct we suggest a modified test for persistent periodicity in stationary short memory models. Our procedure allows an easier application of k-factor Gegenbauer models in empirical analyses and prevents the use of false model specifications that are based on discretionary decisions after a visual analysis of the periodogram. As the example of the Californian electricity load series in Section 3.8 shows, this visual approach can be misleading if it is used as a tool for the selection of the model order without taking the effect of the non-uniform power law into account.
We also gain new insights into the behavior of electricity load series. It turns out that the stochastic variation of the Californian electricity loads can be modeled very well by a 14-factor GARMA process. The fit achieved suggests that it can indeed be a good strategy to model the global series directly instead of fitting separate models for every hour of the day. This is especially important for short term forecasts since the 1-hour-ahead forecast of a local model only uses data up to 23 hours ago and does not utilize the information contained in the most recent observations. These insights demonstrate the potential of k-factor Gegenbauer models for the modeling of periodic time series.

Appendix

T	d/γ	0	.22	.45	.67	.9	1.12	1.35	1.57	1.8	2.02	2.24	2.47	2.69	2.92	π
500	**0**	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	0.05	0.03	0.07	0.05	0.05	0.04	0.05	0.03	0.05	0.04	0.04	0.04	0.05	0.05	0.06	0.03
	0.10	0.08	0.32	0.24	0.19	0.16	0.16	0.15	0.14	0.14	0.15	0.16	0.19	0.25	0.33	0.09
	0.15	0.24	0.72	0.57	0.49	0.42	0.38	0.38	0.37	0.39	0.40	0.43	0.49	0.58	0.72	0.24
	0.20	0.49	0.89	0.83	0.76	0.73	0.69	0.69	0.66	0.67	0.68	0.73	0.76	0.84	0.90	0.46
	0.25	0.71	0.93	0.92	0.92	0.89	0.87	0.87	0.87	0.87	0.88	0.89	0.90	0.93	0.93	0.69
	0.30	0.84	0.93	0.95	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.92	0.82
	0.35	0.90	0.92	0.95	0.96	0.97	0.97	0.97	0.96	0.97	0.97	0.96	0.96	0.95	0.92	0.88
	0.40	0.93	0.92	0.95	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.95	0.91	0.90
	0.45	0.93	0.90	0.94	0.95	0.97	0.97	0.97	0.98	0.97	0.97	0.96	0.95	0.94	0.91	0.91
1000	**0**	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	0.05	0.03	0.08	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.09	0.04
	0.10	0.12	0.47	0.35	0.28	0.23	0.22	0.20	0.20	0.20	0.21	0.23	0.27	0.34	0.46	0.12
	0.15	0.37	0.86	0.74	0.65	0.59	0.55	0.53	0.53	0.52	0.56	0.59	0.65	0.74	0.85	0.36
	0.20	0.66	0.93	0.91	0.90	0.87	0.84	0.83	0.82	0.83	0.85	0.86	0.89	0.92	0.92	0.65
	0.25	0.85	0.94	0.95	0.96	0.95	0.95	0.96	0.95	0.95	0.96	0.96	0.96	0.95	0.94	0.83
	0.30	0.90	0.94	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.94	0.89
	0.35	0.92	0.94	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.96	0.96	0.94	0.91
	0.40	0.94	0.93	0.96	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.96	0.94	0.92
	0.45	0.93	0.93	0.96	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.96	0.95	0.94	0.93
2000	**0**	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.03	0.02
	0.05	0.04	0.11	0.08	0.07	0.06	0.06	0.05	0.06	0.06	0.06	0.06	0.07	0.09	0.11	0.03
	0.10	0.17	0.60	0.46	0.36	0.32	0.29	0.29	0.28	0.28	0.29	0.33	0.37	0.46	0.60	0.18
	0.15	0.52	0.91	0.85	0.80	0.75	0.71	0.71	0.70	0.70	0.72	0.76	0.79	0.87	0.92	0.51
	0.20	0.80	0.94	0.94	0.96	0.94	0.93	0.94	0.93	0.93	0.93	0.95	0.95	0.95	0.94	0.78
	0.25	0.88	0.95	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.95	0.87
	0.30	0.90	0.95	0.97	0.97	0.97	0.97	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.95	0.91
	0.35	0.92	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.96	0.92
	0.40	0.93	0.95	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.92
	0.45	0.93	0.95	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.95	0.93
5000	**0**	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	0.05	0.04	0.16	0.11	0.09	0.08	0.07	0.06	0.08	0.07	0.08	0.08	0.09	0.12	0.17	0.05
	0.10	0.26	0.76	0.63	0.52	0.46	0.42	0.41	0.41	0.41	0.43	0.47	0.53	0.64	0.77	0.27
	0.15	0.71	0.92	0.92	0.93	0.90	0.87	0.89	0.88	0.88	0.89	0.91	0.91	0.93	0.93	0.69
	0.20	0.86	0.95	0.96	0.97	0.97	0.96	0.97	0.97	0.96	0.97	0.97	0.97	0.97	0.96	0.83
	0.25	0.88	0.96	0.96	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.96	0.96	0.88
	0.30	0.91	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.90
	0.35	0.92	0.96	0.97	0.98	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.97	0.97	0.97	0.91
	0.40	0.92	0.97	0.98	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.96	0.91
	0.45	0.93	0.96	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.98	0.96	0.92

Table 3.4: Size and power of the sequential G^*-procedure for $(1-2\cos\gamma L+L^2)^d X_t = \varepsilon_t$ at different frequencies γ.

	ξ	**0.4**				**0.5**				**0.6**				**0.7**			
T	d/k	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
	0	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	0.1	0.09	0.02	0.00	0.00	0.11	0.02	0.00	0.00	0.12	0.02	0.00	0.00	0.13	0.02	0.00	0.00
500	**0.2**	0.49	0.27	0.14	0.07	0.58	0.35	0.18	0.10	0.64	0.41	0.22	0.08	0.66	0.36	0.07	0.04
	0.3	0.77	0.60	0.48	0.35	0.87	0.79	0.67	0.56	0.93	0.88	0.76	0.52	0.95	0.82	0.16	0.16
	0.4	0.84	0.69	0.58	0.46	0.92	0.86	0.81	0.70	0.97	0.95	0.88	0.63	0.98	0.87	0.09	0.11
	0	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	0.1	0.13	0.03	0.01	0.00	0.15	0.03	0.01	0.00	0.18	0.04	0.01	0.00	0.20	0.04	0.01	0.00
1000	**0.2**	0.62	0.43	0.28	0.16	0.74	0.57	0.41	0.29	0.81	0.67	0.50	0.34	0.82	0.66	0.24	0.08
	0.3	0.81	0.67	0.55	0.43	0.91	0.86	0.78	0.74	0.96	0.94	0.90	0.84	0.97	0.95	0.24	0.13
	0.4	0.84	0.71	0.58	0.52	0.93	0.88	0.83	0.77	0.96	0.95	0.92	0.85	0.98	0.96	0.03	0.12
	0	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	0.1	0.18	0.05	0.02	0.01	0.22	0.06	0.02	0.01	0.25	0.07	0.02	0.01	0.27	0.07	0.02	0.00
2000	**0.2**	0.71	0.56	0.41	0.31	0.85	0.76	0.65	0.55	0.91	0.85	0.77	0.68	0.94	0.88	0.76	0.17
	0.3	0.83	0.70	0.60	0.50	0.93	0.88	0.84	0.79	0.96	0.95	0.94	0.91	0.98	0.97	0.88	0.04
	0.4	0.87	0.76	0.65	0.56	0.95	0.90	0.86	0.83	0.97	0.96	0.95	0.93	0.98	0.97	0.80	0.07
	0	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	0.1	0.28	0.10	0.05	0.02	0.33	0.13	0.05	0.02	0.37	0.15	0.05	0.02	0.40	0.17	0.06	0.02
5000	**0.2**	0.78	0.65	0.53	0.43	0.90	0.85	0.78	0.73	0.95	0.93	0.91	0.88	0.97	0.97	0.95	0.90
	0.3	0.87	0.76	0.67	0.58	0.95	0.91	0.88	0.84	0.96	0.96	0.95	0.94	0.98	0.97	0.96	0.85
	0.4	0.91	0.81	0.72	0.64	0.95	0.92	0.89	0.86	0.96	0.97	0.96	0.96	0.98	0.98	0.94	0.72

Table 3.5: Size and power of the sequential G^* procedure for increasing model order k and different bandwidths $m = \lfloor T^\xi \rfloor$. The DGP is $\prod_{a=1}^{k}(1-2\cos\gamma_a L + L^2)^{d_a} X_t = \varepsilon_t$ with $\gamma_a = (\pi a)/(k+1)$.

		non-seasonal														seasonal													
		baseline							$\zeta = 0.25$							baseline							$\zeta = 0.25$						
T	**d/ϕ**	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**
	0	0.02	0.04	0.10	0.24	0.52	0.78	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.16	0.35	0.62	0.86	0.98	0.00	0.01	0.01	0.01	0.01	0.03	0.07
	0.1	0.15	0.13	0.13	0.14	0.27	0.49	0.64	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.14	0.28	0.41	0.36	0.15	0.03	0.00	0.02	0.02	0.03	0.04	0.06	0.11	0.22
500	**0.2**	0.66	0.64	0.57	0.45	0.32	0.27	0.35	0.15	0.14	0.14	0.15	0.14	0.14	0.13	0.67	0.79	0.74	0.47	0.16	0.03	0.00	0.13	0.14	0.17	0.21	0.27	0.34	0.48
	0.3	0.95	0.94	0.89	0.79	0.57	0.32	0.13	0.44	0.45	0.44	0.44	0.43	0.43	0.44	0.95	0.94	0.80	0.48	0.18	0.03	0.01	0.45	0.44	0.49	0.51	0.56	0.65	0.74
	0.4	0.98	0.97	0.94	0.86	0.67	0.41	0.14	0.74	0.74	0.73	0.74	0.72	0.73	0.73	0.98	0.95	0.82	0.50	0.20	0.04	0.01	0.73	0.75	0.77	0.77	0.82	0.86	0.90
	0	0.02	0.04	0.12	0.31	0.60	0.85	0.96	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.07	0.21	0.44	0.75	0.95	1.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02
	0.1	0.21	0.18	0.16	0.16	0.31	0.47	0.50	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.20	0.37	0.47	0.30	0.05	0.00	0.00	0.02	0.02	0.02	0.02	0.03	0.04	0.05
1000	**0.2**	0.82	0.80	0.71	0.55	0.32	0.21	0.24	0.14	0.13	0.14	0.14	0.13	0.14	0.13	0.83	0.88	0.72	0.32	0.05	0.00	0.00	0.13	0.13	0.13	0.15	0.16	0.17	0.23
	0.3	0.97	0.96	0.90	0.76	0.50	0.23	0.08	0.42	0.44	0.42	0.42	0.43	0.43	0.44	0.97	0.94	0.72	0.31	0.05	0.00	0.00	0.44	0.42	0.44	0.44	0.45	0.49	0.53
	0.4	0.98	0.96	0.91	0.79	0.55	0.28	0.10	0.74	0.74	0.73	0.74	0.72	0.73	0.74	0.98	0.94	0.73	0.33	0.05	0.00	0.00	0.73	0.73	0.73	0.74	0.75	0.76	0.80
	0	0.02	0.05	0.14	0.37	0.67	0.87	0.95	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.07	0.23	0.55	0.87	0.99	1.00	0.01	0.01	0.01	0.02	0.02	0.02	0.05
	0.1	0.27	0.26	0.21	0.20	0.33	0.39	0.30	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.26	0.49	0.54	0.23	0.02	0.00	0.00	0.05	0.05	0.05	0.06	0.08	0.13	0.23
2000	**0.2**	0.93	0.90	0.81	0.61	0.32	0.16	0.12	0.32	0.32	0.33	0.34	0.33	0.32	0.34	0.93	0.92	0.69	0.23	0.01	0.00	0.00	0.33	0.34	0.34	0.38	0.44	0.51	0.65
	0.3	0.98	0.96	0.89	0.70	0.41	0.19	0.06	0.78	0.76	0.77	0.77	0.77	0.77	0.78	0.98	0.94	0.67	0.22	0.01	0.00	0.00	0.78	0.76	0.77	0.80	0.82	0.85	0.90
	0.4	0.98	0.97	0.89	0.72	0.46	0.24	0.09	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.98	0.93	0.69	0.22	0.01	0.00	0.00	0.96	0.96	0.95	0.95	0.96	0.96	0.96
	0	0.03	0.05	0.18	0.46	0.74	0.85	0.94	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02	0.08	0.29	0.68	0.96	1.00	1.00	0.02	0.02	0.02	0.02	0.02	0.02	0.04
	0.1	0.40	0.37	0.28	0.23	0.30	0.23	0.12	0.06	0.06	0.06	0.06	0.05	0.06	0.06	0.40	0.65	0.60	0.16	0.00	0.00	0.00	0.06	0.06	0.07	0.07	0.08	0.13	0.23
5000	**0.2**	0.97	0.95	0.83	0.58	0.28	0.13	0.05	0.50	0.50	0.49	0.50	0.50	0.49	0.49	0.97	0.91	0.65	0.15	0.00	0.00	0.00	0.49	0.50	0.50	0.53	0.57	0.64	0.75
	0.3	0.98	0.96	0.85	0.60	0.34	0.18	0.08	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.98	0.93	0.64	0.15	0.00	0.00	0.00	0.92	0.92	0.91	0.92	0.93	0.94	0.95
	0.4	0.98	0.96	0.87	0.64	0.37	0.22	0.09	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.92	0.65	0.15	0.00	0.00	0.00	0.98	0.98	0.98	0.97	0.97	0.97	0.96

Table 3.6: Size and power of the sequential-G^* procedure with $H_T = \lfloor 1 + T^{\zeta} \rfloor$ knots compared to the g-test for a 1-factor Gegenbauer process with non-seasonal autoregressive dynamics $(1-\phi L)(1-2\cos\pi/2L+L^2)^d X_t = \varepsilon_t$ and seasonal autoregressive dynamics $(1-\phi L^4)(1-2\cos\pi/2L+L^2)^d X_t = \varepsilon_t$.

Proofs

Proof of Proposition 3.1:

1. Denote $Q_j = I_j/f(\lambda_j)$ and $\hat{Q}_j = I_j/\hat{f}(\lambda_j)$. For $\lambda \in [0,\pi]$ we have from Assumption 3.3 that $\hat{f}(\lambda) \xrightarrow{p} f(\lambda)$ as $T \to \infty$ if $k^{(i)} - k_0 = 0$. Therefore, by Slutsky's theorem $\hat{Q}_j \overset{d}{\Rightarrow} Q_j$, where $\overset{d}{\Rightarrow}$ denotes weak convergence.

   If $k^{(i)} = k_0$ in equation (3.4), then $\Delta^{k^{(i)}} X_t$ is reduced to the weakly dependent stationary process u_t for all $t = 1,...,T$. Therefore, Theorem 5.3.1 in Giraitis et al. (2012) applies and $(Q_1,...,Q_n)$ converges in distribution to a vector of independent exponentially distributed random variables. Denote by $G_n^M = max\{Q_1,...,Q_n\}$. Then we have $F^{G_n^M}(\tilde{z}) := P(\max_{1\le i\le n} Q_i \le \tilde{z}) = 1 - \prod_{i=1}^n P(Q_i > \tilde{z}) = 1 - (1 - \exp(-\tilde{z}))^n$. Thus, we have for our test statistic $G^* = G_n^M - \log n$ by standard arguments for all $x \in \Re$:

$$\begin{aligned} F^{G^*}(\tilde{z}) &= F^{G_n^M}(\tilde{z} + \log n) \\ &= (1 - (\exp(-\tilde{z} - \log n)))^n 1_{[0,\infty)}(\tilde{z} + \log n) \\ &= (1 - \frac{1}{n}\exp(-\tilde{z}))^n 1_{[-\log n,\infty)}(\tilde{z}) \end{aligned}$$

   and finally $\lim_{T\to\infty} F^{G^*}(\tilde{z}) = \exp(-\exp(-\tilde{z}))$, which proves the first part of the proposition as n depends on T and $-\log n$ converges to $-\infty$ for $n \to \infty$.□

2. To prove the consistency of the G_X^*-statistic if there are $k_0 - k^{(i)} > 0$ poles remaining in $\Delta^{k^{(i)}} X_t$, we follow Hidalgo and Soulier (2004) and consider a Fourier frequency $\gamma_{a,\kappa} = \gamma_a \pm \kappa\pi/T$ in the neighborhood of the cyclical frequency γ_a, where κ is a fixed positive number such that for every sample size $\gamma_{a,\kappa} = \lambda_j$ for some j. Then $\gamma_{a,\kappa} \to \gamma_a$, as $T \to \infty$, according to Theorem 1 in Hidalgo and Soulier (2004). Now write $I(\gamma_{a,\kappa}) = f(\gamma_{a,\kappa})\chi_T$, where χ_T is some random variable so that

$$\hat{Q}(\gamma_{a,\kappa}) = \frac{I(\gamma_{a,\kappa})}{\hat{f}(\gamma_{a,\kappa})} = \frac{f(\gamma_{a,\kappa})\chi_T}{\hat{f}(\gamma_{a,\kappa})}.$$

   From Lemma 2 of Hidalgo and Soulier (2004), $f(\gamma_{a,k}) \ge O(T^{2c_a})$ and χ_T has a distribution with no probability mass at zero. In addition to that, $\hat{f}(\gamma_{a,\kappa})$ is bounded according to Assumption 3.3. Therefore $\lim_{T\to\infty} \frac{\hat{Q}(\gamma_{a,\kappa})}{T^{2c_a}} = \chi_T^*$, where χ_T^* is a rescaling of χ_T. Then $\hat{Q}(\gamma_{a,\kappa}) = O_p(T^{2c_a})$, since for every $\epsilon > 0$ there exists a quantile c^*, such that

$$\lim_{T\to\infty} P\left(\left|\frac{\hat{Q}(\gamma_{a,\kappa})}{T^{2c_a}}\right| > c^*\right) < \epsilon = \lim_{T\to\infty} P\left(\chi_T^* > c^*\right) < \epsilon.$$

   Since $max_j(\hat{Q}_j) \ge \hat{Q}(\gamma_{a,\kappa})$, we have $G_X^* = max_j(\hat{Q}_j) - \log(n) \ge O_p(n^{2c_a} - \log(n))$, so that $\lim_{T\to\infty} G_X^* = \infty$ and $\lim_{T\to\infty} P(G_X^* < g_{crit}) = 0$ for all $g_{crit} = o(n^{2c_a} - \log(n))$. □

Proof of Proposition 3.2: The consistency of the model selection procedure can be established

in analogy to that of the sequential structural break test established in Proposition 11 of Bai (1997).

1. To derive the asymptotic probability to select the correct model order k_0 if a fixed significance level α is used, note that $P(\hat{k} = k_0) = 1 - P(\hat{k} < k_0) - P(\hat{k} > k_0)$. From Proposition 3.1, we have $\lim_{T\to\infty} P(G^*_{\Delta^{k^{(i)}}X_t} > g_{crit}) = 1$, for all $k^{(i)} < k_0$, so that $\lim_{T\to\infty} P(\hat{k} < k_0) = 0$. The second probability is equal to the probability of an α-error of the G^*-test applied to the series $\Delta^{k_0}X_t$, so that $\lim_{T\to\infty} P(\hat{k} > k_0) = \lim_{T\to\infty} P(G^*_{\Delta^{k_0}X_t} > g_{crit}) = \alpha$. Therefore, $\lim_{T\to\infty} P(\hat{k} = k_0) = 1 - \alpha$.□

2. For a decreasing significance level $\alpha(T) \to 0$, we have $\lim_{T\to\infty} P(G^*_{\Delta^{k^{(i)}}X_t} > g_{crit}(\alpha(T))) = 1$ for all $k^{(i)} < k_0$ and $\lim_{T\to\infty} P(\hat{k} > k_0) = \lim_{T\to\infty} P(G^*_{\Delta^{k_0}X_t} > g_{crit}(\alpha(T)) = \alpha(T)$, if the rate of decay of α is according to Proposition 1.2 chosen such that $g_{crit}(\alpha(T)) = o(n^{2c_a} - \log(n))$. Therefore, $\lim_{T\to\infty} P(\hat{k} = k_0) = 1$.□

Chapter 4

On the Memory of Products of Long Range Dependent Time Series

On the Memory of Products of Long Range Dependent Time Series

4.1 Introduction

Products of time series occur frequently in non-linear models such as the bilinear model, random coefficient models, or multiplicative noise models and they also play an important role as interaction terms in time series regressions. Therefore, it is of interest how time series properties such as long range dependence are translated from the factor series x_t and y_t to the product series $z_t = x_t y_t$. In this paper, it is shown that the transmission of memory critically depends on the means of the processes. While the memory of products is the maximum of the memory orders of the factor series if the means are non-zero, the memory order in the product series will be reduced for zero mean processes.

In a related literature Granger and Hallman (1991) and Corradi (1995) have studied the properties of non-linear transformations of integrated variables. For long memory time series Dittmann and Granger (2002) have derived the memory properties for transformations of zero mean time series if the transformation can be expressed as a finite sum of Hermite polynomials. The memory of products of long memory time series, however, has not been covered.[1]

The expectation of the product series $z_t = x_t y_t$ that we are interested in, is given by $E[z_t] = \sigma_x \sigma_y \rho_{xy} + \mu_x \mu_y$, where μ_x and μ_y denote the means of x_t and y_t, respectively, σ_x and σ_y are the standard deviations, and ρ_{xy} denotes the correlation between the two series. For general random variables x and y, with finite first and second moments, Goodman (1960) derived the variance of xy. Later, Bohrnstedt and Goldberger (1969) derived the exact covariance of the products xy and vw, where v and w form a second pair of random variables that fulfills the same moment conditions as x and y. According to Bohrnstedt and Goldberger (1969), the variance of z_t is given by

$$\begin{aligned}\sigma_z^2 = {} & \mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + E[(x-\mu_x)^2(y-\mu_y)^2] \\ & + 2\mu_x E[(x-\mu_x)(y-\mu_y)^2] \\ & + 2\mu_y E[(x-\mu_x)^2(y-\mu_y)] \\ & + 2\mu_x\mu_y\sigma_x\sigma_y\rho_{xy} - \sigma_x^2\sigma_y^2\rho_{xy}^2.\end{aligned}$$

The autocovariance function of the product of two weakly stationary time series x_t and y_t was

[1] Note that the application of a log-transformation does not mitigate this issue. It merely converts the problem into determining the memory of the sum of non-linearly transformed series, but the logarithm cannot be represented as required by Dittmann and Granger (2002). Therefore, the memory of the log-transformed series is unknown - as is that of the product.

derived by Wecker (1978). If both series are Gaussian, it is given by

$$\begin{aligned}\gamma_{xy}(\tau) =&\mu_x^2\gamma_y(\tau)+\mu_y^2\gamma_x(\tau)+\mu_x\mu_y[\xi(\tau)+\xi(-\tau)]\\ &+\gamma_x(\tau)\gamma_y(\tau)+\xi(\tau)\xi(-\tau),\end{aligned} \tag{4.1}$$

where $\xi(\tau)$ denotes the cross-covariance function at lag τ defined as $\xi(\tau) = E[(x_t-\mu_x)(y_{t-\tau}-\mu_y)]$. In the remainder of this paper, the memory properties of the product series z_t will be derived from the asymptotic behavior of (4.1), as $\tau \to \infty$. Definitions, assumptions and the main result are given in Section 4.2. Sections 4.3 and 4.4 extend these results to squares of long memory series and products of variables with common long range dependent factors. Conclusions are drawn in Section 4.5.

4.2 Persistence of Products of Long Memory Time Series

In the following, a time series x_t is a long memory series with parameter d_x if its spectral density $f_x(\lambda)$ at frequency λ obeys the power law

$$f_x(\lambda) \sim g_x(\lambda)\lambda^{-2d_x}, \tag{4.2}$$

as $\lambda \to 0_+$, or if its autocovariance function $\gamma_x(\tau)$ at lag τ is

$$\gamma_x(\tau) \sim G_x(\lambda)\tau^{2d_x-1}, \tag{4.3}$$

for $\tau \to \infty$. Here, $g_x(\lambda)$ and $G_x(\lambda)$ denote functions that are slowly varying at zero and infinitiy, respectively. As Beran et al. (2013) show, these definitions are equivalent under fairly general conditions. Hereafter, we write $x_t \sim LM(d_x)$ if x_t fulfills at least one of (4.2) or (4.3). For simplicity, we will treat g_x and G_x as constants. The properties of any x_t that is $LM(d_x)$ depend on the value of $d_x \in (-1/2, 1/2)$. For $d_x < 0$, the process is antipersistent, and $f_x(0) = 0$. If $d_x = 0$, $f_x(0) = g_x$ and the process has short memory. Finally, for $d_x > 0$, x_t is long range dependent.
Here, we follow Dittmann and Granger (2002) and distinguish between fractional integration and long memory. The reason is, that we derive the memory of $z_t = x_t y_t$ based on the behavior of $\gamma_{xy}(\tau)$ for large τ that is of the form specified in (4.3), so that its spectral density is of the form given in (4.2). A fractionally integrated process $\tilde{z}_t$, on the other hand, has spectral density $f_{\tilde{z}}(\lambda) = |1-e^{i\lambda}|^{-2d_{\tilde{z}}} g_{\tilde{z}}(\lambda)$, so that $f_{\tilde{z}}(\lambda) \sim g_{\tilde{z}}|\lambda|^{-2d_{\tilde{z}}}$, as $\lambda \to 0_+$, since $|1-e^{i\lambda}| \to \lambda$, as $\lambda \to 0_+$. While fractional integration is therefore a special case of long memory, the results given here only allow to draw conclusions about the memory properties of the product series.
For the main result we require the following assumptions.

Assumption 4.1. *$x_t \sim LM(d_x)$ and $y_t \sim LM(d_y)$ are weakly stationary and causal Gaussian processes, with $0 \le d_x, d_y < 0.5$ and finite second order moments.*

Assumption 4.2. *If $x_t, y_t \sim LM(d)$, then $x_t - \psi_0 - \psi_1 y_t \sim LM(d)$ for all $\psi_0, \psi_1 \in \mathbb{R}$.*

Assumption 4.1 is a simple regularity condition, whereas Assumption 4.2 precludes the presence of fractional cointegration. This will be relaxed in Section 4.4, where the case of a common long memory factor driving x_t and y_t is considered.
Given these assumptions, the memory of the product series $x_t y_t$ is characterized by the following proposition.

Proposition 4.1. *Under Assumptions 4.1 and 4.2 the product series $z_t = x_t y_t$ is $LM(d_z)$, with*

$$d_z = \begin{cases} \max(d_x, d_y), & \text{for } \mu_x, \mu_y \neq 0 \\ d_x, & \text{for } \mu_x = 0, \mu_y \neq 0 \\ d_y, & \text{for } \mu_y = 0, \mu_x \neq 0 \\ \max\left\{d_x + d_y - 1/2, 0\right\}, & \text{for } \mu_y = \mu_x = 0 \text{ and } S_{xy} \neq 0 \\ d_x + d_y - 1/2, & \text{for } \mu_y = \mu_x = 0 \text{ and } S_{xy} = 0, \end{cases}$$

where $S_{xy} = \sum_{\tau=-\infty}^{\infty} \gamma_x(\tau)\gamma_y(\tau)$.

Proof. *The autocovariance function of any $x_t y_t$ satisfying Assumption 4.1 is given by (4.1). This is a linear combination of the autocovariance functions $\gamma_x(\tau)$, $\gamma_y(\tau)$, the cross-covariance function $\xi(\tau)$ and interaction terms between them. Since long memory is defined in (4.3) by the shape of the autocovariance function for $\tau \to \infty$, we can determine the memory of $x_t y_t$ by finding the limit of $\gamma_{xy}(\tau)$. For $\tau \to \infty$, we can substitute $\gamma_x(\tau)$ and $\gamma_y(\tau)$ with $G_x \tau^{2d_x - 1}$ and $G_y \tau^{2d_y - 1}$ from (4.3). The asymptotic properties of the cross-covariance function $\xi(\tau)$ can be derived from results of Phillips and Kim (2007). In Theorem 1, they show that the autocovariance matrix $\Gamma_{XX}(\tau)$ of a q-dimensional multivariate fractionally integrated process X_t is*

$$[\Gamma_{XX}(\tau)]_{ab} = \frac{2 f_{u_a u_b}(0)\Gamma(1 - d_a - d_b)\sin(\pi d_b)}{\tau^{1 - d_a - d_b}} + O\left(\frac{1}{\tau^{2 - d_a - d_b}}\right),$$

where A_{ab} denotes the element in the ath row and bth column of the matrix A. The asymptotic expansion of the Fourier integral used to derive this result is not specific to fractionally integrated processes, but holds for long memory processes in general. It therefore follows, that $\xi(\tau) = G_{xy}\tau^{d_a + d_b - 1} + o(\tau^{d_a + d_b - 1})$. Furthermore, since by Assumption 4.1 both x_t and y_t are causal, $\xi(-\tau) = 0$, so that the last term in (4.1) drops out.
Therefore, as $\tau \to \infty$, we have

$$\gamma_{xy}(\tau) = \mu_x^2 G_y \tau^{2d_y - 1} + \mu_y^2 G_x \tau^{2d_x - 1} + \mu_x \mu_y G_{xy} \tau^{d_x + d_y - 1} + G_x G_y \tau^{2(d_x + d_y - 1)} + o(\tau^{d_x + d_y - 1}).$$

Now, considering the exponents and setting $d_x + d_y - 1 = 2\bar{d}_3 - 1$ *and* $2(d_x + d_y - 1) = 2\bar{d}_4 - 1$ *gives* $\bar{d}_3 = (d_x + d_y)/2$ *and* $\bar{d}_4 = (d_x + d_y - 1/2)$*, so that*

$$\gamma_{xy}(\tau) = \mu_x^2 G_y \tau^{2d_y - 1} + \mu_y^2 G_x \tau^{2d_x - 1} + \mu_x \mu_y G_{xy} \tau^{2\bar{d}_3 - 1} + G_x G_y \tau^{2\bar{d}_4 - 1} + o(\tau^{d_x + d_y - 1}). \tag{4.4}$$

Since $O(\tau^p) + O(\tau^q) = O(\tau^{\max(p,q)})$*, the autocovariance function* $\gamma_{xy}(\tau)$ *is dominated by the term with the largest memory parameter, as* $\tau \to \infty$*. The approximation error* $o(\tau^{d_x + d_y - 1})$ *vanishes, because* $d_x, d_y < 1/2$*. Depending on the values of* μ_x *and* μ_y*, different cases can be distinguished.*

1. *If* $\mu_x = \mu_y = 0$*, (4.4) is reduced to* $\gamma_{xy}(\tau) \approx G_x G_y \tau^{2\bar{d}_4 - 1}$*. Therefore, the memory of* $x_t y_t$ *would be given by* $\bar{d}_4 = (d_x + d_y - 1/2)$*, which can be negative so that the decay rate of the autocovariance function is that of an antipersistent LM process. However, in this case the long memory definition is only fulfilled if the spectral density is zero at the origin, which is equivalent to* $S_{xy} = 0$*. Otherwise the process is* $LM(0)$*.*

2. *If* $\mu_x = 0 \neq \mu_y$*, (4.4) becomes* $\gamma_{xy}(\tau) \approx \mu_y^2 G_x \tau^{2d_x - 1} + G_x G_y \tau^{2\bar{d}_4 - 1}$ *and the dominating term is the maximum of* d_x *and* $\bar{d}_4 = (d_x + d_y - 1/2)$*. This is* d_x*, because* $d_y < 1/2$*.*

3. *If* $\mu_y = 0 \neq \mu_x$*, by the same arguments, the memory is* d_y*.*

4. *Finally, if* $\mu_x, \mu_y \neq 0$*, the memory order is the maximum of* d_x*,* d_y*,* $(d_x + d_y)/2$ *and* $d_x + d_y - 1/2$*. Furthermore, since* $d_x, d_y < 1/2$*,* $\max\{d_x, d_y\}$ *will always be at least as large as the other two terms.*

If $\rho_{xy} = 0$*, we have* $G_{xy} = 0$*, so that the third term in (4.4) is zero. However, this does not affect the memory properties of the product series compared to the case when* $\rho_{xy} \neq 0$*, because the memory of the third term is always dominated by that of the first two terms since* $\max\{d_x, d_y\} \geq (d_x + d_y)/2$*.* □

It can immediately be seen from Proposition 4.1, that the means of x_t and y_t play a crucial role. If both means are non-zero, the asymptotic autocovariance function is dominated by the input series with the larger memory parameter. If both means are zero, the first three terms in (4.4) drop out, and the memory in z_t is $\bar{d}_4 = (d_x + d_y - 1/2)$. If one of the processes is mean zero, the memory is equal to that of this process, because $d_x, d_y > \bar{d}_4$.

Intuitively, if one thinks of long range dependence in terms of the persistence of deviations from the mean, it is obvious that this persistence vanishes if the series is multiplied with a zero mean *iid*-series. On the other hand, if the series is multiplied by a series with a non-zero mean, consecutive observations of the product series are likely to be located at the same side of the mean, so that the persistence remains intact.

Since the decay of the autocovariances of antipersistent *LM*-processes is initially very fast, the condition $S_{xy} = 0$ is basically a condition on the first autocovariances $\gamma_{xy}(\tau)$. As can be seen from (4.1), this is equal to the product $\gamma_{xy}(\tau) = \gamma_x(\tau)\gamma_y(\tau)$, if $\mu_x = \mu_y = 0$. The process can therefore only be *LM* with negative d, if $\gamma_x(\tau)$ and $\gamma_y(\tau)$ have different signs and the sum of their product

over the first leads and lags is close to the process variance. If $\gamma_{xy}(\tau)$ is negative but $S_{xy} \neq 0$, the process has antipersistent short memory. Finally, if $\gamma_{xy}(\tau)$ is positive, the process is simply $LM(0)$ if $d_x + d_y - 1/2$ is negative.[2]

4.3 Memory of Squared Series

An important special case of Proposition 4.1 arises if $x_t = y_t$, so that the product becomes the square of one series. This gives the following corollary.

Corollary 1. *For x_t satisfying Assumption 4.1, the square $z_t = x_t^2$ is $LM(d_z)$ with*

$$d_z = \begin{cases} d_x, & \text{if } \mu_x \neq 0 \\ \max\{2d_x - 1/2, 0\}, & \text{if } \mu_x = 0. \end{cases}$$

Proof. *Wecker (1978) shows that for x_t^2 equation (4.1) simplifies to*

$$\gamma_{xx}(\tau) = 4\mu_x^2\gamma_x(\tau) + 2\gamma_x^2(\tau).$$

For $\tau \to \infty$, this gives

$$\gamma_{xx}(\tau) \approx 4\mu_x^2 G_x \tau^{2d_x - 1} + 2G_x \tau^{2(2d_x - 1)}. \tag{4.5}$$

Again, from equating $2(2d_x - 1) = 2d_{sq} - 1$ we have $d_{sq} = 2d_x - 1/2$ so that $\gamma_{xx}(\tau) \approx 4\mu_x^2 G_x \tau^{2d_x - 1} + 2G_x \tau^{2d_{sq} - 1}$. Since $d_x < 1/2$, by Assumption 4.1, $d_{sq} < d_x$ so that the first term in (4.5) always dominates the second if $\mu_x \neq 0$. For $\mu_x = 0$, the memory is determined by the second term. This is $\max\{d_{sq}, 0\}$, for the reasons discussed in the proof of Proposition 4.1, above. Since $\gamma_x(\tau)^2 \geq 0$, the process cannot become antipersistent and the lower bound for the memory order of the square is always zero. □

Corollary 1 shows, that the memory of a squared series will be reduced if the series is mean zero, whereas the memory is unaffected by this non-linear transformation if it has a non-zero mean. In case of a reduction, the memory will be zero, for $d_x \leq 0.25$.

The memory of the squared zero mean series is also a special case of the results in Dittmann and Granger (2002). As discussed above, they derive the properties of non-linear transformations of fractionally integrated time series with zero mean and unit variance under the restriction that the non-linear transformation can be expressed as a finite sum of Hermite polynomials. Corollary 1 shows that the mean zero assumption is critical, since squaring the series does not cause a reduction in memory if it is not fulfilled.

[2] A similar issue arises in Dittmann and Granger (2002). However, they consider pure fractionally integrated processes and do not allow for short run dynamics. In this case all autocovariances are positive and the lower bound for the memory of the transformed series is zero.

Note that it is not possible to apply Proposition 4.1 together with Corollary 1 to determine the memory of higher order power transformations such as x_t^3, because the Gaussianity requirement in Assumption 4.1 is no longer satisfied by the squared series x_t^2.

4.4 Products of Fractionally Cointegrated Time Series

Another interesting special case is the product of series with a common factor, such as fractionally cointegrated series. Consider the following model

$$x_t = \beta_x + \delta_x u_t + \eta_t \tag{4.6}$$

$$y_t = \beta_y + \delta_y u_t + \varepsilon_t, \tag{4.7}$$

where u_t is $LM(d_u)$, $\eta_t \sim LM(d_\eta)$ and $\varepsilon_t \sim LM(d_\varepsilon)$ fulfill the conditions imposed in Assumption 4.1, and β_x and β_y are finite real constants. Furthermore, let η_t and ε_t be mean zero, with $d_\eta = d_u - b \geq 0$ and $d_\varepsilon = d_u - b - \epsilon \geq 0$, for some constants $b, \epsilon \geq 0$. Clearly, both y_t and x_t are driven by the common long range dependent factor u_t and they are both $LM(d_u)$. We therefore refer to them as series with common long memory. As before $0 \leq d_u < 0.5$ and of course δ_x, $\delta_y \neq 0$. We then obtain the following result.

Proposition 4.2. *Let y_t and x_t have common long memory of order d_u, so that they can be represented as in equations (4.6) and (4.7). Then the product $x_t y_t$ is $LM(d_u)$ if either $(\beta_x\delta_y + \beta_y\delta_x)$ or μ_u is unequal zero and it is $LM(\max\{2d_u - 1/2, 0\})$ if $\mu_x = \mu_y = \mu_u = 0$.*

Proof. *By substituting the relationships in (4.6) and (4.7) and rearranging the terms, the product $x_t y_t$ is given by*

$$x_t y_t = \underbrace{\beta_x\beta_y}_{I} + \underbrace{(\beta_x\delta_y + \beta_y\delta_x)u_t}_{II} + \underbrace{\delta_x\delta_y u_t^2}_{III} + \underbrace{(\beta_x + \delta_x u_t)\varepsilon_t}_{IV} + \underbrace{(\beta_y + \delta_y u_t)\eta_t}_{V} + \underbrace{\eta_t\varepsilon_t}_{VI}. \tag{4.8}$$

Proposition 3 in Chambers (1998) states that a linear combination of fractionally integrated processes is itself fractionally integrated, with an order of integration equal to the maximum of those in the linear combination. Chamber's arguments extend readily to long memory processes in general because they only make use of the long memory properties. More specifically, let $f_X(\lambda)$ be the spectral density matrix of the multivariate long memory process X_t that fulfills $[Re(f_X(\lambda))]_{ab} \sim g_{ab}|\lambda|^{-d_a-d_b}$, as $\lambda \to 0$. Than the linear combination $S_t = w'X_t$ has spectral density

$$f_S(\lambda) = w' f_X(\lambda) w \sim \sum_{a=1}^{q} w_a^2 G_{aa}|\lambda|^{-2d_a} + \sum_{a\neq b} w_a w_b G_{ab}|\lambda|^{-d_a-d_b}, \tag{4.9}$$

as $\lambda \to 0$. Since $O(|\lambda|^{-2d_a}) + O(|\lambda|^{-2d_b}) = O(|\lambda|^{-2\max(d_a,d_b)})$, as $\lambda \to 0$, and $2\max(d_a,d_b) > d_a + d_b$, the spectral density is proportionate to the largest term in the first sum on the right hand side of (4.9).

The memory order of $x_t y_t$ is therefore the maximum of the memory orders in terms I to VI in (4.8) and the memory of these individual terms can be determined from the results in Proposition 4.1 and Corollary 1.
In Proposition 4.2, two cases are distinguished. In the first one, either μ_u or $\beta_x\delta_y + \beta_y\delta_x$ are unequal zero. In the second, $\mu_x = \mu_y = \mu_u = 0$. These cases are considered separately.

1. *Whenever $\mu_u \neq 0$, III is $LM(d_u)$, by Corollary 1. Since we know from Proposition 4.1 that none of the other terms can have stronger memory than the original series, it follows directly from the result of Chambers (1998) that $x_t y_t$ will be $LM(d_u)$. The same holds true if $\beta_x\delta_y + \beta_y\delta_x \neq 0$, since u_t is $LM(d_u)$.*

2. *If $\mu_x = \mu_y = \mu_u = 0$, β_x and β_y in II are zero from (4.6) and (4.7) and term III is $LM(\max\{2d_u - 1/2, 0\})$, by Corollary 1. From Proposition 3 in Chambers (1998) the lower bound for the memory of the linear combination in (4.8) is therefore zero. By Proposition 4.1, the reduced memory of the terms IV, V and VI is $d_u + d_\varepsilon - 1/2$, $d_u + d_\eta - 1/2$ and $d_u + d_\varepsilon - 1/2$, respectively. Since $d_\varepsilon, d_\eta < d_u$, by definition, all of these terms are smaller than $2d_u - 1/2$. The memory of $x_t y_t$ is therefore determined by that of term III.* □

When comparing Propositions 4.1 and 4.2, one can see that the memory in the product series is less fugacious if the factor series have common long memory. As before, the nature of the transmission depends on the means of the series. However, since the square in term *III* determines the memory of $x_t y_t$ in equation (4.8), it is now the mean of the common factor that is crucial. Therefore, z_t can be d_u, even if $\mu_x = \mu_y = 0$.

4.5 Conclusion

This paper derives the memory of products of long memory time series. It is found that the nature of the transmission critically hinges on the means of the factor series. While the memory in the product series will be reduced if the means are zero, the memory of the more persistent factor series will be directly propagated if the factor series have non-zero means. These findings show that the property of long range dependence is less fugacious than it might seem from previous results on non-linear transformations of long memory series obtained by Dittmann and Granger (2002) that rely on a mean zero assumption.
One may conjecture that the transmission of long memory to other non-linear transformations may similarly depend on the means. Further research on the memory properties of a broader class of non-linear transformations would therefore be of great interest.

Chapter 5

Comparing Predictive Accuracy under Long Memory - With an Application to Volatility Forecasting -

Comparing Predictive Accuracy under Long Memory - With an Application to Volatility Forecasting -

Co-authored with Robinson Kruse and Michael Will.

5.1 Introduction

If the accuracy of competing forecasts is to be evaluated in a (pseudo-)out-of-sample setup, it has become standard practice to employ the test of Diebold and Mariano (1995) (hereafter DM test). Let $\widehat{y}_{1t}$ and $\widehat{y}_{2t}$ denote two competing forecasts for the forecast objective series y_t and let the loss function of the forecaster be given by $g(\cdot) \geq 0$. Forecast errors are defined as $e_{it} = y_t - \widehat{y}_{it}$ for $i = 1,2$ and the corresponding forecast error loss differential is denoted by $z_t = g(e_{1t}) - g(e_{2t})$. By only imposing restrictions on the loss differential z_t, instead of the forecast objective and the forecasts, Diebold and Mariano (1995) test the null hypothesis of equal predictive accuracy, i.e. $H_0 : E(z_t) = 0$, by means of a simple t-statistic for the mean of the loss differentials. In order to account for serial correlation, a long-run variance estimator such as the heteroscedasticity and autocorrelation consistent (HAC) estimator is applied (see Newey and West (1987), Andrews (1991) and Andrews and Monahan (1992)). For weakly dependent and second-order stationary processes this leads to an asymptotic standard normal distribution of the t-statistic.

Apart from the development of other forecast comparison tests such as those of West (1996) or Giacomini and White (2006), several direct extensions and improvements of the DM test have been proposed. Harvey et al. (1997) suggest a version that corrects for the bias of the long-run variance estimation in finite samples. A multivariate DM test is derived by Mariano and Preve (2012). To mitigate the well known size issues of HAC-based tests in finite samples of persistent short memory processes, Choi and Kiefer (2010) construct a DM test using the so-called fixed-bandwidth (or in short, fixed-b) asymptotics, originally introduced in Kiefer and Vogelsang (2005) (see also Patton (2015) and Li and Patton (2013)). Another extension of the DM test is proposed by Rossi (2005), who develops a DM test under near unit root asymptotics. However, all of these extensions fall into the classical $I(0)/I(1)$ framework.

In this paper, we study the situation if these assumptions on the loss differential do not apply and instead z_t follows a long memory process. Our first contribution is to show that long memory can be transmitted from the forecasts and the forecast objective to the forecast errors and subsequently to the forecast error loss differentials. We consider the case of a mean squared error (MSE) loss function and give conditions under which the transmission occurs and characterize the memory properties of the forecast error loss differential. As a second contribution, we show that the original DM test is invalidated in this case and suffers from severe upward size distortions. Third, we study two simple extensions of the DM statistic that allow valid inference under long (and short) memory. These extensions are the memory and autocorrelation consistent (MAC) estimator of Robinson (2005) (see also Abadir et al. (2009)) and the extended fixed-b

asymptotics (EFB) of McElroy and Politis (2012). The performance of these modified statistics is analyzed in a Monte Carlo study. Since these tests build on a restriction on the mean, the results allow broader conclusions about the relative performance of the MAC and the extended fixed-b approach, which is an interesting topic in its own right. We compare several bandwidth and kernel choices that allow recommendations for practical applications.

Our fourth contribution is an empirical application where we reconsider two recent extensions of the heterogeneous autoregressive model for realized volatility (HAR-RV) by Corsi (2009). First, we test whether forecasts obtained from HAR-RV type models can be improved by including information on model-free risk-neutral implied volatility which is measured by the CBOE volatility index (VIX). We find that short memory approaches (classic Diebold-Mariano test and fixed-b versions) reject the null hypothesis of equal predictive ability in favor of models including implied volatility. On the contrary, our long memory robust statistics do not indicate a significant improvement in forecast performance which implies that previous rejections might be spurious. The second issue we tackle relates to earlier work by Andersen et al. (2007) and Corsi et al. (2010), among others, who consider the decomposition of the quadratic variation of the log-price process into a continuous integrated volatility component and a discrete jump component. Here, we find that the separate treatment of continuous components and jump components significantly improves forecasts of realized variance for short forecast horizons even if the memory in the loss differentials is accounted for.

The rest of this paper is organized as follows. Section 5.2 reviews the classic Diebold-Mariano test and presents the fixed-b approach for the short memory case. Section 5.3 covers the case of long range dependence and contains our theoretical results on the transmission of long memory to the loss differential series. Two distinct approaches to design a robust t-statistic are discussed in Section 5.4. Section 5.5 contains our Monte Carlo study and in Section 5.6 we present our empirical results. Conclusions are drawn in Section 5.7. All proofs are contained in the Appendix.

5.2 Diebold-Mariano Test

Diebold and Mariano (1995) construct a test for $H_0 : E\left[g(e_{1t}) - g(e_{2t})\right] = E(z_t) = 0$, solely based on assumptions on the loss differential series z_t. Suppose that z_t follows the weakly stationary linear process

$$z_t = \mu_z + \sum_{j=0}^{\infty} \theta_j v_{t-j}\,, \tag{5.1}$$

where it is required that $|\mu_z| < \infty$ and $\sum_{j=0}^{\infty} \theta_j^2 < \infty$ hold. For simplicity of the exposition we additionally assume that $v_t \sim iid(0, \sigma_v^2)$. If $\widehat{y}_{1t}$ and $\widehat{y}_{2t}$ are performing equally good in terms of $g(\cdot)$, $\mu_z = 0$ holds, otherwise $\mu_z \neq 0$. The corresponding t-statistic is based on the sample mean

$\bar{z} = T^{-1}\sum_{t=1}^{T} z_t$ and an estimate ($\widehat{V}$) of the long-run variance $V = \lim_{T\to\infty} \mathrm{Var}(T^{\delta}(\bar{z}-\mu_z))$. The DM statistic is given by

$$t_{DM} = T^{\delta}\frac{\bar{z}}{\sqrt{\widehat{V}}}\,. \tag{5.2}$$

Under stationary short memory, we have $\delta = 1/2$, while the rate changes to $\delta = 1/2 - d$ under stationary long memory, with $0 < d < 1/2$ being the long memory parameter. The (asymptotic) distribution of this t-statistic hinges on the autocorrelation properties of the loss differential series z_t. In the following, we shall distinguish two cases: (1) z_t is a stationary short-memory process and (2) strong dependence in form of a long memory process is present in z_t as presented in Section 5.3.

5.2.1 Conventional Approach: HAC

For the estimation of the long-run variance V, Diebold and Mariano (1995) suggest to use the truncated long-run variance of an MA($h-1$) process for an h-step-ahead forecast. This is motivated by the fact that optimal h-step-ahead forecast errors of a linear time series process follow an MA($h-1$) process. Nevertheless, as pointed out by Diebold (2015), among others, the test is readily extendable to more general situations, if for example, HAC estimators are used (see also Clark (1999) for some early simulation evidence). The latter have become the standard estimators for the long-run variance. In particular,

$$\widehat{V}_{HAC} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{B}\right)\widehat{\gamma}_z(j)\,, \tag{5.3}$$

where $k(\cdot)$ is a user-chosen kernel function, B denotes the bandwidth and

$$\widehat{\gamma}_z(j) = \frac{1}{T}\sum_{t=|j|+1}^{T}(z_t - \bar{z})\left(z_{t-|j|} - \bar{z}\right)$$

is the usual estimator for the autocovariance of process z_t at lag j. The Diebold-Mariano statistic is given by

$$t_{HAC} = T^{1/2}\frac{\bar{z}}{\sqrt{\widehat{V}_{HAC}}}. \tag{5.4}$$

If z_t is weakly stationary with absolutely summable autocovariances $\gamma_z(j)$, it holds that $V = \sum_{j=-\infty}^{\infty}\gamma_z(j)$. Suppose that a functional central limit theorem applies for partial sums of z_t, so that $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} z_t \Rightarrow \sqrt{V}W(r)$ where $W(r)$ is a standard Brownian motion. Then, the t_{HAC}-statistic is asymptotically standard normal under the null hypothesis, i.e.

$$t_{HAC} \Rightarrow \mathcal{N}(0,1),$$

as $\sqrt{V}$ in (5.2) cancels out as long as $\widehat{V} \xrightarrow{p} V$ holds. For the sake of a comparable notation to the long memory case, note that $V = 2\pi f_z(0)$, where $f_z(0)$ is the spectral density function of z_t at frequency zero.

5.2.2 Fixed-bandwidth Approach

Even though nowadays the application of HAC estimators is standard practice, related tests are often found to be seriously size-distorted in finite samples, especially under strong persistence. It is assumed that the ratio $b = B/T \to 0$ as $T \to \infty$ in order to achieve a consistent estimation of the long-run variance V (see for instance Andrews (1991) for additional technical details). Kiefer and Vogelsang (2005) develop a new asymptotic framework in which the ratio B/T approaches a fixed constant $b \in (0,1]$ as $T \to \infty$. Therefore, it is called fixed-b inference as opposed to the classical small-b HAC approach where $b \to 0$.
In the case of fixed-b (FB), the estimator $\widehat{V}(k,b)$ does not converge to V any longer. Instead, $\widehat{V}(k,b)$ converges to V multiplied by a functional of a Brownian bridge process. In particular, $\widehat{V}(k,b) \Rightarrow VQ(k,b)$. Therefore, the corresponding t-statistic

$$t_{FB} = T^{1/2} \frac{\bar{z}}{\sqrt{\widehat{V}(k,b)}} \tag{5.5}$$

has a non-normal and non-standard limiting distribution, i.e.

$$t_{FB} \Rightarrow \frac{W(1)}{\sqrt{Q(k,b)}} .$$

Here, $W(r)$ is a standard Brownian motion on $r \in [0,1]$. Both, the choice of the bandwidth parameter b and the (twice continuously differentiable) kernel k appear in the limit distribution. For example, for the *Bartlett* kernel we have

$$Q(k,b) = \frac{2}{b}\left(\int_0^1 \widetilde{W}(r)^2 \mathrm{d}r - \int_0^{1-b} \widetilde{W}(r+b)\widetilde{W}(r)\mathrm{d}r \right),$$

with $\widetilde{W}(r) = W(r) - rW(1)$ denoting a standard Brownian bridge. Thus, critical values reflect the user choices on the kernel and the bandwidth even in the limit. In many settings, fixed-b inference is more accurate than the conventional HAC estimation approach. An example of its application to forecast comparisons are the aforementioned articles of Choi and Kiefer (2010), Patton (2015) and Li and Patton (2013), who apply both techniques (HAC and fixed-b) to compare exchange rate forecasts. Our Monte Carlo simulation study sheds additional light on their relative empirical performance.

5.3 Long Memory in Forecast Error Loss Differentials

5.3.1 Preliminaries

Under long-range dependence in z_t, one has to expect that neither conventional HAC estimators nor the fixed-b approach can be applied in this context without any further modification, since strong dependence such as fractional integration is ruled out by assumption. In particular, we show that HAC-based tests reject with probability one in the limit (as $T \to \infty$) if z_t has long memory. This claim is proven in our Proposition 5.5 (at the end of this section). As our finite-sample simulations clearly demonstrate, this implies strong upward size distortions and invalidates the use of the classic DM test statistic. Before we actually state these results formally, we first show that the loss differential z_t may exhibit long memory in various situations.
We start with a basic definition of stationary long memory time series.

Definition 5.1. *A time series a_t with spectral density $f_a(\lambda)$, with $\lambda \in [-\pi, \pi]$, has long memory with memory parameter $d_a \in (0, 1/2)$, if $f_a(\lambda) \sim L_f|\lambda|^{-2d_a}$ for $d_a \in (0, 1/2)$ as $\lambda \to 0$. $L_f(\cdot)$ is slowly varying at the origin. We write $a_t \sim LM(d_a)$.*

This is the usual definition of a stationary long memory process and Theorem 1.3 of Beran et al. (2013) states that under this restriction and mild regularity conditions, Definition 5.1 is equivalent to $\gamma_a(j) \sim L_\gamma|j|^{2d_a-1}$ as $j \to \infty$, where $\gamma_a(j)$ is the autocovariance function of a_t at lag j and $L_\gamma(\cdot)$ is slowly varying at infinity. If $d_a = 0$ holds, the process has short memory. Our results build on the asymptotic behavior of the autocovariances that have the long memory property from Definition 5.1. Whether this memory is generated by fractional integration can not be inferred. However, this does not affect the validity of the test statistics introduced in Section 5.4. We therefore adopt Definition 5.1 which covers fractional integration as a special case. A similar approach is taken by Dittmann and Granger (2002).[1]
Given Definition 5.1, we now state some assumptions regarding the long memory structure of the forecast objective and the forecasts.

Assumption 5.1 (Long Memory)**.** *The time series y_t, $\widehat{y}_{1t}$, $\widehat{y}_{2t}$ with expectations $E(y_t) = \mu_y$, $E(\widehat{y}_{1t}) = \mu_1$ and $E(\widehat{y}_{2t}) = \mu_2$ are causal Gaussian long memory processes (according to Definition 5.1) of orders d_y, d_1 and d_2, respectively.*

Similar to Dittmann and Granger (2002), we rely on the assumption of Gaussianity since no results for the memory structure of squares and cross-products of non-Gaussian long memory processes are available in the existing literature. It shall be noted that Gaussianity is only assumed for the derivation of the memory transmission from the forecasts and the forecast objective to the loss differential, but not for the subsequent results.

[1] Sometimes the terms long memory and fractional integration are used interchangeably. However, a stationary fractionally integrated process a_t has spectral density $f_a(\lambda) = |1 - e^{i\lambda}|^{-2d_a} G_a(\lambda)$, so that $f_a(\lambda) \sim G(\lambda)|\lambda|^{-2d_a}$ as $\lambda \to 0$, since $|1 - e^{i\lambda}| \to \lambda$ as $\lambda \to 0$. Therefore, fractional integration is a special case of long memory, but many other processes would satisfy Definition 5.1, too. Examples include non-causal processes and processes with trigonometric power law coefficients, as recently discussed in Kechagias and Pipiras (2015).

In the following, we make use of the concept of common long memory in which a linear combination of long memory series has reduced memory. The amount of reduction is labeled as b in accordance with the literature (similar to the symbol b in "fixed-b", but no confusion shall arise).

Definition 5.2 (Common Long Memory)**.** *The time series a_t and b_t have common long memory (CLM) if both a_t and b_t are $LM(d)$ and there exists a linear combination $c_t = a_t - \psi_0 - \psi_1 b_t$ with $\psi_0 \in \mathbb{R}$ and $\psi_1 \in \mathbb{R}\backslash 0$ such that $c_t \sim LM(d-b)$, for some $d \geq b > 0$. We write $a_t, b_t \sim CLM(d, d-b)$.*

For simplicity and ease of exposition, we first exclude the possibility of common long memory among the series. This assumption is relaxed later on.

Assumption 5.2 (No Common Long Memory)**.** *If $a_t, b_t \sim LM(d)$, then $a_t - \psi_0 - \psi_1 b_t \sim LM(d)$ for all $\psi_0 \in \mathbb{R}, \psi_1 \in \mathbb{R}$ and $a_t, b_t \in \{y_t, \widehat{y}_{1t}, \widehat{y}_{2t}\}$.*

In order to derive the long memory properties of the forecast error loss differential, we make use of a result in Leschinski (2016) that characterizes the memory structure of the product series $a_t b_t$ for two long memory time series a_t and b_t. Such products play an important role in the following analysis. The result is therefore shown as Proposition 5.1 below, for convenience.

Proposition 5.1 (Memory of Products)**.** *Let a_t and b_t be long memory series according to Definition 5.1 with memory parameters d_a and d_b, and means μ_a and μ_b, respectively. Then*

$$a_t b_t \sim \begin{cases} LM(\max\{d_a, d_b\}), & \text{for } \mu_a, \mu_b \neq 0 \\ LM(d_a), & \text{for } \mu_a = 0, \mu_b \neq 0 \\ LM(d_b), & \text{for } \mu_b = 0, \mu_a \neq 0 \\ LM(\max\{d_a + d_b - 1/2, 0\}), & \text{for } \mu_a = \mu_b = 0 \text{ and } S_{a,b} \neq 0 \\ LM(d_a + d_b - 1/2), & \text{for } \mu_a = \mu_b = 0 \text{ and } S_{a,b} = 0, \end{cases}$$

where $S_{a,b} = \sum_{j=-\infty}^{\infty} \gamma_a(j)\gamma_b(j)$ with $\gamma_a(\cdot)$ and $\gamma_b(\cdot)$ denoting the autocovariance functions of a_t and b_t, respectively.

Proposition 5.1 shows that the memory of products of long memory time series critically depends on the means μ_a and μ_b of the series a_t and b_t. If both series are mean zero, the memory of the product is either the maximum of the sum of the memory parameters of both factor series minus one half - or it is zero - depending on the sum of autocovariances. Since $d_a, d_b < 1/2$, this is always smaller than any of the original memory parameters. If only one of the series is mean zero, the memory of the product $a_t b_t$ is determined by the memory of this particular series. Finally, if both series have non-zero means, the memory of the product is equal to the maximum of the memory orders of the two series.
It should be noted, that Proposition 5.1 makes a distinction between antipersistent series and short memory series, if the processes have zero means and $d_a + d_b - 1/2 < 0$. Our results below,

however, do not require this distinction. The reason for this is that a linear combination involving the square of at least one of the series appears in each case, and these cannot be anti-persistent long memory processes (cf. the proofs of Propositions 5.2 and 5.4 for details).
As discussed in Leschinski (2016), Proposition 5.1 is related to the results in Dittmann and Granger (2002), who consider the memory of non-linear transformations of zero mean long memory time series that can be represented through a finite sum of Hermite polynomials. Their results include the square a_t^2 of a time series which is also covered by Proposition 5.1 if $a_t = b_t$. If the mean is zero ($\mu_a = 0$), we have $a_t^2 \sim LM(\max\{2d_a - 1/2, 0\})$. Therefore, the memory is reduced to zero if $d \leq 1/4$. However, as can be seen from Proposition 5.1, this behavior depends critically on the expectation of the series.
Since it is the most widely used loss function in practice, we focus on the MSE loss function. Let $e_{it} = y_t - \widehat{y}_{it}$ denote the i-th forecast error, then the quadratic forecast error loss differential is given by

$$\begin{aligned} z_t &= e_{1t}^2 - e_{2t}^2 = (y_t - \widehat{y}_{1t})^2 - (y_t - \widehat{y}_{2t})^2 \\ &= \widehat{y}_{1t}^2 - \widehat{y}_{2t}^2 - 2y_t(\widehat{y}_{1t} - \widehat{y}_{2t}). \end{aligned} \tag{5.6}$$

Note that even though the forecast objective y_t as well as the forecasts $\widehat{y}_{it}$ in (5.6), have time index t, the representation is quite versatile. It allows for forecasts generated from time series models where $\widehat{y}_{it} = \sum_{s=1}^{T} \phi_s y_{t-s}$ as well as predictive regressions with $\widehat{y}_{it} = \beta' x_{t-s}$, where β is a $w \times 1$ parameter vector and x_{t-s} is a vector of w explanatory variables lagged by s periods. In addition to that, even though estimation errors are not considered explicitly, they would be reflected by the fact that $E[y_t|\Psi_{t-h}] \neq \widehat{y}_{it|t-h}$, where Ψ_{t-h} is the information set available at the forecast origin $t-h$. This means that forecasts are biased in presence of estimation error, even if the model employed corresponds to the true data generating process. The forecasts are also not restricted to be obtained from a linear model. Similar to the Diebold-Mariano test, which is solely based on a single assumption on the forecast error loss differential (5.6), the following results are derived by assuming certain properties of the forecasts and the forecast objective. Therefore, we follow Diebold and Mariano (1995) and do not impose direct restrictions on the way forecasts are generated.

5.3.2 Transmission of Long Memory to the Loss Differential

Following the introduction of the necessary definitions and a preliminary result, we now present the result for the memory order of z_t defined via (5.6) in Proposition 5.2. It is based on the memory of y_t, $\widehat{y}_{1t}$ and $\widehat{y}_{2t}$ and assumes the absence of common long memory for simplicity.

Proposition 5.2 (Memory Transmission without CLM)**.** *Under Assumptions 5.1 and 5.2, the*

forecast error loss differential in (5.6) is $z_t \sim LM(d_z)$*, where*

$$d_z = \begin{cases} \max\left\{d_y,\ d_1,\ d_2\right\}, & \text{if } \mu_1 \neq \mu_2 \neq \mu_y \\ \max\left\{d_1,\ d_2\right\}, & \text{if } \mu_1 = \mu_2 \neq \mu_y \\ \max\left\{2d_1 - 1/2,\ d_2,\ d_y\right\}, & \text{if } \mu_1 = \mu_y \neq \mu_2 \\ \max\left\{2d_2 - 1/2,\ d_1,\ d_y\right\}, & \text{if } \mu_1 \neq \mu_y = \mu_2 \\ \max\left\{2\max\left\{d_1,\ d_2\right\} - 1/2,\ d_y + \max\left\{d_1,\ d_2\right\} - 1/2,\ 0\right\}, & \text{if } \mu_1 = \mu_2 = \mu_y. \end{cases}$$

Proof: *See the Appendix.*

The basic idea of the proof relates to Proposition 3 of Chambers (1998). It shows that the long-run behavior of a linear combination of long memory series is dominated by the series with the strongest memory. Since we know from Proposition 5.1 that the means μ_1,μ_2 and μ_y play an important role for the memory of a squared long memory series, we set $y_t = y_t^* + \mu_y$ and $\widehat{y}_{it} = \widehat{y}_{it}^* + \mu_i$, so that the starred series denote the demeaned series and μ_i denotes the expected value of the respective series. Straightforward algebra yields

$$z_t = \widehat{y}_{1t}^{*2} - \widehat{y}_{2t}^{*2} - 2\left[y_t^*(\mu_1 - \mu_2) + \widehat{y}_{1t}^*(\mu_y - \mu_1) + \widehat{y}_{2t}^*(\mu_y - \mu_2)\right] - 2\left[y_t^*(\widehat{y}_{1t}^* - \widehat{y}_{2t}^*)\right] + const. \tag{5.7}$$

From (5.7) it is apparent that z_t is a linear combination of (i) the squared forecasts $\widehat{y}_{1t}^{*2}$ and $\widehat{y}_{2t}^{*2}$, (ii) the forecast objective y_t, (iii) the forecast series $\widehat{y}_{1t}^*$ and $\widehat{y}_{2t}^*$ and (iv) products of the forecast objective with the forecasts, i.e. $y_t^*\widehat{y}_{1t}^*$ and $y_t^*\widehat{y}_{2t}^*$. The memory of the squared series and the product series is determined in Proposition 5.1, from which the zero mean product series $y_t^*\widehat{y}_{it}^*$ is $LM(\max\left\{d_y + d_i - 1/2,\ 0\right\})$ or $LM(d_y + d_i - 1/2)$. Moreover, the memory of the squared zero mean series $\widehat{y}_{it}^{*2}$ is $\max\{2d_i - 1/2,\ 0\}$. By combining these results with that of Chambers (1998), the memory of the loss differential z_t is the maximum of all memory parameters of the components in (5.7). Proposition 5.2 then follows from a case-by-case analysis.

Proposition 5.2 demonstrates the transmission of long memory from the forecasts $\widehat{y}_{1t}$, $\widehat{y}_{2t}$ and the forecast objective y_t to the loss differential z_t. The nature of this transmission, however, critically hinges on the (un)biasedness of the forecasts. If both forecasts are unbiased (i.e. if $\mu_1 = \mu_2 = \mu_y$), the memory from all three input series is reduced and the memory of the loss differential z_t is equal to the maximum of the maximum of (i) these reduced orders and (ii) zero. Therefore, only if memory parameters are small enough such that $d_y + \max\{d_1 + d_2\} < 1/2$, the memory of the loss differential z_t is reduced to zero. In all other cases, there is a transmission of dependence from the forecast and/or the forecast objective to the loss differential. The reason for this can immediately be seen from (5.7). Note that the terms in the first bracket have larger memory than the remaining ones, because $d_i > 2d_i - 1/2$ and $\max\left\{d_y, d_i\right\} > d_y + d_i - 1/2$. Therefore, these terms dominate the memory of the products and squares whenever biasedness is present,

i.e. $\mu_i - \mu_y \neq 0$ holds. Interestingly, the transmission of memory from the forecast objective y_t is prevented, if both forecasts have equal bias - that is $\mu_1 = \mu_2$. On the contrary, if $\mu_1 \neq \mu_2$, d_z is at least as high as d_y.

5.3.3 Memory Transmission under Common Long Memory

The results in Proposition 5.2 are based on Assumption 5.2 that precludes common long memory among the series. Of course, in practice it is likely that such an assumption is violated. In fact, it can be argued that reasonable forecasts of long memory time series should have common long memory with the forecast objective. Therefore, we relax this restrictive assumption and replace it with Assumption 5.3, below.

Assumption 5.3 (Common Long Memory)**.** *The causal Gaussian process x_t has long memory according to Definition 5.1 of order d_x with expectation $E(x_t) = \mu_x$. If $a_t, b_t \sim CLM(d_x, d_x - b)$, then they can be represented as $y_t = \beta_y + \xi_y x_t + \eta_t$ for $a_t, b_t = y_t$ and $\widehat{y}_{it} = \beta_i + \xi_i x_t + \varepsilon_{it}$, for $a_t, b_t = \widehat{y}_{it}$, with $\xi_y, \xi_i \neq 0$. η_t and ε_{it} are mean zero causal Gaussian long memory processes with parameters d_η and $d_{\varepsilon_{it}}$ fulfilling $1/2 > d_x > d_\eta, d_{\varepsilon_i} \geq 0$, for $i = 1, 2$.*

Assumption 5.3 restricts the common long memory to be of a form so that both series a_t and b_t can be represented as linear functions of their joint factor x_t. This excludes more complicated forms of dependence that are sometimes considered in the cointegration literature such as non-linear or time-varying cointegration.
We know from Proposition 5.2 that the transmission of memory critically depends on the biasedness of the forecasts which leads to a complicated case analysis. If common long memory according to Assumption 5.3 is allowed for, this leads to an even more complex situation since there are several possible relationships: CLM of y_t with one of the $\widehat{y}_{it}$, CLM of both $\widehat{y}_{it}$ with each other, but not with y_t, and CLM of each $\widehat{y}_{it}$ with y_t. Each of these situations has to be considered with all possible combinations of the ξ_a and the μ_a for all $a \in \{y, 1, 2\}$. To deal with this complexity, we focus on two important special cases: (i) at least one forecast is biased and (ii) all forecasts are unbiased (and $\xi_a = \xi_b$ if a_t and b_t are in a common long memory relationship).
Situation (i) is similar to the first four cases considered in Proposition 5.2. By substituting the linear relations from Assumption 5.3 for those series involved in the CLM relationship in the loss differential $z_t = \widehat{y}_{1t}^2 - \widehat{y}_{2t}^2 - 2y_t(\widehat{y}_{1t} - \widehat{y}_{2t})$ and again setting $a_t = a_t^* + \mu_a$ for those series that are not involved in the CLM relationship, it is possible to find expressions that are analogous to (5.7). Since analogous terms to those in the first bracket of (5.7) appear in each case, it is possible to focus on the transmission of memory from the forecasts and the objective function to the loss differential. We therefore obtain the following result.

Proposition 5.3 (Memory Transmission with Biased Forecasts and CLM)**.** *Under Assumptions 5.1 and 5.3, the forecast error loss differential in (5.6) is* $z_t \sim LM(d_z)$*, where*

$$d_z \geq \begin{cases} d_1, & \text{if } \mu_1 \neq \mu_y \\ d_2, & \text{if } \mu_2 \neq \mu_y \\ d_y, & \text{if } \mu_1 \neq \mu_2. \end{cases}$$

Proof: *See the Appendix.*

Proposition 5.3 states that the transmission of memory remains the same as in the absence of common long memory, given that the forecasts are biased. As in (5.7) before, if the forecasts are biased (or have different biases) the memory of the de-meaned series y_t^*, $\widehat{y}_{1t}^*$ and $\widehat{y}_{2t}^*$ dominate that of the other terms. However, if two of those terms appear, it is unclear which one of them is larger - therefore the inequalities in Proposition 5.3.
The second special case (ii) refers to a situation of unbiasedness similar to the last case in Proposition 5.2. In addition to that, it is assumed that $\xi_a = \xi_b$, if a_t and b_t are in a common long memory relationship. To understand the role of the coefficients ξ_a and ξ_b of the common long memory factor x_t driving both series, note that the forecast errors $y_t - \widehat{y}_{it}$ impose a cointegrating vector of $(1,-1)$. A different scaling of the forecast objective and the forecasts is not possible. In the case of CLM between y_t and $\widehat{y}_{it}$, for example, we have from Assumption 5.3 that $y_t - \widehat{y}_{it} = \beta_y - \beta_i + x_t(\xi_y - \xi_i) + \eta_t - \varepsilon_{it}$, so that x_t does not disappear from the linear combination if the scaling parameters ξ_y and ξ_i are different from each other. Hence, we have the following result.

Proposition 5.4 (Memory Transmission with Unbiased Forecasts and CLM)**.** *Under Assumptions 5.1 and 5.3, and if* $\mu_y = \mu_1 = \mu_2$ *and* $\xi_y = \xi_a = \xi_b$*, then* $z_t \sim LM(d_z)$*, with*

$$d_z = \begin{cases} \max\left\{d_2 + \max\left\{d_x, d_\eta\right\} - 1/2,\ 2\max\left\{d_x, d_2\right\} - 1/2,\ d_{\varepsilon_1}\right\}, & \text{if } y_t, \widehat{y}_{1t} \sim CLM(d_x, d_x - \widetilde{b}) \\ \max\left\{d_1 + \max\left\{d_x, d_\eta\right\} - 1/2,\ 2\max\left\{d_x, d_1\right\} - 1/2,\ d_{\varepsilon_2}\right\}, & \text{if } y_t, \widehat{y}_{2t} \sim CLM(d_x, d_x - \widetilde{b}) \\ \max\left\{\max\left\{d_x,\ d_y\right\} + \max\left\{d_{\varepsilon_1},\ d_{\varepsilon_2}\right\} - 1/2,\ 0\right\}, & \text{if } \widehat{y}_{1t}, \widehat{y}_{2t} \sim CLM(d_x, d_x - \widetilde{b}) \\ \max\left\{d_\eta + \max\left\{d_{\varepsilon_1},\ d_{\varepsilon_2}\right\} - 1/2,\ 2\max\left\{d_{\varepsilon_1},\ d_{\varepsilon_2}\right\} - 1/2,\ 0\right\}, & \text{if } y_t, \widehat{y}_{1t} \sim CLM(d_x, d_x - \widetilde{b}) \\ & \text{and } y_t, \widehat{y}_{2t} \sim CLM(d_x, d_x - \widetilde{b}). \end{cases}$$

Here, $0 < \widetilde{b} \leq 1/2$ denotes a generic constant for the reduction in memory.

Proof: *See the Appendix.*

Proposition 5.4 shows that the memory of the forecasts and the objective variable can indeed cancel out if the forecasts are unbiased and if they have the same factor loading on x_t (i.e. if

$\xi_1 = \xi_2 = \xi_y$). However, in the first two cases, the memory of the error series ε_{1t} and ε_{2t} imposes a lower bound on the memory of the loss differential. Furthermore, even though the memory can be reduced to zero in the third and fourth case, this situation only occurs if the memory orders of x_t, y_t and the error series are sufficiently small. Otherwise, the memory is reduced, but does not vanish.

The results in Propositions 5.2, 5.3 and 5.4 show that long memory can be transmitted from forecasts or the forecast objective to the forecast error loss differentials. This situation can arise naturally in many practical situations. First, of course the forecast objective might be a long memory time series. Second, from Proposition 3 in Chambers (1998), forecasts that are based on linear combinations - such as predictive regressions - exhibit long memory if they include a long memory variable.

Our results also show that the biasedness of the forecasts plays an important role for the transmission of dependence to the loss differentials. In practical situations, it might be overly restrictive to impose exact unbiasedness (under which memory would be reduced according to Proposition 5.4). Our empirical application regarding the predictive ability of the VIX serves as an example since it is a biased forecast of future quadratic variation due to the existence of a variance risk premium (see Section 5.6).

It is well established that estimation errors might imply biased forecasts. This issue might be of less importance in a setup where the estimation period grows at a faster rate than the (pseudo-) out-of-sample period that is used for forecast evaluation. For the DM test however, it is usually assumed that this is not the case. Otherwise, it could not be used for the comparison of forecasts from nested models due to a degenerated limiting distribution (cf. Giacomini and White (2006) for a discussion). Instead, the sample of size T^* is split into an estimation period T_E and a forecasting period T such that $T^* = T_E + T$ and it is assumed that T grows at a faster rate than T_E so that $T_E/T \to 0$ as $T^* \to \infty$. Therefore, the estimation error shrinks at a lower rate than the growth rate of the evaluation period and it remains relevant, asymptotically.

Finally, even optimal forecasts can be strongly persistent for long forecast horizons. It is well known that the forecast errors of an optimal h-step-ahead forecast follow an MA$(h-1)$ process. The coefficients of this process are given by the first $h-1$ coefficients in the MA(∞) representation of the objective series y_t. Therefore, forecast errors of the h-step forecast for $h \to \infty$ have long memory if the underlying series y_t has long memory as well. For larger forecast horizons long memory processes are therefore a better approximation to the true dependence structure than short memory processes.

5.3.4 Asymptotic and Finite-Sample Behaviour under Long Memory

After confirming that forecast error loss differentials can exhibit long memory, we now consider the effect of long memory on the HAC-based Diebold-Mariano test. The following Proposition 5.5 establishes that the size of the test approaches unity, as $T \to \infty$. Thus, the test indicates with probability one that one of the forecasts is superior to the other one, even if both tests

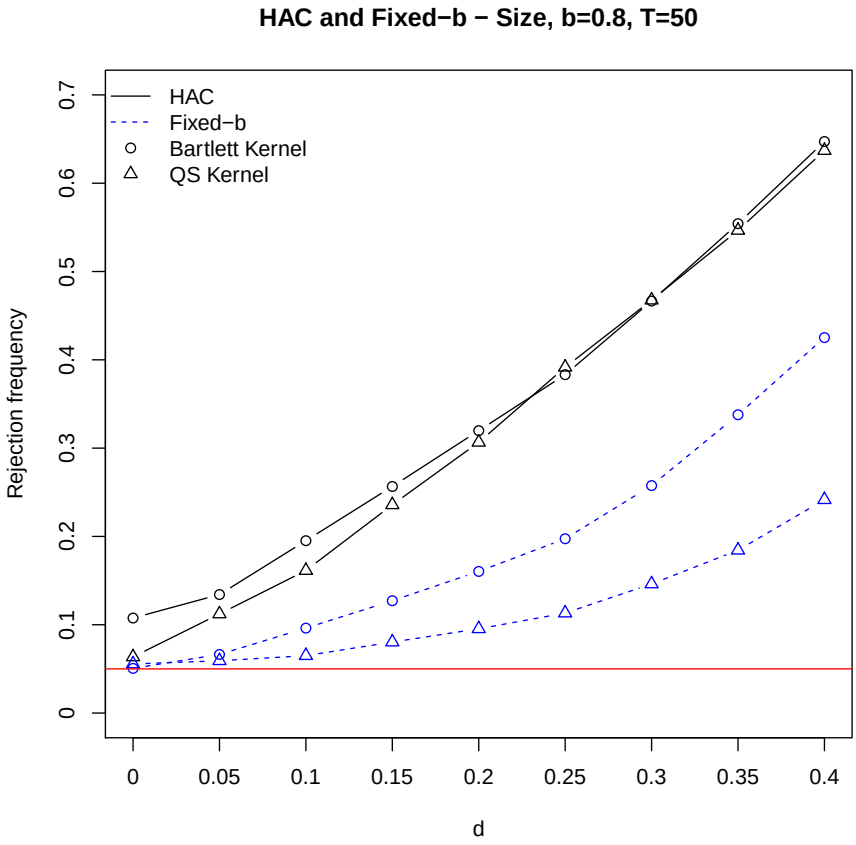

Figure 5.1: Size of the t_{HAC}- and t_{FB}-tests with $T = 50$ for different values of the memory parameter d.

perform equally in terms of $g(\cdot)$.

Proposition 5.5 (DM under Long Memory)**.** *For $z_t \sim LM(d)$ with $d \in (0,1/4) \cup (1/4,1/2)$, the asymptotic size of the t_{HAC}-statistic equals unity as $T \to \infty$.*

Proof: *See the Appendix.*

This result shows that inference based on HAC estimators is asymptotically invalid under long memory. At the point $d = 1/4$, the asymptotic distribution of the t_{HAC}-statistic changes from normality to a Rosenblatt-type distribution which explains the discontinuity, see Abadir et al. (2009). In order to explore to what extent this finding also affects the finite-sample performance of the t_{HAC}- and t_{FB}-statistics, we conduct a small-scale Monte Carlo experiment as an illustration. The results shown in Figure 5.1 are obtained with $M = 5000$ Monte Carlo repetitions. We simulate samples of $T = 50$ observations from a fractionally integrated process using different values of the memory parameter d in the range from 0 to 0.4. The HAC estimator and the fixed-b approach are implemented with the commonly used Bartlett- and Quadratic Spectral (QS) kernels.[2]

[2]The bandwidth parameter of the fixed-b estimator is set to $b = 0.8$, since using a larger fraction of the autocor-

As demonstrated by Kiefer and Vogelsang (2005), the fixed-b approach works exceptionally well in the case of $d = 0$, with the Bartlett and QS kernel achieving approximately equal size control. The t_{HAC}-statistic behaves more liberal than the fixed-b approach and, as stated in Andrews (1991), better size control is provided if the Quadratic Spectral kernel is used. If the memory parameter d is positive, we observe that both tests severely over-reject the null hypothesis. For $d = 0.4$, the size of the HAC-based test is approximately 65% and that of the fixed-b version using the Bartlett kernel is around 40%. We therefore find that the size distortions are not only an asymptotic phenomenon, but they are already severe in samples of just $T = 50$ observations. Moreover, even for small deviations of d from zero, both tests are over-sized. These findings motivate the use of long memory robust procedures.

5.4 Long-Run Variance Estimation under Long Memory

Since conventional HAC estimators lead to spurious rejections under long memory, it is necessary to consider memory robust long-run variance estimators. To the best of our knowledge only two extensions of this kind are available in the literature: the memory and autocorrelation consistent (MAC) estimator of Robinson (2005) and an extension of the fixed-b estimator from McElroy and Politis (2012). Note that we do not assume that forecasts are obtained from some specific class of model. We merely extend the typical assumptions of Diebold and Mariano (1995) on the loss differentials so that long memory is allowed.

5.4.1 MAC Estimator

The MAC estimator is developed by Robinson (2005) and further explored and extended by Abadir et al. (2009). Albeit stated in a somewhat different form, the same result is derived independently by Phillips and Kim (2007), who consider the long-run variance of a multivariate fractionally integrated process. Robinson (2005) assumes that z_t is linear and that for $\lambda \to 0$ its spectral density fulfills

$$f(\lambda) = b_0|\lambda|^{-2d} + o(|\lambda|^{-2d}),$$

with $b_0 > 0$, $|\lambda| \leq \pi$, $d \in (-1/2, 1/2)$ and $b_0 = \lim_{\lambda \to 0} |\lambda|^{2d} f(\lambda)$ (cf. Assumption L, Abadir et al. (2009)).[3] Among others, this assumption covers stationary and invertible ARFIMA processes. A key result for the MAC estimator is that as $T \to \infty$

$$\mathrm{Var}\left(T^{1/2-d}\bar{z}\right) \to b_0 p(d)$$

relations provides a higher emphasis on size control (cf. Kiefer and Vogelsang (2005)). Other bandwidth choices lead to similar results.

[3] For notational convenience, here we drop the index z from the spectral density and the memory parameter.

with

$$p(d) = \begin{cases} \frac{2\Gamma(1-2d)\sin(\pi d)}{d(1+2d)} & \text{if } d \neq 0\,, \\ 2\pi & \text{if } d = 0\,. \end{cases}$$

The case of short memory ($d = 0$) yields the familiar result that the long-run variance of the sample mean equals $2\pi b_0 = 2\pi f(0)$. Hence, estimation of the long-run variance requires estimation of $f(0)$ in the case of short memory. If long memory is present in the data generating process, estimation of the long-run variance additionally hinges on the estimation of d. The MAC estimator is therefore given by

$$\widehat{V}(\widehat{d}, m_d, m) = \widehat{b}_m(\widehat{d})p(\widehat{d})\,.$$

In more detail, the estimation of V works as follows: First, if the estimator for d fulfills the condition $\widehat{d} - d = o_p(1/\log T)$, plug-in estimation is valid (cf. Abadir et al. (2009)). Thus, $p(d)$ can simply be estimated through $p(\widehat{d})$. A popular estimator that fulfills this rather weak requirement is the local Whittle estimator with bandwidth $m_d = [T^q]$, where $0 < q < 1$ denotes a generic bandwidth parameter. Many other estimation approaches (e.g. log-periodogram estimation, etc.) would be a possibility as well. Next, b_0 can be estimated consistently by

$$\widehat{b}_m(\widehat{d}) = m^{-1} \sum_{j=1}^{m} \lambda_j^{2\widehat{d}} I_T(\lambda_j)\,,$$

where $I_T(\lambda_j)$ is the periodogram (which is independent of $\widehat{d}$),

$$I_T(\lambda_j) = (2\pi T)^{-1} \left| \sum_{t=1}^{T} \exp(\mathrm{i}t\lambda_j) z_t \right|^2$$

and $\lambda_j = 2\pi j/T$ are the Fourier frequencies for $j = 1, ..., \lfloor T/2 \rfloor$. Here, $\lfloor \cdot \rfloor$ denotes the largest integer smaller than its argument. The bandwidth m is determined according to $m = \lfloor T^q \rfloor$ such that $m \to \infty$ and $m = o(T/(\log T)^2)$.

The MAC estimator is consistent as long as $\widehat{d} \xrightarrow{p} d$ and $\widehat{b}_m(\widehat{d}) \xrightarrow{p} b_0$. These results hold under very weak assumptions - neither linearity of z_t nor Gaussianity are required. Under somewhat stronger assumptions the t_{MAC}-statistic is also normal distributed (see Theorem 3.1. of Abadir et al. (2009)):

$$t_{MAC} \Rightarrow \mathcal{N}(0,1)\,.$$

The t-statistic using the feasible MAC estimator can be written as

$$t_{MAC} = T^{1/2-\widehat{d}} \frac{\bar{z}}{\sqrt{\widehat{V}(\widehat{d}, m_d, m)}},$$

with m_d and m being the bandwidths for estimation of d and b_0, respectively.
It shall be noted that Abadir et al. (2009) also consider long memory versions of the classic HAC estimators. However, these extensions have two important shortcomings. First, asymptotic normality is lost for $1/4 < d < 1/2$ which complicates inference remarkably as d is generally unknown. Second, the extended HAC estimator is very sensitive towards the bandwidth choice as the MSE-optimal rate depends on d. On the contrary, the MAC estimator is shown to lead to asymptotically standard normally distributed t-ratios for the whole range of values $d \in (-1/2, 1/2)$. Moreover, the MSE-optimal bandwidth choice $m = [T^{4/5}]$ is independent of d. Thus, we focus on the MAC estimator and do not consider extended HAC estimators further.

5.4.2 Extended Fixed-Bandwidth Approach

Following up on the work by Kiefer and Vogelsang (2005), McElroy and Politis (2012) extend the fixed-bandwidth approach to long range dependence. Their approach is similar to the one of Kiefer and Vogelsang (2005) in many respects, as can be seen below. The test statistic suggested by McElroy and Politis (2012) is given by

$$t_{EFB} = T^{1/2} \frac{\bar{z}}{\sqrt{\widehat{V}(k,b)}}.$$

In contrast to the t_{MAC}-statistic, the t_{EFB}-statistic involves a scaling of $T^{1/2}$. This has an effect on the limit distribution which depends on the memory parameter d. Analogously to the short memory case, the limiting distribution is derived by assuming that a functional central limit theorem for the partial sums of z_t applies, so that

$$t_{EFB} \Rightarrow \frac{W_d(1)}{\sqrt{Q(k,b,d)}},$$

where $W_d(r)$ is a fractional Brownian motion and $Q(k,b,d)$ depends on the fractional Brownian bridge $\widetilde{W}_d(r) = W_d(r) - rW_d(1)$. Furthermore, $Q(k,b,d)$ depends on the first and second derivatives of the kernel $k(\cdot)$. In more detail, for the *Bartlett* kernel we have

$$Q(k,b,d) = \frac{2}{b}\left(\int_0^1 \widetilde{W}_d(r)^2 \mathrm{d}r - \int_0^{1-b} \widetilde{W}_d(r+b)\widetilde{W}_d(r)\mathrm{d}r\right)$$

and thus, a similar structure as for the short memory case. Further details and examples can be found in McElroy and Politis (2012). The joint distribution of $W_d(1)$ and $\sqrt{Q(k,b,d)}$ is found through their joint Fourier-Laplace transformation, see Fitzsimmons and McElroy (2010). It is symmetric around zero and has a cumulative distribution function which is continuous in d.
Besides the similarities to the short memory case, there are some important conceptual differences to the MAC estimator. First, the MAC estimator belongs to the class of "small-b" estimators in the sense that it estimates the long-run variance directly, whereas the fixed-b approach leads also in the long memory case to an estimate of the long-run variance multiplied by a

functional of a *fractional* Brownian bridge. Second, the limiting distribution of the t_{EFB}-statistic is not a standard normal, but rather depending on the chosen kernel k, the fixed-bandwidth parameter b and the long memory parameter d. While the first two are user-specific, the latter one requires a plug-in estimator, as does the MAC estimator. As a consequence, the critical values are depending on d. McElroy and Politis (2012) offer response curves for various kernels.[4]

5.5 Monte Carlo Study

In this section we analyze the finite-sample performance of the procedures discussed above by means of a simulation study. As in our motivating example, we conduct all size and power simulations for the t_{MAC}- and t_{EFB}-tests with $M = 5000$ Monte Carlo repetitions and the nominal significance level is set to 5%. For both tests, the plug-in estimation of d is done via local Whittle (LW) with $m_d = \lfloor T^{0.65} \rfloor$ which is similar to the simulation setup in Abadir et al. (2009). In the case of the extended fixed-b approach, we consider the Bartlett and the Modified Quadratic Spectral (MQS) kernel as used in Politis and McElroy (2009) and McElroy and Politis (2012).[5] Note that even though the theoretical results in Section 5.3 are based on assumptions on the forecasts and the forecast objective, the modified DM tests proposed in Section 5.4 are based solely on assumptions on the time series properties of the loss differentials. Since these tests are the subject of this Monte Carlo study, we also take this perspective for the simulation design and generate the loss differential series z_t directly from standard time series models.

The results reported below are generated for the following two DGPs. DGP1 is a fractional Gaussian white noise process with memory parameter $d = \{0, 0.05, 0.1, ..., 0.45\}$, while DGP2 contains an additional first-order autoregressive component with parameter $\phi = 0.6$.

If the loss differential series has zero mean, this represents a situation where both forecasts are equally good. For non-zero means one of the forecasts outperforms the other. Since the DM test is essentially a test on the mean, the results presented below can not only be interpreted with regard to forecast comparisons. Instead, they can also be considered as a general comparison of size and power between statistics using the MAC estimator and tests employing the extended fixed-b asymptotics. To the best of our knowledge, such a comparison has not been conducted in the existing literature before.

[4] All common kernels (e.g. Bartlett, Parzen) as well as others considered in Kiefer and Vogelsang (2005) can be used. In addition to the aforementioned, McElroy and Politis (2012) use the Daniell, the Trapezoid, the Modified Quadratic Spectral, the Tukey-Hanning and the Bohman kernel.

[5] The MQS kernel is a modified version of the usual QS kernel used in Kiefer and Vogelsang (2005), but restricted to $x \in [-1, 1]$. The kernel is given by $k(x) = 3(\sin(\pi x)/(\pi x) - \cos(\pi x))/(\pi x)^2$ for $x \in [-1, 1]$ and $k(x) = 0$ for $|x| > 1$, where $x = j/B$, if the kernel is employed for the long-run variance estimation as in (5.3). Further kernels, including flat-top tapers, are analyzed as well but yield slightly inferior results to those reported here.

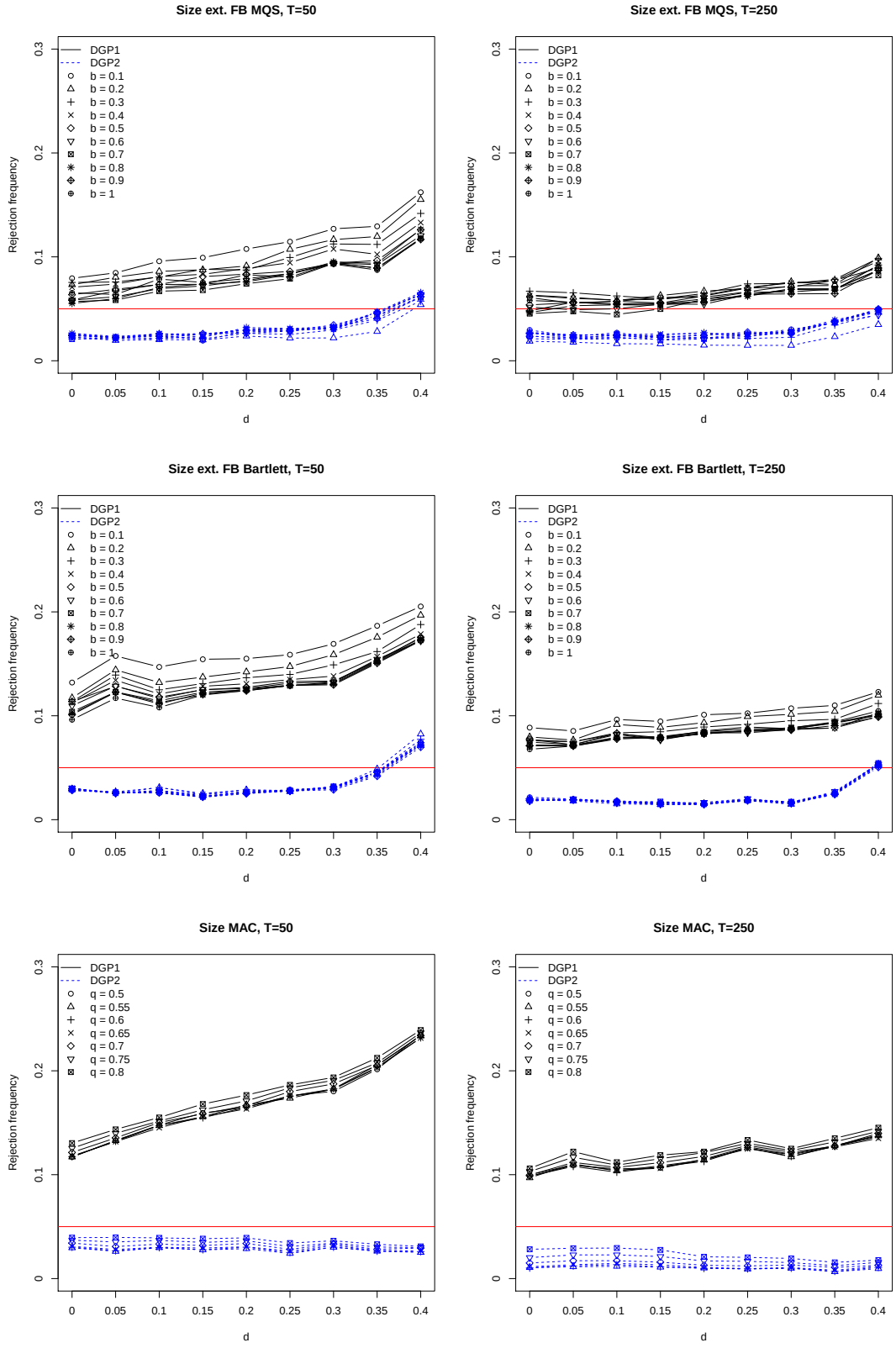

Figure 5.2: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes T and bandwidth parameters q and b.

In regard of the fact that optimal forecasts are MA processes the attentive reader might wonder why the results presented do not include MA dynamics. However, the derivative of the spectral density of MA processes in the vicinity of the zero frequency tends to be much smaller than that of AR processes, so that the spectral density at the origin is more flat and has a less severe effect on the finite-sample performance of the estimators for the long memory parameters. We therefore decide to present the results for the situation that is more challenging for the methods employed, but additional results under MA dynamics are available from the authors upon request. In addition to that, the important special case of optimal one-step-ahead forecasts is represented by DGP1 for $d = 0$.

Figure 5.2 shows a comparison of the size of both tests for different degrees of long memory and sample sizes of $T \in \{50, 250\}$. In the case of DGP1, all tests are liberal and the size tends to increase with increasing d. The t_{EFB}-statistic obtained with the MQS kernel gives the best size control, whereas the t_{MAC}-statistic shows the highest rejection frequencies.

In larger samples of $T = 250$ observations the dependence of the size on d is reduced and both tests approach their nominal significance level. However, for both sample sizes the t_{EFB}-statistics are notably closer to their nominal level of 5% than the MAC-based statistic and among the t_{EFB}-statistics, the one obtained with the MQS kernel performs best. As observed by Kiefer and Vogelsang (2005), there is a trade-off in terms of size and power in the choice of the bandwidth parameter b. Larger bandwidths generally improve the size and reduce the power. However, as can be seen from the results below, the kernel choice has a more severe effect than the bandwidth choice. Especially in larger samples the size is nearly identical for all bandwidths.

DGP2 contains short memory influences and the results shown here are obtained with an autoregressive coefficient of $\phi = 0.6$. Interestingly, in the presence of short memory components, the results change notably. Already in small samples of $T = 50$ observations both tests are conservative. For large values of d, we can observe that the t_{EFB}-test becomes slightly liberal again. Several further simulations are conducted, considering a wide range of different autoregressive and moving average components. Qualitatively, this does not alter the findings, even though the conservativeness of the procedure becomes stronger with increasing ϕ and moving average components tend to have a less severe impact compared to autoregressive components for the reasons discussed above.

Additional simulation studies using the true - but in practical applications unknown - memory parameter d, reveal that the tests are no longer conservative. Consequently, the effect can be attributed to the finite-sample bias of the local Whittle estimator that occurs if short memory dynamics are present. This is in line with the results of McElroy and Politis (2012), who also note that their results are strongly influenced by the performance of the estimation procedure for the memory parameter d.

Our simulation study suggests that a fixed-bandwidth choice of $b = 0.8$ provides a good balance in the size-power trade-off under both DGPs, for both kernels, and for all considered memory parameters. Concerning the MAC estimator, the MSE-optimal choice $m_{opt} = \lfloor T^{0.8} \rfloor$ derived in Abadir et al. (2009) indeed provides the best results under DGP2. In this situation, it gives a

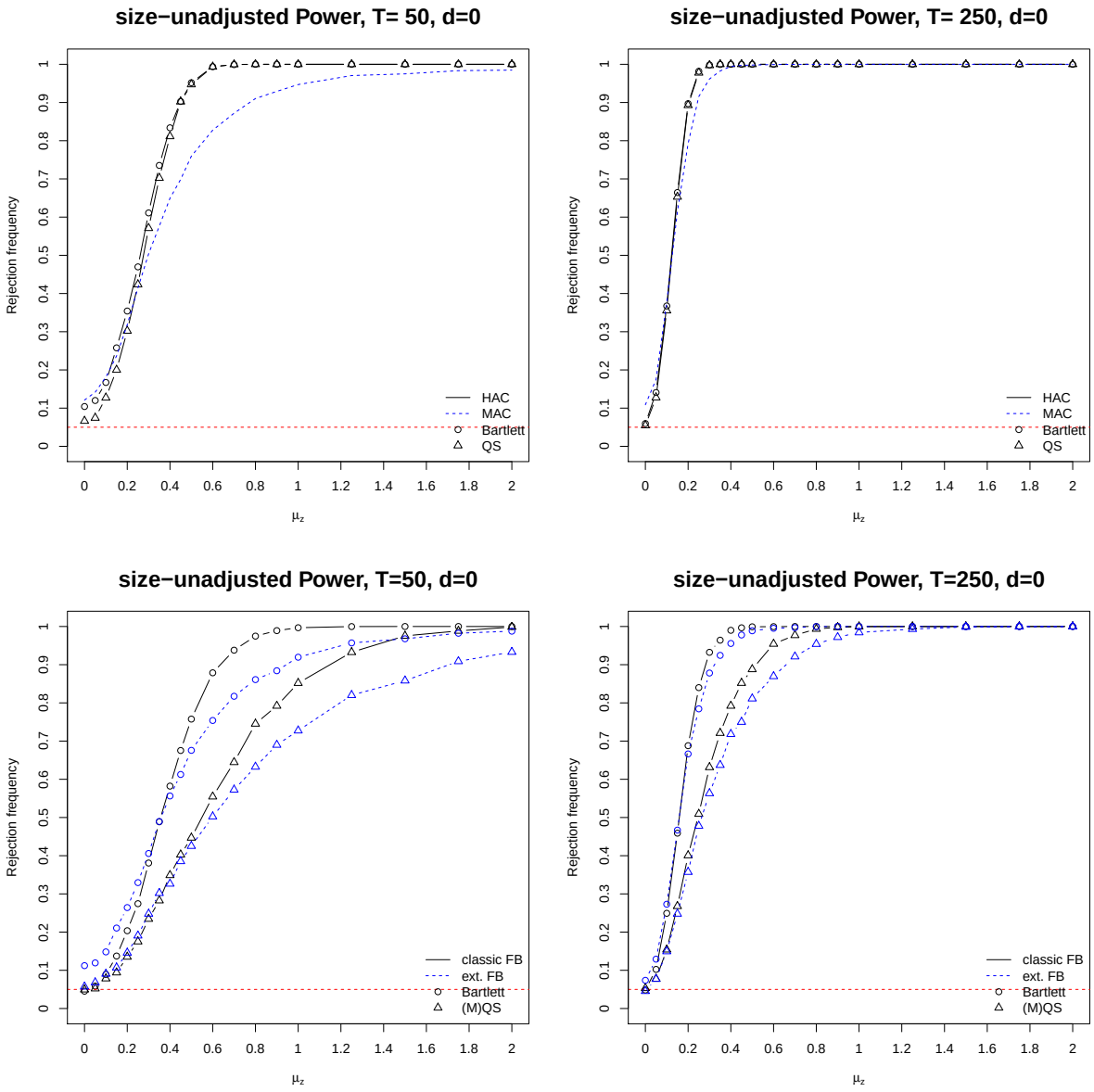

Figure 5.3: Power comparison of the robust statistics t_{EFB} and t_{MAC} with their short memory counterparts when $d = 0$.

size close to the nominal level and is better than that of the t_{EFB}-test. However, in the case of DGP1 the bandwidth $m = \lfloor T^q \rfloor$ with $q = 0.7$ seems more adequate which can also observed in the simulation study of Abadir et al. (2009).

In a next step, we consider the potential losses in power arising from the use of the robust t_{EFB}- and t_{MAC}-statistics when the additional flexibility is not needed, because the series does not have long memory ($d = 0$). Results are presented in Figure 5.3. Since it is our objective to evaluate the potential loss in power if one would generally use memory robust tests in practice, we consider size-unadjusted power here. We compare the t_{HAC}- and t_{MAC}-tests in the top row and the t_{FB}- with the t_{EFB}-statistics in the bottom row of Figure 5.3 under $T \in \{50, 250\}$ and DGP1, setting

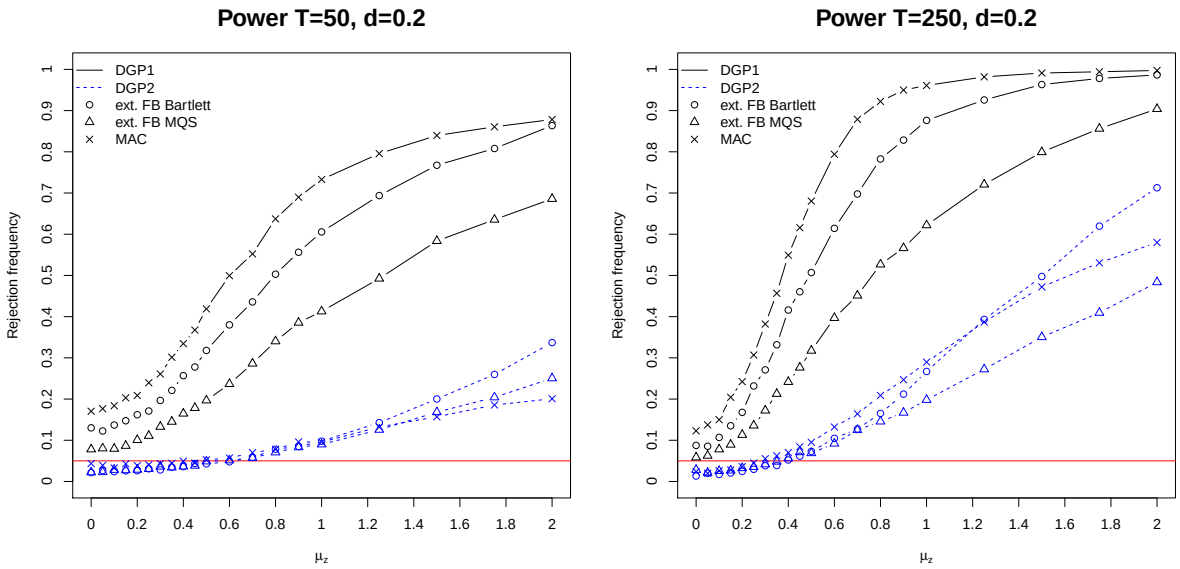

Figure 5.4: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample sizes $T \in \{50, 250\}$.

$d = 0$. Although as expected some power loss can be observed, the cost of using the long memory robust procedures is more than acceptable, especially for $T = 250$.

Finally, we analyze the power of the t_{MAC}- and t_{EFB}-statistics under both DGPs, for the case of $d = 0.2$. This setup matches the memory orders in our empirical application (cf. Section 5.6) closely. With regard to the previous results, we chose the bandwidth parameter of $b = 0.8$ for the extended fixed-b and $m = \lfloor T^{0.7} \rfloor$ ($m = \lfloor T^{0.8} \rfloor$) for the MAC approach under DGP1 (DGP2). To control for the increase in the variance of the process (which depends on the memory parameter d), each loss differential series is standardized before the mean is added and the respective test is applied. The results are shown in Figure 5.4. As expected, the power increases with the sample size.

With regard to the ranking, we observe that in case of DGP1 the t_{MAC}-statistic clearly outperforms the t_{EFB}-statistics among which the one obtained using a Bartlett kernel performs best. The t_{EFB}-statistic obtained with the MQS kernel, on the other hand, has the lowest power under both DGPs. Since the t_{MAC}-statistic is clearly more liberal than its two competitors, we also provide size-adjusted power curves in Figure 5.9 in the Appendix. Due to the different employment of the d estimator in both methods, such a comparison is only valid for known d. For both DGPs, it can be clearly seen that the power advantages of the t_{MAC}-statistic go beyond the effect of the upward size distortion. Interestingly, further simulations have shown that these power advantages tend to increase with increasing d.

By comparing the results for DGP1 with those of DGP2 in Figure 5.4, one can observe that the power of both tests suffers if short memory components are present. Different from the size effect of the short memory dynamics discussed above, simulations with known d show that this cannot be explained by the effect of autoregressive dynamics on the estimation of d alone. Instead, the

presence of short memory dynamics increases the finite-sample variance of the estimated means - similar to the effect of an increase in d.

To robustify the procedures against the effect of short memory dynamics discussed above, one could consider to apply the adaptive local polynomial Whittle (ALPW) estimator of Andrews and Sun (2004). Figures 5.8 and 5.10 in the Appendix shows the results of this exercise. In smaller samples, the size obtained using the ALPW estimator now becomes similarly liberal for all procedures and both DGPs. In larger samples of $T = 250$, all tests reach a satisfactory size, however, the size of the t_{EFB}-statistic using the MQS kernel remains the best and the t_{MAC}-statistic performs better if a smaller bandwidth, say $m = \lfloor T^{0.55} \rfloor$ is used. The power, on the other hand, is remarkably reduced and the t_{EFB}-statistic using the Bartlett kernel now has the highest power.

We find that the t_{EFB}-tests generally provide better size control than the t_{MAC}-test, whereas the latter has better power properties. Among the extended fixed-b procedures, the MQS kernel has better size but less power compared to the Bartlett kernel. In presence of short memory dynamics both procedures become quite conservative. This effect can be mitigated if the ALPW estimator is employed for the plug-in estimation of the memory parameter d. However, this comes at the cost of an additional loss in power.

Since there is no dominant procedure in terms of size control and power, we conclude that it is beneficial for forecast comparisons in practice to consider both statistics and to compare the outcomes. This also applies to other inference problems involving the sample mean.

5.6 An Application to Realized Volatility Forecasting

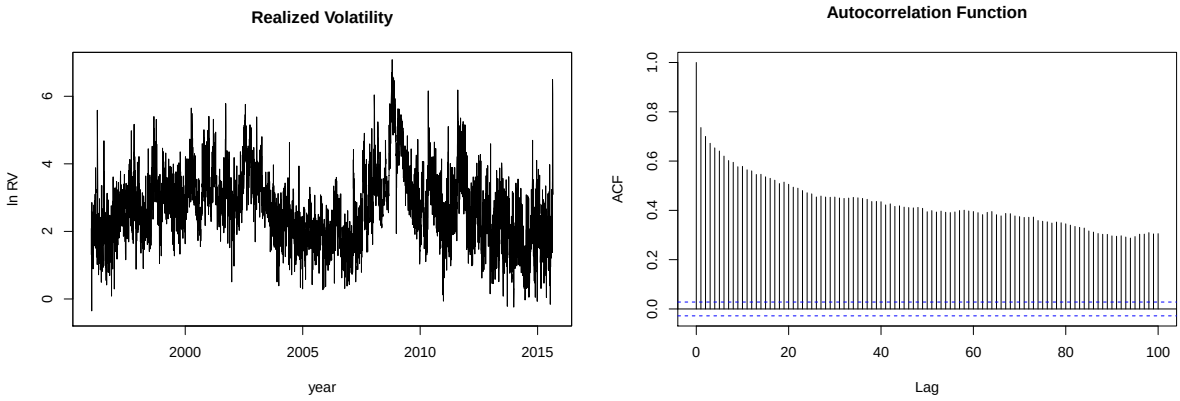

Figure 5.5: Daily log-realized volatility of the S&P500 index and their autocorrelation function.

Due to its relevance for risk management and derivative pricing, volatility forecasting is of vital importance and is also one of the fields in which long memory models are applied most often (cf., e.g., Deo et al. (2006), Martens et al. (2009) and Chiriac and Voev (2011)). Since intraday data on financial transactions has become widely available, the focus has shifted from GARCH-

type models to the direct modelling of realized volatility series. In particular the heterogeneous autoregressive model (HAR-RV) of Corsi (2009) and its extensions have emerged as one of the most popular approaches.

As an empirical application we therefore re-evaluate some recent results from the related literature using traditional Diebold-Mariano tests as well as the long memory robust versions from Section 5.4. We use a data set of 5-minute log-returns of the S&P 500 Index from January 2, 1996 to August 31, 2015 and we include close-to-open returns. The raw data is obtained from the Thomson Reuters Tick History Database.

Before we turn to the forecast evaluations in Sections 5.6.1 and 5.6.2, we use the remainder of this section to define the relevant volatility variables and to introduce the data and the employed time series models. Define the j-th intraday return on day t by $r_{t,j}$ and let there be N intraday returns per day, than following Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) the daily realized variance is defined as

$$RV_t = \sum_{j=1}^{N} r_{t,j}^2 \,.$$

If $r_{t,j}$ is sampled with an ever-increasing frequency such that $N \to \infty$, RV_t provides a consistent estimate of the quadratic variation of the log-price process. Therefore, RV_t is usually treated as a direct observation of the stochastic volatility process. The HAR-RV model of Corsi (2009), for example, explains log-realized variance by an autoregression involving overlapping averages of past realized variances.

$$\ln RV_t^{(h)} = \alpha + \rho_{22} \ln RV_{t-h}^{(22)} + \rho_5 \ln RV_{t-h}^{(5)} + \rho_1 \ln RV_{t-h}^{(1)} + \varepsilon_t \,, \tag{5.8}$$

where

$$RV_t^{(M)} = \frac{22}{M} \sum_{j=0}^{M-1} RV_{t-j} \,,$$

and ε_t is a white noise process. Although this is formally not a long memory model, this simple process provides a good approximation to the slowly decaying autocorrelation functions of long memory processes in finite samples. Forecast comparisons show that the HAR-RV model performs similar to ARFIMA models (cf. Corsi (2009)).

Motivated by developments in derivative pricing that highlighted the importance of jumps in price processes, Andersen et al. (2007) extend the HAR-RV model to consider jump components in realized volatility. Here, the underlying model for the continuous time log-price process $p(t)$ is given by

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t) \,,$$

where $0 \leq t \leq T$, $\mu(t)$ has locally bounded variation, $\sigma(t)$ is a strictly positive stochastic volatility process that is càdlàg and $W(t)$ is a standard Brownian motion. The counting process $q(t)$ takes the value $dq(t) = 1$, if a jump is realized and it is allowed to have time varying intensity. Finally, the process $\kappa(t)$ determines the size of discrete jumps, if these are realized. Therefore, the quadratic variation of the cumulative return process can be decomposed into integrated volatility plus the sum of squared jumps:

$$[r]_t^{t+h} = \int_t^{t+h} \sigma^2(s)ds + \sum_{t<s\leq t+h} \kappa^2(s) \, .$$

In order to measure the integrated volatility component, Barndorff-Nielsen and Shephard (2004, 2006) introduce the concept of bipower variation (BPV) as an alternative estimator that is robust to the presence of jumps. Here, we use threshold bipower variation (TBPV) as suggested by Corsi et al. (2010), who showed that BPV can be severely biased in finite samples. TBPV is defined as follows:

$$TBPV_t = \frac{\pi}{2} \sum_{j=2}^{N} |r_{t,j}||r_{t,j-1}| \mathbb{I}(|r_{t,j}|^2 \leq \zeta_j) \mathbb{I}(|r_{t,j-1}|^2 \leq \zeta_{j-1}) \, ,$$

where ζ_j is a strictly positive, random threshold function as specified in Corsi et al. (2010) and $\mathbb{I}(\cdot)$ is an indicator function.[6] Since

$$TBPV_t \xrightarrow{p} \int_t^{t+1} \sigma^2(s)ds$$

for $N \to \infty$, one can decompose the realized volatility into the continuous integrated volatility component C_t and the jump component J_t as

$$\begin{aligned} J_t &= \max\{RV_t - TBPV_t, 0\} \mathbb{I}(\text{C-Tz} > 3.09) \, , \\ C_t &= RV_t - J_t \, . \end{aligned}$$

The argument of the indicator function $\mathbb{I}(\text{C-Tz} > 3.09)$ ensures that the jump component is set to zero if it is insignificant at the nominal 0.1% level, so that J_t is not contaminated by measurement error, see also Corsi and Renò (2012). For details on the C-Tz statistic, see Corsi et al. (2010). Different from previous studies that find an insignificant or negative impact of jumps, Corsi et al. (2010) show that the impact of jumps on future realized volatility is significant and positive. Here, we use the HAR-RV-TCJ model that is studied in Bekaert and Hoerova (2014):

$$\begin{aligned} \ln RV_t^{(h)} = \alpha &+ \rho_{22} \ln C_{t-h}^{(22)} + \rho_5 \ln C_{t-h}^{(5)} + \rho_1 \ln C_{t-h}^{(1)} \\ &+ \varpi_{22} \ln\left(1 + J_{t-h}^{(22)}\right) + \varpi_5 \ln\left(1 + J_{t-h}^{(5)}\right) + \varpi_1 \ln\left(1 + J_{t-h}^{(1)}\right) + \varepsilon_t \, . \end{aligned} \tag{5.9}$$

[6] To calculate ζ_j, we closely follow Corsi et al. (2010).

q	$\widehat{d}_{LW}$	$\widehat{d}_{HP}$	s.e.	W	$\widehat{d}_{(0,0)}$		$\widehat{d}_{(1,0)}$		$\widehat{d}_{(1,1)}$	
0.55	0.554	0.493	(0.048)	0.438	0.613	(0.088)	0.612	(0.132)	0.689	(0.163)
0.60	0.553	0.522	(0.039)	0.568	0.567	(0.074)	0.577	(0.110)	0.692	(0.131)
0.65	0.573	0.573	(0.032)	0.544	0.573	(0.059)	0.570	(0.089)	0.570	(0.118)
0.70	0.549	0.532	(0.026)	0.449	0.573	(0.048)	0.578	(0.072)	0.588	(0.093)
0.75	0.539	0.518	(0.021)	0.515	0.564	(0.039)	0.574	(0.058)	0.593	(0.075)

Table 5.1: Long memory estimation and testing results for S&P 500 log-realized volatility. Local Whittle estimates for the d parameter and results of the Qu (2011) test (W statistic) for true versus spurious long memory are reported for various bandwidth choices $m_d = \lfloor T^q \rfloor$. Critical values are 1.118, 1.252 and 1.517 at the nominal significance level of 10%, 5% and 1%, respectively. Asymptotic standard errors for $\widehat{d}_{LW}$ and $\widehat{d}_{HP}$ are given in parentheses. The indices of the LPWN estimators indicate the orders of the polynomials used. For details, see Frederiksen et al. (2012).

The daily log-realized variance series ($\ln RV_t$) is depicted in Figure 5.5.[7] It is common to use log-realized variance to avoid non-negativity constraints on the parameters and to have a better approximation to the normal distribution, as advocated by Andersen et al. (2001). As can be seen from Figure 5.5, the series shows the typical features of a long memory time series, namely a hyperbolically decaying autocorrelation function, as well as local trends.

Estimates of the memory parameter are shown in Table 5.1. Local Whittle estimates ($\widehat{d}_{LW}$) exceed 0.5 slightly and thus indicate non-stationarity. Since there is a large literature on the potential of spurious long memory in volatility time series, we carry out the test of Qu (2011). To avoid issues due to non-stationarity and to increase the power of the test, we follow Kruse (2015) and apply the test to the fractional difference of the data. The necessary degree of differencing is determined using the estimator by Hou and Perron (2014) ($\widehat{d}_{HP}$) that is robust to low-frequency contaminations. As one can see, the memory estimates are fairly stable and the Qu test fails to reject the null hypothesis of true long memory.

Since N is finite in practice, RV_t might contain a measurement error and is therefore often modeled as the sum of the quadratic variation and an *iid* perturbation process such that $RV_t = [r]_t^{t+1} + u_t$, where $u_t \sim iid(0, \sigma_u^2)$. Furthermore, it is well known that local Whittle estimates can be biased in presence of short run dynamics. We therefore also report results of the local polynomial Whittle plus noise (LPWN) estimator of Frederiksen et al. (2012). Similar to the ALPW estimator of Andrews and Sun (2004), the LPWN estimator reduces the bias due to short memory dynamics by approximating the log-spectral density of the short memory component with a polynomial, but it additionally includes a second polynomial to account for the downward bias induced by perturbations. As one can see, the estimates remain remarkably stable - irrespective of the choice of the estimator. The downward bias of the local Whittle

[7] For a better comparison, all variables in this section are scaled towards a monthly basis.

estimator due to the measurement error in realized variance is therefore moderate.
Altogether, the realized variance series appears to be a long memory process. Consequently, if forecasts of the series are evaluated, a transmission of long range dependence to the loss differentials as implied by Propositions 5.2, 5.3 and 5.4 can occur.

5.6.1 Predictive Ability of the VIX for Quadratic Variation

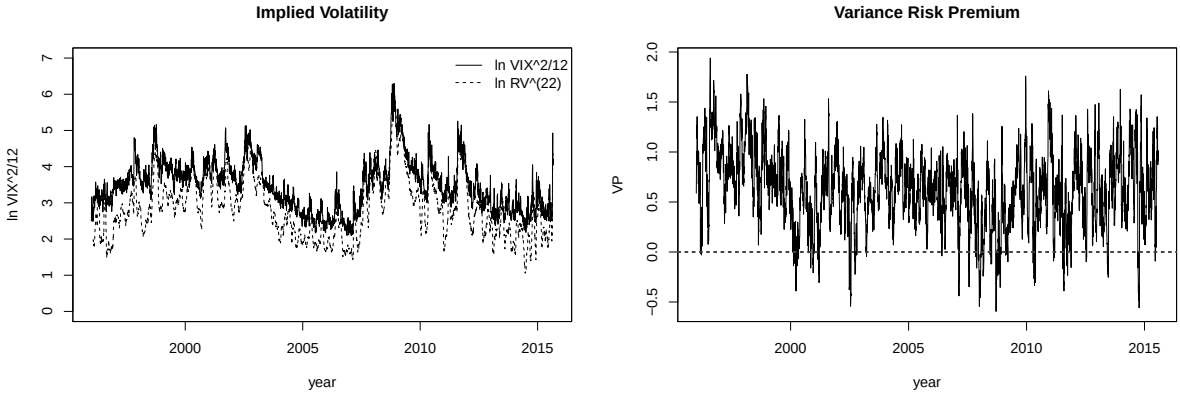

Figure 5.6: Log squared implied volatility and log cumulative realized volatility of the S&P 500 (left panel) and variance risk premium $VP_t = \ln(VIX_t^2/12) - \ln RV_{t+22}^{(22)}$ (right panel).

The predictive ability of implied volatility for future realized volatility is an issue that has received a lot of attention in the related literature. The CBOE VIX represents the market expectation of quadratic variation of the S&P 500 over the next month, derived under the assumption of risk neutral pricing. Both, $\ln(VIX_t^2/12)$ and $\ln RV_{t+22}^{(22)}$ are depicted in Figure 5.6. As one can see, both series behave fairly similar and are quite persistent. As for the log-realized volatility series, the Qu (2011) test does not reject the null hypothesis of true long memory for the VIX after appropriate fractional differencing following Kruse (2015).
Chernov (2007) investigates the role of a variance risk premium in the market for volatility forecasting. The variance risk premium is given by $VP_t = \ln(VIX_t^2/12) - \ln RV_{t+22}^{(22)}$ and displayed on the right hand side of Figure 5.6. The graph clearly suggests that the VIX tends to overestimate the realized variance and the sample average of the variance risk premium is 0.623. Furthermore, the linear combination of realized and implied volatility is rather persistent and has a significant memory of $\widehat{d}_{LPWN} = 0.2$. This is consistent with the existence of a fractional cointegration relationship between $\ln(VIX_t^2/12)$ and $\ln RV_{t+22}^{(22)}$ which has been considered in several contributions including Christensen and Nielsen (2006), Nielsen (2007b) and Bollerslev et al. (2013). Bollerslev et al. (2009), Bekaert and Hoerova (2014) and Bollerslev et al. (2013) additionally extend the analysis towards the predictive ability of VP_t for stock returns.
While the aforementioned articles test the predictive ability of the VIX itself and the "implied-realized-parity", there has also been a series of studies that analyze whether the inclusion of implied volatility can improve model-based forecasts. On the one hand, Becker et al. (2007)

Models		Summary statistics				Short memory inference			Long memory inference						
									t_{MAC}			t_{EFB}			
Model vs. Model+VIX	$\bar{z}/\widehat{\sigma}_z$	$MSE1$	$MSE2$	$\widehat{d}_{LW}$	$\widehat{d}_{LPWN}$	t_{DM}	t_{HAC}	t_{FB}	0.7	0.75	0.8	0.2	0.4	0.6	0.8
HAR-RV	0.135	0.292	0.269	**0.219***	**0.234***	**2.968**	**3.032**	**2.494**	0.929	1.038	1.188	2.494 (3.404)	2.754 (4.064)	2.985 (4.750)	2.849 (5.388)
HAR-RV-TCJ	0.109	0.285	0.268	**0.175***	0.138	**2.421**	**2.455**	**2.097**	1.397	1.610	**1.892**	2.097 (2.610)	2.503 (3.154)	2.889 (3.693)	2.724 (4.228)
HAR-RV-TCJ-L	0.082	0.282	0.269	**0.182***	0.163	**1.784** (1.645)	**1.786** (1.645)	1.819 (2.092)	0.889	1.016 (1.645)	1.192	1.819 (3.404)	2.153 (4.064)	2.430 (4.750)	2.317 (5.388)

Table 5.2: Predictive ability of the VIX for future RV. Models excluding the VIX are tested against models including the VIX. Reported are the standardized mean ($\bar{z}/\widehat{\sigma}_z$) and estimated memory parameter ($\widehat{d}$) of the forecast error loss differential. Furthermore, the respective out-of-sample MSEs of the models and the results of various DM test statistics. Bold-faced values indicate significance at the nominal 5% level; an additional star indicates significance at the nominal 1% level. Critical values of the tests are given in parentheses.

conclude that the VIX does not contain any incremental information on future volatility relative to an array of forecasting models. On the other hand, Becker et al. (2009) show that the VIX is found to subsume information on past jump activity and contains incremental information on future jumps if continuous components and jump components are considered separately. Similarly, Busch et al. (2011) study a HAR-RV model with continuous components and jumps and propose a VecHAR-RV model. They find that the VIX has incremental information and partially predicts jumps.

Motivated by these findings, we test whether the inclusion of $\ln(VIX_t^2/12)$ improves model-based forecasts from HAR-RV-type models, using Diebold-Mariano statistics. Since the VIX can be seen as a forecast of future quadratic variation over the next month, we consider a 22-step forecast horizon. Consecutive observations of multi-step forecasts of stock variables, such as integrated realized volatility, can be expected to exhibit relatively persistent short memory dynamics. The empirical autocorrelations of these loss differentials reveal an MA structure with linearly decaying coefficients. We therefore base all our robust statistics on the local polynomial Whittle plus noise (LPWN) estimator of Frederiksen et al. (2012) discussed above.[8] Since Chen and Ghysels (2011) and Corsi and Renò (2012) show that leverage effects improve forecasts, we also include a comparison of the HAR-RV-TCJ-L model and the HAR-RV-TCJ-L-VIX model. For details on the HAR-RV-TCJ-L model, see Corsi and Renò (2012) and equation (2) in Bekaert and Hoerova (2014).

Table 5.2 reports the results. Models are estimated using a rolling window of $T_w = 1000$ obser-

[8] We choose $R_y = 1$ and $R_w = 0$ concerning the polynomial degrees and a bandwidth $m_d = \lfloor T^{0.8} \rfloor$ (see Frederiksen et al. (2012) for details on the estimator).

vations.[9] All DM tests are conducted with one-sided alternatives. We test that a more complex model outperforms its parsimonious version. For the sake of a better comparability, all kernel-based tests use the Bartlett kernel. In accordance with the previous literature, the t_{DM}-statistic is implemented using an MA approximation with 44 lags for the forecast horizon of 22 days, c.f. for instance Bekaert and Hoerova (2014). For the t_{HAC}-statistic we use an automatic bandwidth selection procedure and the t_{FB}-statistic is computed by using $b = 0.2$ which offers a good trade-off between size control and power, as confirmed in the simulation studies of Sun et al. (2008).

Table 5.2 reveals that the forecast error loss differentials have long memory with d parameters between 0.138 and 0.234. The results are very similar for the local Whittle and the LPWN estimator. Standard DM statistics (t_{DM}, t_{HAC} and t_{FB}) reject the null hypothesis of equal predictive ability, thereby confirming the findings in the previous literature.

However, if the memory robust statistics in the right panel of Table 5.2 are taken into account, all evidence for a superior predictive ability of models including the VIX vanishes. Therefore, the previous rejections might be spurious and reflect the theoretical findings in Proposition 5.5. In regard of the persistence in the loss differential series the improvements are too small to be considered significant. These findings highlight the importance of long memory robust tests for forecast comparisons in practice.

5.6.2 Separation of Continuous Components and Jump Components

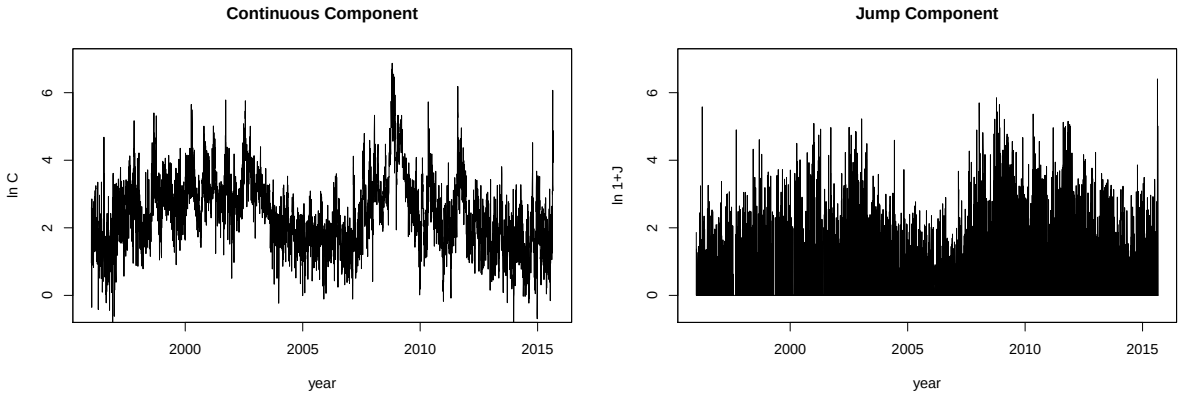

Figure 5.7: Log continuous component $\ln C_t$ and jump component $\ln(1 + J_t)$ of RV_t.

As a second empirical application, we revisit the question whether the HAR-RV-TCJ model from equation (5.9) leads to a significant improvement in forecast performance compared to the standard HAR-RV-model (5.8) from a purely out-of-sample perspective.

The continuous components and jump components - separated using the approach described above - are shown in Figure 5.7. The occurrence of jumps is often associated with macroeconomic

[9] As a robustness check, we repeat the analysis for a larger window of 2500 observations and obtain qualitatively similar results.

events (cf. Barndorff-Nielsen and Shephard (2006) and Andersen et al. (2007)) and they are observed relatively frequently at about 40% of the days in the sample. The trajectory of the log-continuous component closely follows that of the log-realized volatility series.

Models	Summary statistics					Short memory inference			Long memory inference						
									t_{MAC}			t_{EFB}			
HAR-RV vs.	$\bar{z}/\widehat{\sigma}_z$	MSE_1	MSE_2	$\widehat{d}_{LW}$	$\widehat{d}_{LPWN}$	t_{DM}	t_{HAC}	t_{FB}	0.7	0.75	0.8	0.2	0.4	0.6	0.8
HAR-RV-TCJ, $h=1$	0.122	0.409	0.375	**0.094***	0.127	**6.932**	**7.631**	**3.995**	**3.243**	**3.144**	**3.091**	**3.995** (2.610)	**4.068** (3.154)	**4.468** (3.693)	**4.947** (4.228)
HAR-RV-TCJ, $h=5$	0.092	0.263	0.247	**0.072**	0.009	**3.666**	**3.790**	**2.789**	**3.620**	**3.853**	**4.277**	**2.789** (2.050)	**3.981** (2.522)	**5.093** (2.975)	**5.848** (3.386)
HAR-RV-TCJ, $h=22$	0.045	0.292	0.285	**0.359***	**0.343***	0.776 (1.645)	0.912 (1.645)	0.666 (2.092)	0.140	0.152 (1.645)	0.171	0.666 (4.701)	0.925 (5.551)	1.064 (6.413)	1.164 (7.281)

Table 5.3: Separation of Continuous and Jump Components. Reported are the standardized mean ($\bar{z}/\widehat{\sigma}_z$) and estimated memory parameter ($\widehat{d}$) of the forecast error loss differential. Furthermore, the respective out-of-sample MSEs of the models and the results of various DM test statistics. Bold-faced values indicate significance at the 5% level and an additional star indicates significance at the 1% level. Critical values of the tests are given in parentheses.

Table 5.3 shows the results of our forecasting exercise for $h \in \{1,5,22\}$ steps. Similar to the previous analysis, the t_{DM}-statistic is implemented using an MA approximation including 5, 10 or 44 lags for forecast horizons $h = 1,5$ and 22, respectively, as is customary in this literature. All other specifications are the same as before. As one can see, the standard tests (t_{DM}, t_{HAC} and t_{FB}) agree upon rejection of the null hypothesis of equal predictive ability in favour of a better performance of the HAR-RV-TCJ model for $h = 1$ and $h = 5$, but not for $h = 22$.

If we consider estimates of the memory parameter, strong (stationary) long memory of 0.34 is only found for $h = 22$. For smaller forecast horizons of $h = 1$ and $h = 5$, LPWN estimates are no longer significantly different from zero, since the asymptotic variance is inflated by a multiplicative constant which is also larger for smaller values of d. However, local Whittle estimates remain significant at $\widehat{d}_{LW} = 0.094$ and $\widehat{d}_{LW} = 0.070$ which is qualitatively similar to the results obtained using the LPWN estimator. Therefore, the rejections of equal predictive accuracy obtained using standard tests might be spurious due to the neglected effect of long range dependence. Nevertheless, the improvement in forecast accuracy is large enough, so that the long memory robust t_{MAC}- and t_{EFB}-statistics reject across the board for $h = 1$ and $h = 5$. We can therefore confirm that the separation of continuous and jump components indeed improves the forecast performance on daily to weekly horizons.

5.7 Conclusion

This paper deals with forecast evaluation under long range dependence. We show in Section 5.3 that long memory can be transmitted from the forecasts $\widehat{y}_{it}$ and the forecast objective y_t to the forecast error loss differential series z_t. We demonstrate that the popular test of Diebold and Mariano (1995) is invalidated in these cases. Rejections of the null hypothesis of equal predictive accuracy might therefore be spurious if the series of interest has long memory.

Two methods to robustify DM tests against long memory are discussed in Section 5.4 - the MAC estimator of Robinson (2005) and Abadir et al. (2009), as well as the extended fixed-b approach of McElroy and Politis (2012).

The finite sample performance of both of these methods is studied using Monte Carlo simulations. While the extended fixed-b approach allows a better size control, the MAC performs better in terms of power. With regard to kernel and bandwidth choices for the t_{EFB}-statistic, we find that $b = 0.8$ gives good results and that the kernel choice has a larger impact on the size and power of the procedure than the bandwidth selection. In general the MQS kernel gives a better size control, whereas the Bartlett kernel is superior in terms of power. An important issue remains the impact of short memory dynamics on the plug-in estimation of the memory parameter. However, our results using the ALPW estimator of Andrews and Sun (2004) indicate that bias-corrected local Whittle estimators successfully improve the results - at least in larger samples. As to be expected, this comes at the price of a power loss.

An important example of long memory time series is the realized variance of the S&P 500. It has been the subject of various forecasting exercises. We therefore consider this series in our empirical application. In contrast to previous studies, we do not find statistical evidence for the hypothesis that the inclusion of the VIX index in HAR-RV-type models leads to an improved forecast performance. Taking the memory of the loss differentials into account reverses the test decisions and suggests that the corresponding findings might be spurious. With regard to the separation of continuous components and jump components, as suggested by Andersen et al. (2007), on the other hand, the improvements in forecast accuracy remain significant. These examples stress the importance of long memory robust statistics in practice.

Appendix

Proofs

Proof (Proposition 5.2). *By defining* $a_t^* = a_t - \mu_a$, *for* $a_t \in \{y_t, \widehat{y}_{1t}, \widehat{y}_{2t}\}$, *the loss differential* z_t *in (5.6) can be re-expressed as*

$$\begin{aligned}
z_t &= -2y_t(\widehat{y}_{1t} - \widehat{y}_{2t}) + \widehat{y}_{1t}^2 - \widehat{y}_{2t}^2 \\
&= -2(y_t^* + \mu_y)[\widehat{y}_{1t}^* + \mu_1 - \widehat{y}_{2t}^* - \mu_2] + (\widehat{y}_{1t}^* + \mu_1)^2 - (\widehat{y}_{2t}^* + \mu_2)^2 \\
&= -2\{y_t^*\widehat{y}_{1t}^* + \mu_1 y_t^* - y_t^*\widehat{y}_{2t}^* - y_t^*\mu_2 + \mu_y\widehat{y}_{1t}^* + \mu_y\mu_1 - \widehat{y}_{2t}^*\mu_y - \mu_2\mu_y\} \\
&\quad + \widehat{y}_{1t}^{*2} + 2\widehat{y}_{1t}^*\mu_1 + \mu_1^2 - \widehat{y}_{2t}^{*2} - 2\widehat{y}_{2t}^*\mu_2 - \mu_2^2 \\
&= \underbrace{-2[y_t^*(\mu_1 - \mu_2) + \widehat{y}_{1t}^*(\mu_y - \mu_1) - \widehat{y}_{2t}^*(\mu_y - \mu_2)]}_{I} \underbrace{-2[y_t^*(\widehat{y}_{1t}^* - \widehat{y}_{2t}^*)]}_{II} + \underbrace{\widehat{y}_{1t}^{*2} - \widehat{y}_{2t}^{*2}}_{III} + const.
\end{aligned} \tag{5.10}$$

Proposition 3 in Chambers (1998) states that the memory of a linear combination of fractionally integrated processes is equal to the maximum of the memory orders of the components. As discussed in Leschinski (2016), this result also applies for long memory processes in general, since the proof is only based on the long memory properties of the fractionally integrated processes. We can therefore also apply it to (5.10). In order to determine the memory of the forecast error loss differential z_t, *we have to determine the memory orders of the three individual components I, II and III in the linear combination.*
Regarding I, we have $y_t^* \sim LM(d_y)$, $\widehat{y}_{1t}^* \sim LM(d_1)$ *and* $\widehat{y}_{2t}^* \sim LM(d_2)$. *For terms II and III, we refer to Proposition 5.1 from Leschinski (2016). We thus have for* $i \in \{1,2\}$

$$y_t^*\widehat{y}_{it}^* \sim \begin{cases} LM(\max\{d_y + d_i - 1/2, 0\}), & \text{if } S_{y,\widehat{y}_i} \neq 0 \\ LM(d_y + d_i - 1/2), & \text{if } S_{y,\widehat{y}_i} = 0 \end{cases} \tag{5.11}$$

$$\text{and} \qquad \widehat{y}_{it}^{*2} \sim LM(\max\{2d_i - 1/2, 0\}). \tag{5.12}$$

Further note that

$$d_y > d_y + d_i - 1/2 \;\; \text{and} \;\; d_i > d_y + d_i - 1/2 \tag{5.13}$$

and

$$d_i > 2d_i - 1/2, \tag{5.14}$$

since $0 \leq d_a < 1/2$ *for* $a \in \{y, 1, 2\}$.
Using these properties, we can determine the memory d_z *in (5.10) via a case-by-case analysis.*

1. *First, if* $\mu_1 \neq \mu_2 \neq \mu_y$ *the memory of the original terms dominates because of (5.13) and (5.14) and we obtain* $d_z = \max\{d_y,\ d_1,\ d_2\}$.

2. *Second, if* $\mu_1 = \mu_2 \neq \mu_y$*, then* y_t^* *drops out from (5.10), but the two forecasts* $\widehat{y}_{1t}$ *and* $\widehat{y}_{2t}$ *remain. From (5.13) and (5.14), we have that* d_1 *and* d_2 *dominate their transformations leading to the result* $d_z = \max\{d_1,\ d_2\}$.

3. *Third, if* $\mu_1 = \mu_y \neq \mu_2$*, the forecast* $\widehat{y}_{1t}^*$ *vanishes and* d_2 *and* d_y *dominate their reduced counterparts by (5.13) and (5.14), so that* $d_z = \max\left\{2d_1 - 1/2,\ d_2,\ d_y\right\}$.

4. *Fourth, by the same arguments just as before,* $d_z = \max\left\{2d_2 - 1/2,\ d_1,\ d_y\right\}$ *if* $\mu_2 = \mu_y \neq \mu_1$.

5. *Finally, if* $\mu_1 = \mu_2 = \mu_y$*, the forecast objective* y_t^* *as well as both forecasts* $\widehat{y}_{1t}^*$ *and* $\widehat{y}_{2t}^*$ *drop from (5.10). The memory of the loss differential is therefore the maximum of the memory orders in the remaining four terms in II and III that are given in (5.11) and (5.12). Furthermore, the memory of the squared series given in (5.12) is always non-negative from Corollary 1 in Leschinski (2016) and a linear combination of an antipersistent process with an* $LM(0)$ *series is* $LM(0)$*, from Proposition 3 of Chambers (1998). Therefore, the lower bound for* d_z *is zero and*

$$d_z = \max\left\{2\max\{d_1,\ d_2\} - 1/2,\ d_y + \max\{d_1,\ d_2\} - 1/2,\ 0\right\}.$$

□

Proof (Proposition 5.3)**.** *For the case that common long memory is permitted, we consider three possible situations: CLM between the forecasts* $\widehat{y}_{1t}$ *and* $\widehat{y}_{2t}$*, CLM between the forecast objective* y_t *and one of the forecasts* $\widehat{y}_{1t}$ *or* $\widehat{y}_{2t}$ *and finally CLM between* y_t *and each* $\widehat{y}_{1t}$ *and* $\widehat{y}_{2t}$.

First, note that as a direct consequence of Assumption 5.3, we have

$$\mu_i = \beta_i + \xi_i \mu_x \tag{5.15}$$

and

$$\mu_y = \beta_y + \xi_y \mu_x. \tag{5.16}$$

We can now re-express the forecast error loss differential z_t *in (5.10) for each possible CLM relationship. In all cases, tedious algebraic steps are not reported to save space.*

1. *In the case of CLM between* $\widehat{y}_{1t}$ *and* $\widehat{y}_{2t}$*, we have*

$$\begin{aligned} z_t = &-2\{y_t^*(\mu_1 - \mu_2) + x_t^*[\xi_1(\mu_y - \mu_1) - \xi_2(\mu_y - \mu_2)] + x_t^* y_t^*(\xi_1 - \xi_2) - x_t^*(\xi_1 \varepsilon_{1t} - \xi_2 \varepsilon_{2t}) \\ &+ \varepsilon_{1t}(\mu_y - \mu_1) - \varepsilon_{2t}(\mu_y - \mu_2) + \mu_x(\varepsilon_{1t}\xi_1 - \varepsilon_{2t}\xi_2) + y_t^*(\varepsilon_{1t} - \varepsilon_{2t})\} \\ &+ x_t^{*^2}(\xi_1^2 - \xi_2^2) + \varepsilon_{1t}^2 - \varepsilon_{2t}^2 + 2\mu_x(\varepsilon_{1t}\xi_1 - \varepsilon_{2t}\xi_2) + const. \end{aligned} \tag{5.17}$$

2. *If the forecast objective y_t and one of the $\widehat{y}_{it}$ have CLM, we have for $\widehat{y}_{1t}$:*

$$\begin{aligned} z_t = &-2\{x_t^*[(\mu_y-\mu_1)\xi_1+\xi_y(\mu_1-\mu_2)]-\widehat{y}_{2t}^*[\mu_y-\mu_2]-\xi_y x_t^*\widehat{y}_{2t}^*+x_t^*[\varepsilon_{1t}(\xi_y-\xi_1)+\xi_1\eta_t] \\ &+\varepsilon_{1t}(\xi_y\mu_x-\mu_1)+\eta_t(\mu_1-\mu_2)+\varepsilon_{1t}\eta_t-\widehat{y}_{2t}^*\eta_t\} \\ &-(2\xi_1\xi_y-\xi_1^2)x_t^{*^2}+\varepsilon_{1t}^2-\widehat{y}_{2t}^{*^2}-2\beta_y\varepsilon_{1t}+const. \end{aligned} \tag{5.18}$$

The result for CLM between y_t and $\widehat{y}_{2t}$ is entirely analogous, but with index "1" being replaced by "2".

3. *Finally, if y_t has CLM with both $\widehat{y}_{1t}$ and $\widehat{y}_{2t}$, we have:*

$$\begin{aligned} z_t = &-2\{x_t^*[\xi_1(\mu_y-\mu_1)-\xi_2(\mu_y-\mu_2)+\xi_y(\mu_1-\mu_2)] \\ &+x_t^*[(\xi_y-\xi_1)\varepsilon_{1t}-(\xi_y-\xi_2)\varepsilon_{2t}+(\xi_1-\xi_2)\eta_t] \\ &+x_t^{*2}[\xi_y(\xi_1-\xi_2)-\frac{1}{2}(\xi_1^2-\xi_2^2)] \\ &+\varepsilon_{1t}(\mu_y-\mu_1)-\varepsilon_{2t}(\mu_y+\mu_2)+\mu_x(\xi_1\varepsilon_{1t}+\xi_2\varepsilon_{2t})+\eta_t(\varepsilon_{1t}-\varepsilon_{2t})+\eta_t[\mu_1-\mu_2]\} \\ &+\varepsilon_{1t}^2-\varepsilon_{2t}^2+2\mu_x(\xi_1\varepsilon_{1t}-\xi_2\varepsilon_{2t})+const. \end{aligned} \tag{5.19}$$

As in the proof of Proposition 5.2, we can now determine the memory orders of z_t in (5.17), (5.18) and (5.19) by first considering the memory of each term in each of the linear combinations and then by applying Proposition 3 of Chambers (1998) thereafter. Note, however, that

$$y_t^*(\mu_1-\mu_2)+x_t^*[\xi_1(\mu_y-\mu_1)-\xi_2(\mu_y-\mu_2)] \text{ in (5.17)},$$

$$x_t^*[(\mu_y-\mu_1)\xi_1+\xi_y(\mu_1-\mu_2)]-\widehat{y}_{2t}^*(\mu_y-\mu_2) \text{ in (5.18)}$$

and

$$x_t^*[\xi_1(\mu_y-\mu_1)-\xi_2(\mu_y-\mu_2)+\xi_y(\mu_1-\mu_2)] \text{ in (5.19)}$$

have the same structure as

$$y_t^*(\mu_1-\mu_2)+\widehat{y}_{1t}^*(\mu_y-\mu_1)-\widehat{y}_{2t}^*(\mu_y-\mu_2) \text{ in (5.10)}$$

and that all of the other non-constant terms in (5.17), (5.18) and (5.19) are either squares or products of demeaned series, so that their memory is reduced according to Proposition 5.1 from Leschinski (2016). From Assumption 5.3, x_t^* is the common factor driving the series with CLM and from $d_x > d_{\varepsilon_1}, d_{\varepsilon_2}, d_\eta$ and the dominance of the largest memory in a linear combination from Proposition 3 in Chambers (1998), x_t^* has the same memory as the series involved in the CLM relationship. Now from (5.13) and (5.14), the reduced memory of the product series and the squared series is dominated by that of either x_t^*, y_t^*, $\widehat{y}_{1t}^*$ or $\widehat{y}_{2t}^*$. Therefore, whenever a bias term is non-zero, the memory of the linear combination can be no smaller than that of the respective

original series. □

Proof (Proposition 5.4)**.** *First note that under the assumptions of Proposition 5.4, (5.17) is reduced to*

$$\begin{aligned} z_t &= -2\{-x_t^*(\xi_1\varepsilon_{1t}-\xi_2\varepsilon_{2t})+y_t^*(\varepsilon_{1t}-\varepsilon_{2t})\}+\varepsilon_{1t}^2-\varepsilon_{2t}^2+const, \\ &= -2\{-\underbrace{\xi_1 x_t^*\varepsilon_{1t}}_{I}+\underbrace{\xi_2 x_t^*\varepsilon_{2t}}_{II}+\underbrace{y_t^*\varepsilon_{1t}}_{III}-\underbrace{y_t^*\varepsilon_{2t}}_{IV}\}+\underbrace{\varepsilon_{1t}^2}_{V}-\underbrace{\varepsilon_{2t}^2}_{VI}+const, \end{aligned} \tag{5.20}$$

(5.18) becomes

$$\begin{aligned} z_t &= -2\{-x_t^*(\xi_y\widehat{y}_{2t}^*-\xi_1\eta_t)+(\varepsilon_{1t}-\widehat{y}_{2t}^*)\eta_t+\varepsilon_{1t}(\xi_y\mu_x-\mu_1)\}+\varepsilon_{1t}^2-\widehat{y}_{2t}^{*^2}-2\beta_y\varepsilon_{1t}-\xi_1\xi_y x_t^{*^2}+const, \\ &= -2\{-\underbrace{\xi_y x_t^*\widehat{y}_{2t}^*}_{I}+\underbrace{\xi_1 x_t^*\eta_t}_{II}+\underbrace{\varepsilon_{1t}\eta_t}_{III}-\underbrace{\widehat{y}_{2t}^*\eta_t}_{IV}+\underbrace{\varepsilon_{1t}(\xi_y\mu_x-\mu_1)}_{V}\}+\underbrace{\varepsilon_{1t}^2}_{VI}-\underbrace{\widehat{y}_{2t}^{*^2}}_{VII}-\underbrace{2\beta_y\varepsilon_{1t}}_{VIII}-\underbrace{\xi_1\xi_y x_t^{*^2}}_{IX}+const, \end{aligned} \tag{5.21}$$

and finally (5.19) is

$$\begin{aligned} z_t &= -2(\varepsilon_{1t}-\varepsilon_{2t})\eta_t+\varepsilon_{1t}^2-\varepsilon_{2t}^2+const, \\ &= -2\underbrace{\varepsilon_{1t}\eta_t}_{I}+\underbrace{2\varepsilon_{2t}\eta_t}_{II}+\underbrace{\varepsilon_{1t}^2}_{III}-\underbrace{\varepsilon_{2t}^2}_{IV}+const. \end{aligned} \tag{5.22}$$

We can now proceed as in the proof of Proposition 5.2 and infer the memory orders of each term in the respective linear combination from Proposition 5.1 and then determine the maximum as in Proposition 3 in Chambers (1998).
In the following, we label the terms appearing in each of the equations by consecutive letters with the equation number as an index. For the terms in (5.20), we have

$$I_{5.20} \sim \begin{cases} LM(\max\{d_x+d_{\varepsilon_1}-1/2,\ 0\}), & \text{if } S_{x,\varepsilon_1}\neq 0 \\ LM(d_x+d_{\varepsilon_1}-1/2), & \text{if } S_{x,\varepsilon_1}=0 \end{cases}$$

$$II_{5.20} \sim \begin{cases} LM(\max\{d_x+d_{\varepsilon_2}-1/2,\ 0\}), & \text{if } S_{x,\varepsilon_2}\neq 0 \\ LM(d_x+d_{\varepsilon_2}-1/2), & \text{if } S_{x,\varepsilon_2}=0 \end{cases}$$

$$III_{5.20} \sim \begin{cases} LM(\max\{d_y+d_{\varepsilon_1}-1/2,\ 0\}), & \text{if } S_{y,\varepsilon_1}\neq 0 \\ LM(d_y+d_{\varepsilon_1}-1/2), & \text{if } S_{y,\varepsilon_1}=0 \end{cases}$$

$$IV_{5.20} \sim \begin{cases} LM(\max\{d_y+d_{\varepsilon_2}-1/2,\ 0\}), & \text{if } S_{y,\varepsilon_2}\neq 0 \\ LM(d_y+d_{\varepsilon_2}-1/2), & \text{if } S_{y,\varepsilon_2}=0 \end{cases}$$

$$V_{5.20} \sim LM(\max\{2d_{\varepsilon_1}-1/2,\ 0\})$$

$$\text{and} \qquad VI_{5.20} \sim LM(\max\{2d_{\varepsilon_2}-1/2,\ 0\}).$$

Since by definition $d_x > d_{\varepsilon_i}$, the memory of $V_{5.20}$ and $VI_{5.20}$ is always of a lower order than that of

$I_{5.20}$ *and* $II_{5.20}$*. As in the proof of Proposition 5.2, the squares in terms* $V_{5.20}$ *and* $VI_{5.20}$ *establish zero as the lower bound of* d_z*. Therefore, we have*

$$d_z = \max\Big\{\max\Big\{d_x,\ d_y\Big\} + \max\{d_{\varepsilon_1},\ d_{\varepsilon_2}\} - 1/2,\ 0\Big\}.$$

Similarly, in (5.21), we have

$$I_{5.21} \sim \begin{cases} LM(\max\{d_x + d_2 - 1/2,\ 0\}), & \text{if } S_{x,\widehat{y}_2} \neq 0 \\ LM(d_x + d_2 - 1/2), & \text{if } S_{x,\widehat{y}_2} = 0 \end{cases}$$

$$II_{5.21} \sim \begin{cases} LM(\max\Big\{d_x + d_\eta - 1/2,\ 0\Big\}), & \text{if } S_{x,\eta} \neq 0 \\ LM(d_x + d_\eta - 1/2), & \text{if } S_{x,\eta} = 0 \end{cases}$$

$$III_{5.21} \sim \begin{cases} LM(\max\Big\{d_{\varepsilon_1} + d_\eta - 1/2,\ 0\Big\}), & \text{if } S_{\varepsilon_1,\eta} \neq 0 \\ LM(d_{\varepsilon_1} + d_\eta - 1/2), & \text{if } S_{\varepsilon_1,\eta} = 0 \end{cases}$$

$$IV_{5.21} \sim \begin{cases} LM(\max\Big\{d_2 + d_\eta - 1/2,\ 0\Big\}), & \text{if } S_{\widehat{y}_2,\eta} \neq 0 \\ LM(d_2 + d_\eta - 1/2), & \text{if } S_{\widehat{y}_2,\eta} = 0 \end{cases}$$

$$V_{5.21} \sim LM(d_{\varepsilon_1})$$

$$VI_{5.21} \sim LM(\max\{2d_{\varepsilon_1} - 1/2,\ 0\})$$

$$VII_{5.21} \sim LM(\max\{2d_2 - 1/2,\ 0\})$$

$$VIII_{5.21} \sim LM(d_{\varepsilon_1})$$

$$\text{and} \qquad IX_{5.21} \sim LM(\max\{2d_x - 1/2, 0\}).$$

Here, $V_{5.21}$ *can be disregarded since it is of the same order as* $VIII_{5.21}$*.* $VIII_{5.21}$ *dominates* $VI_{5.21}$*, because* $d_{\varepsilon_1} < 1/2$*. Finally, as* $d_{\varepsilon_1} < d_x$ *holds by assumption,* $III_{5.21}$ *is dominated by* $II_{5.21}$ *and* $d_\eta < d_x$*, so that* $IX_{5.21}$ *dominates* $II_{5.21}$*. Therefore,*

$$d_z = \max\Big\{d_2 + \max\Big\{d_x, d_\eta\Big\} - 1/2,\ 2\max\{d_x, d_2\} - 1/2,\ d_{\varepsilon_1}\Big\}.$$

As before, for the case of CLM between y_t *and* $\widehat{y}_{2t}$*, the proof is entirely analogous, but with index* "1" *replaced by* "2" *and vice versa.*

Finally, in (5.22), we have

$$
\begin{aligned}
I_{5.22} &\sim \begin{cases} LM(\max\left\{d_\eta + d_{\varepsilon_1} - 1/2,\ 0\right\}), & \text{if } S_{\eta,\varepsilon_1} \neq 0 \\ LM(d_\eta + d_{\varepsilon_1} - 1/2), & \text{if } S_{\eta,\varepsilon_1} = 0 \end{cases} \\
II_{5.22} &\sim \begin{cases} LM(\max\left\{d_\eta + d_{\varepsilon_2} - 1/2,\ 0\right\}), & \text{if } S_{\eta,\varepsilon_1} \neq 0 \\ LM(d_\eta + d_{\varepsilon_2} - 1/2), & \text{if } S_{\eta,\varepsilon_2} = 0 \end{cases} \\
III_{5.22} &\sim LM(\max\{2d_{\varepsilon_1} - 1/2,\ 0\}) \\
IV_{5.22} &\sim LM(\max\{2d_{\varepsilon_2} - 1/2,\ 0\}).
\end{aligned}
$$

Here, no further simplifications can be made, since we do not impose restrictions on the relationship between d_η, d_{ε_1} and d_{ε_2}, so that

$$
d_z = \max\left\{d_\eta + \max\{d_{\varepsilon_1},\ d_{\varepsilon_2}\} - 1/2,\ 2\max\{d_{\varepsilon_1},\ d_{\varepsilon_2}\} - 1/2,\ 0\right\},
$$

where again the zero is established as the lower bound by the squares in $III_{5.22}$ and $IV_{5.22}$.

Proof (Proposition 5.5)**.** *First note that under short memory, the t_{HAC}-statistic is given by*

$$
t_{HAC} = T^{1/2}\frac{\bar{z}}{\sqrt{\widehat{V}_{HAC}}},
$$

with $\widehat{V}_{HAC} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{B}\right)\widehat{\gamma}_z(j)$ and B being the bandwidth satisfying $B \to \infty$ and $B = O(T^{1-\epsilon})$ for some $\epsilon > 0$. From Abadir et al. (2009), the appropriately scaled long-run variance estimator for a long memory processes is given by $B^{-1-2d}\sum_{i,j=1}^{B}\widehat{\gamma}_z(|i-j|)$, see equation (2.2) in Abadir et al. (2009). Corresponding long memory robust HAC-type estimators (with a Bartlett kernel, for instance) take the form

$$
\widehat{V}_{HAC,d} = B^{-2d}\left(\widehat{\gamma}_z(0) + 2\sum_{j=1}^{B}(1 - j/B)\widehat{\gamma}_z(j)\right).
$$

The long memory robust $t_{HAC,d}$-statistic is then given by

$$
t_{HAC,d} = T^{1/2-d}\frac{\bar{z}}{\sqrt{\widehat{V}_{HAC,d}}}.
$$

We can therefore write

$$
t_{HAC,d} = T^{1/2}T^{-d}\frac{\bar{z}}{\sqrt{B^{-2d}\widehat{V}_{HAC}}} = \frac{T^{-d}}{B^{-d}}t_{HAC}
$$

and thus,

$$t_{HAC} = \frac{T^d}{B^d} t_{HAC,d}.$$

The short memory t_{HAC}-statistic is inflated by the scaling factor $T^d/B^d = O(T^{d\epsilon})$. This leads directly to the divergence of the $_{HAC}$-statistic ($t_{HAC} \to \infty$ as $T \to \infty$) which implies that

$$\lim_{T\to\infty} P(|t_{HAC}| > c_{1-\alpha/2,d}) = 1$$

for all values of $d \in (0,1/4) \cup (1/4,1/2)$. For $0 < d < 1/4$, $c_{1-\alpha/2,d}$ is the critical value from the $N(0,1)$-distribution, while for $1/4 < d < 1/2$, the critical value (depending with d) stems from the well-defined Rosenblatt distribution, see Abadir et al. (2009). The proof is analogous for other kernels and thus omitted. □

Additional Simulation Results

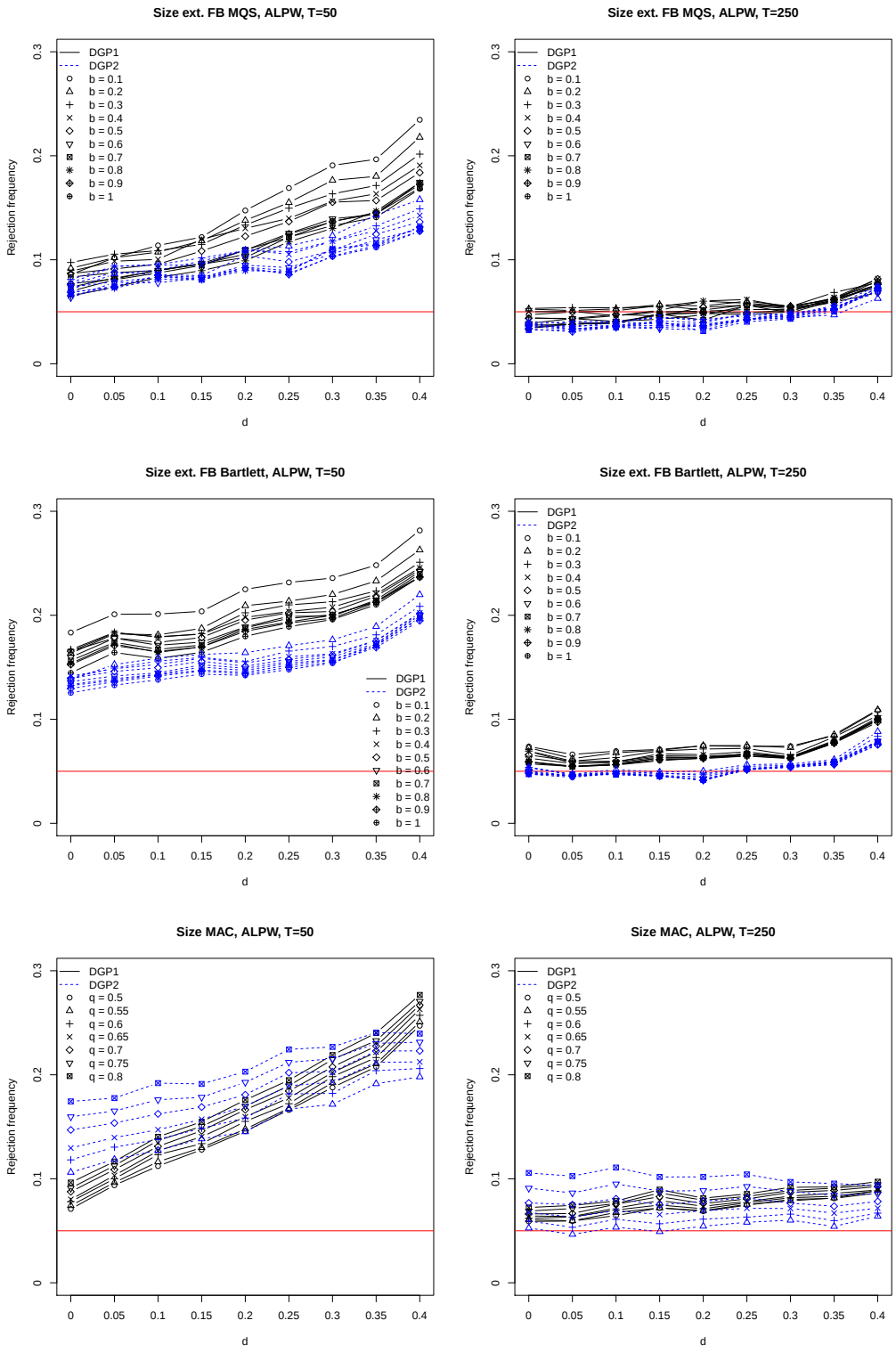

Figure 5.8: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes T and bandwidth parameters q and b, if the ALPW estimator is used for the plug-in estimation of d.

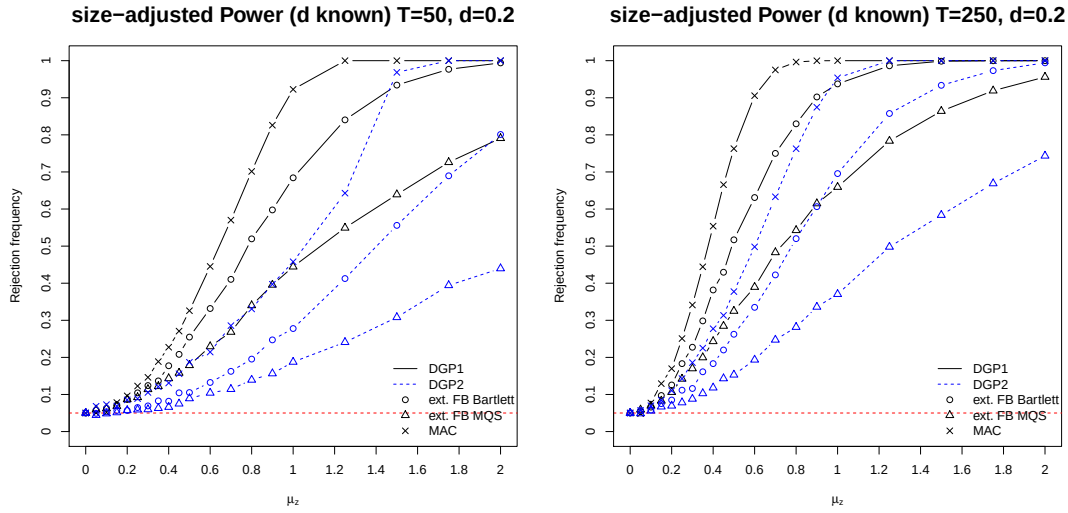

Figure 5.9: Power comparison of the t_{MAC}- and t_{EFB}-statistics, for different memory parameters d and sample sizes T, adjusted for size and with known memory parameter d.

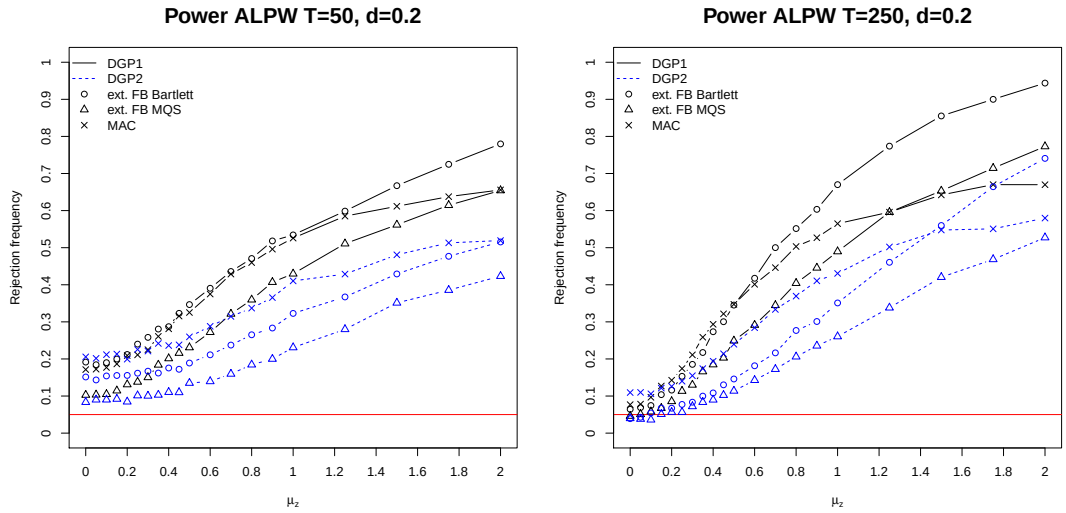

Figure 5.10: Power comparison of the t_{MAC}- and t_{EFB}-statistics, for different memory parameters d and sample sizes T if the ALPW estimator is used for the plug-in estimation of d.

Bibliography

Abadir, K. M., Distaso, W., and Giraitis, L. (2007). Nonstationarity-extended local Whittle estimation. *Journal of Econometrics*, 141(2):1353–1384.

Abadir, K. M., Distaso, W., and Giraitis, L. (2009). Two estimators of the long-run variance: Beyond short memory. *Journal of Econometrics*, 150(1):56–70.

Abeysinghe, T. (1991). Inappropriate use of seasonal dummies in regression. *Economics Letters*, 36(2):175–179.

Abeysinghe, T. (1994). Deterministic seasonal models and spurious regressions. *Journal of Econometrics*, 61(2):259–272.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2):115–158.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.

Andrews, D. W. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Andrews, D. W. K. and Sun, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica*, 72(2):569–614.

Arteche, J. (2002). Semiparametric robust tests on seasonal or cyclical long memory time series. *Journal of Time Series Analysis*, 23(3):251–285.

Arteche, J. and Robinson, P. M. (2000). Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis*, 21(1):1–25.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory*, 13(03):315–352.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30.

Becker, R., Clements, A. E., and McClelland, A. (2009). The jump component of S&P 500 volatility and the VIX index. *Journal of Banking & Finance*, 33(6):1033–1038.

Becker, R., Clements, A. E., and White, S. I. (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking & Finance*, 31(8):2535–2549.

Bekaert, G. and Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57:659–672.

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

Beran, J. and Terrin, N. (1994). Estimation of the long-memory parameter, based on a multivariate central limit theorem. *Journal of Time Series Analysis*, 15(3):269–278.

Berkes, I., Rorvath, L., Kokoszka, P., and Shao, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. *The Annals of Statistics*, 34(3):1140–1165.

Billingsley, P. (2009). *Convergence of probability measures*. Wiley.

Bisaglia, L., Bordignon, S., and Lisi, F. (2003). k-Factor GARMA models for intraday volatility forecasting. *Applied Economics Letters*, 10(4):251–254.

Bohrnstedt, G. W. and Goldberger, A. S. (1969). On the exact covariance of products of random variables. *Journal of the American Statistical Association*, 64(328):1439–1442.

Bollerslev, T., Osterrieder, D., Sizova, N., and Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108(2):409–424.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463–4492.

Bordignon, S., Caporin, M., and Lisi, F. (2008). Periodic long-memory GARCH models. *Econometric Reviews*, 28(1-3):60–82.

Busch, T., Christensen, B. J., and Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160(1):48–57.

Chambers, M. J. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review*, 39(4):1053–1072.

Chen, X. and Ghysels, E. (2011). News–good or bad–and its impact on volatility predictions over multiple horizons. *Review of Financial Studies*, 24(1):46–81.

Chernov, M. (2007). On the role of risk premia in volatility forecasting. *Journal of Business & Economic Statistics*, 25(4):411–426.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Choi, H.-S. and Kiefer, N. M. (2010). Improving robust model selection tests for dynamic models. *The Econometrics Journal*, 13(2):177–204.

Christensen, B. J. and Nielsen, M. Ø. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, 133(1):343–371.

Clark, T. E. (1999). Finite-sample properties of tests for equal forecast accuracy. *Journal of Forecasting*, 18(7):489–504.

Cogburn, R. and Davis, H. T. (1974). Periodic splines and spectral estimation. *The Annals of Statistics*, 2(6):1108–1126.

Corradi, V. (1995). Nonlinear transformations of integrated time series: a reconsideration. *Journal of Time Series Analysis*, 16(6):539–549.

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

Corsi, F., Pirino, D., and Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288.

Corsi, F. and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30(3):368–380.

Davidson, J. and Rambaccussing, D. (2015). A test of the long memory hypothesis based on self-similarity. *Journal of Time Series Econometrics*, 7(2):115–141.

De Boor, C. (1978). A practical guide to splines. *Mathematics of Computation*.

Deo, R., Hurvich, C., and Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics*, 131(1):29–58.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–8.

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

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Diongue, A. K., Guegan, D., and Vignal, B. (2009). Forecasting electricity spot market prices with a k-factor GIGARCH process. *Applied Energy*, 86(4):505–510.

Dittmann, I. and Granger, C. W. J. (2002). Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics*, 110(2):113–133.

Dolado, J. J., Gonzalo, J., and Mayoral, L. (2005). What is what?: A simple time-domain test of long-memory vs. structural breaks. Unpublished Manuscript, Department of Economics, Universidad Carlos III de Madrid.

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

Fitzsimmons, P. and McElroy, T. (2010). On joint Fourier–Laplace transforms. *Communications in Statistics – Theory and Methods*, 39(10):1883–1885.

Franses, P. H., Hylleberg, S., and Lee, H. S. (1995). Spurious deterministic seasonality. *Economics Letters*, 48(3):249–256.

Frederiksen, P., Nielsen, F. S., and Nielsen, M. Ø. (2012). Local polynomial Whittle estimation of perturbed fractional processes. *Journal of Econometrics*, 167(2):426–447.

García-Enríquez, J., Hualde, J., Arteche, J., and Murillas-Maza, A. (2014). Spatial integration in the spanish mackerel market. *Journal of Agricultural Economics*, 65(1):234–256.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Gil-Alana, L. A. (2002). Seasonal long memory in the aggregate output. *Economics Letters*, 74(3):333–337.

Gil-Alana, L. A. (2007). Testing the existence of multiple cycles in financial and economic time series. *Annals of Economics & Finance*, 8(1):1–20.

Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large sample inference for long memory processes*. World Scientific Publishing Company Incorporated.

Giraitis, L. and Leipus, R. (1995). A generalized fractionally differencing approach in long-memory modeling. *Lithuanian Mathematical Journal*, 35(1):53–65.

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55(292):708–713.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14(2):227–238.

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

Granger, C. W. J. and Hallman, J. (1991). Nonlinear transformations of integrated time series. *Journal of Time Series Analysis*, 12(3):207–224.

Granger, C. W. J. 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.

Gray, H. L., Zhang, N.-F., and Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, 10(3):233–257.

Haldrup, N. and Kruse, R. (2014). Discriminating between fractional integration and spurious long memory. Unpublished Manuscript, Department of Economics, University of Aarhus.

Haldrup, N. and Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, 135(1):349–376.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hassler, U. (1994). (Mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis*, 15(1):19–30.

Hassler, U. and Olivares, M. (2013). Semiparametric inference and bandwidth choice under long memory: Experimental evidence. *İstatistik, Journal of the Turkish Statistical Association*, 6(1):27–41.

Hassler, U., Rodrigues, P. M. M., and Rubia, A. (2009). Testing for general fractional integration in the time domain. *Econometric Theory*, 25:1793–1828.

Henry, M. (2001). Robust automatic bandwidth for long memory. *Journal of Time Series Analysis*, 22(3):293–316.

Henry, M. and Robinson, P. M. (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In *Athens Conference on Applied Probability and Time Series Analysis*, pages 220–232. Springer.

Hidalgo, J. and Soulier, P. (2004). Estimation of the location and exponent of the spectral singularity of a long memory process. *Journal of Time Series Analysis*, 25(1):55–81.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrica*, 68(1):165–176.

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.

Hualde, J., Robinson, P. M., et al. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *The Annals of Statistics*, 39(6):3152–3181.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions*, volume 2. Wiley New York, 2 edition.

Kechagias, S. and Pipiras, V. (2015). Definitions and representations of multivariate long-range dependent time series. *Journal of Time Series Analysis*, 36(1):1–25.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(6):1130–1164.

Künsch, H. (1987). Statistical aspects of self-similar processes. In Prohorov, Y. and Sazanov, V. V., editors, *Proceedings 1st world congress of the Bernoulli society*, volume 1, pages 67–74.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995). Rate of convergence for logspline spectral density estimation. *Journal of Time Series Analysis*, 16(4):389–401.

Kruse, R. (2015). A modified test against spurious long memory. *Economics Letters*, 135:34–38.

Kruse, R., Leschinski, C., and Will, M. (2016). Comparing predictive accuracy under long memory - with an application to volatility forecasting. Technical Report 571.

Leccadito, A., Rachedi, O., and Urga, G. (2015). True versus spurious long memory: Some theoretical results and a Monte Carlo comparison. *Econometric Reviews*, 34(4):452–479.

Leschinski, C. (2016). On the memory of products of long range dependent time series. Technical Report 569, Leibniz University of Hannover.

Li, J. and Patton, A. J. (2013). Asymptotic inference about predictive accuracy using high frequency data. *Economic Research Initiatives at Duke (ERID) Working Paper*, (163).

Lobato, I. N. (1997). Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis*, 18(2):137–155.

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

Lobato, I. N. and Savin, N. E. (1998). Real and spurious long-memory properties of stock-market data. *Journal of Business & Economic Statistics*, 16(3):261–268.

Lu, Y. K. and Perron, P. (2010). Modeling and forecasting stock return volatility using a random level shift model. *Journal of Empirical Finance*, 17(1):138–156.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Springer.

Mariano, R. S. and Preve, D. (2012). Statistical tests for multiple forecast comparison. *Journal of Econometrics*, 169(1):123–130.

Marinucci, D. and Robinson, P. M. (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference*, 80(1):111–122.

Martens, M., Van Dijk, D., and De Pooter, M. (2009). Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting*, 25(2):282–303.

McCloskey, A. and Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory*, 29(06):1196–1237.

McElroy, T. and Politis, D. N. (2012). Fixed-b asymptotics for the studentized mean from time series with short, long, or negative memory. *Econometric Theory*, 28(2):471–481.

Mikosch, T. and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86(1):378–390.

Murphy, K. M. and Topel, R. H. (2002). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, 20(1):88–97.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

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

Nielsen, M. Ø. (2007b). 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. (2015). Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time-series models. *Journal of Time Series Analysis*, 36(2):154–188.

Ohanissian, A., Russell, J. R., and Tsay, R. S. (2008). True or spurious long memory? A new test. *Journal of Business & Economic Statistics*, 26(2):161–175.

Patton, A. J. (2015). Comment on "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests" by Francis X. Diebold. *Journal of Business and Economic Statistics*, 33(1):22–24.

Perron, P. and Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics*, 28(2):275–290.

Phillips, P. C. B. and Kim, C. S. (2007). Long-run covariance matrices for fractionally integrated processes. *Econometric Theory*, 23(6):1233–1247.

Politis, D. N. and McElroy, T. S. (2009). Fixed-b asymptotics for the studentized mean from time series with short, long or negative memory. *Department of Economics, UCSD*.

Porter-Hudak, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of the American Statistical Association*, 85(410):338–344.

Priestley, M. B. (1981). *Spectral analysis and time series*. Academic press.

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

Ramanathan, R., Engle, R., Granger, C. W. J., Vahid-Araghi, F., and Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, 13(2):161–174.

Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association*, 89(428):1420–1437.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):1630–1661.

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(4):1048–1072.

Robinson, P. M. (2005). Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory*, 21(1):171–180.

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

Robinson, P. M. and Henry, M. (1999). Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels. *Econometric Theory*, 15(03):299–336.

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

Rossi, B. (2005). Testing long-horizon predictive ability with high persistence, and the Meese–Rogoff puzzle*. *International Economic Review*, 46(1):61–92.

Rossi, E. and Fantazzini, D. (2015). Long memory and periodicity in intraday volatility. *Journal of Financial Econometrics*, 13(4):1–40.

Sela, R. J. and Hurvich, C. M. (2009). Computationally efficient methods for two multivariate fractionally integrated models. *Journal of Time Series Analysis*, 30(6):631–651.

Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Unpublished Manuscript, Department of Economics, Queens's University.

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

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

Shimotsu, K. and Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, 33(4):1890–1933.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754.

Soares, L. J. and Souza, L. R. (2006). Forecasting electricity demand using generalized long memory. *International Journal of Forecasting*, 22(1):17–28.

Sowell, F. (1989). Maximum likelihood estimation of fractionally integrated time series models. Unpublished Manuscript, Department of Economics, Carnegie Mellon University.

Sun, Y., Phillips, P. C. B., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, 76(1):175–194.

Varneskov, R. T. and Perron, P. (2011). Combining long memory and level shifts in modeling and forecasting the volatility of asset returns. Unpublished Manuscript, Department of Economics, Boston University.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1):87–127.

Walker, G. (1914). On the criteria for the reality of relationships or periodicities. *Calcutta Ind. Met. Memo.*

Wecker, W. E. (1978). A note on the time series which is the product of two stationary time series. *Stochastic Processes and their Applications*, 8(2):153–157.

Weron, R. and Misiorek, A. (2008). Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, 24(4):744–763.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084.

Woodward, W. A., Cheng, Q. C., and Gray, H. L. (1998). A k-factor GARMA long-memory model. *Journal of Time Series Analysis*, 19(4):485–504.

Xu, J. and Perron, P. (2014). Forecasting return volatility: Level shifts with varying jump probability and mean reversion. *International Journal of Forecasting*, 30(3):449–463.

Yajima, Y. (1996). Estimation of the frequency of unbounded spectral densities. In *Proceedings of the Business and Economic Statistical Section*, pages 4–7.

Yau, C. Y. and Davis, R. A. (2012). Likelihood inference for discriminating between long-memory and change-point models. *Journal of Time Series Analysis*, 33(4):649–664.