Essays on Robust Long Memory Inference

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. Michael Wolfgang Will

geboren am 20. Januar 1988 in Gehrden

2018

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

Korreferent: Prof. Dr. Michael Massmann, WHU - Otto Beisheim School of Management, Vallendar

Tag der Promotion: 11.01.2018

Acknowledgements

First, I want to thank my advisor and co-author Prof. Dr. Philipp Sibbertsen for his continuous support, guidance and for always providing me a good working atmosphere to pursue my research projects. I also want to thank Prof. Dr. Michael Massmann for taking the time to be my second examiner and Prof. Dr. Stefan Wielenberg for chairing my examination board.

Moreover, I am deeply grateful to my co-authors: Christian Leschinski, for his advice and many inspiring discussions, as well as Robinson Kruse, Matei Demetrescu and Kai Wenger, for all their effort and companionship. I truly learned a lot from each of them.

In my time at the Institute of Statistics in Hannover, which started 2011 as a student assistant and continued from 2014 onwards as a research assistant, I was fortunate to meet many people who left so many memories and impressions. Also I thank all my current and former colleagues, who always kept an open door and were always willing to help.

Finally, I want to thank my friends and family - especially my mother and Martin, for their everlasting support, patience and love, for always being understanding, and for always believing in me even during setbacks along the way.

Hannover, January 2018
Michael Will

Kurzzusammenfassung

Diese Doktorarbeit enthält drei Aufsätze über robuste statistische Inferenz unter langem Gedächtnis. Nach einer Einleitung in Kapitel 1 erweitert der erste Aufsatz in Kapitel 2 den populären Test von Diebold und Mariano (1995) auf Situationen, in denen das Prognosefehlerverlustdifferenzial langes Gedächtnis aufweist. Es wird theoretisch gezeigt, dass diese Situation häufig auftritt, da das lange Gedächtnis von den Prognosereihen und von der zu prognostizierenden Reihe auf das Prognosefehlerverlustdifferenzial übertragen werden kann. Wir zeigen, dass der Mechanismus der Übertragung hauptsächlich von der (Un)Verzerrtheit der Prognosereihen abhängt und ob die involvierten Reihen (un)balanciertes gemeinsames Gedächtnis teilen. Weitere theoretische und simulationsbasierte Resultate zeigen, dass der konventionelle Diebold und Mariano (1995) Test in dieser Situation invalidiert wird.
Wir verwenden daher robuste Statistiken, welche einerseits auf dem gedächtnis- und autokorrelationskonsistenten Schätzer von Robinson (2005) und andererseits auf dem erweiterten fixierten-Bandbreiten Ansatz von McElroy und Politis (2012) beruhen. Die folgende Monte Carlo Analyse liefert einen neuartigen Vergleich dieser robusten Statistiken und adressiert ferner viele Anwenderfragen wie Kern- und Bandbreitenwahl, sowie Auswahl des Plug-in Schätzers.
Als empirische Anwendungen führen wir Prognosevergleichstests für realisierte Volatilität des Standard und Poors 500 Index unter Verwendung kürzlich weiterentwickelter Modelle des heterogenen autoregressiven Modells (siehe Corsi (2009), Corsi et al. (2010) und Corsi und Renò (2012)) durch. Während wir herausfinden, dass Prognosen sich signifikant verbessern, wenn Sprünge im log-Preisprozess separiert von der kontinuierlichen Komponente berücksichtigt werden, so zeigt sich auch, dass Verbesserungen, die durch das Berücksichtigen von implizierter Volatilität herbeigeführt werden, in den meisten Situationen insignifikant sind.

Im zweiten Aufsatz in Kapitel 3 schlagen wir einen nichtparametrischen Lagrange Multiplier basierten Test auf langes Gedächtnis im Zeitbereich vor der Robustheit gegen fixe und wechselnde Kurzfristdynamiken und Heteroskedastizität von recht allgemeiner Form aufweist. Unser Test basiert auf dem Verfahren von Harris et al. (2008), das eine nichtparametrische kurze Gedächtnis-Korrektur verwendet. Dieses wird von uns in einen Regressionsrahmen umgestaltet im Stile von Breitung und Hassler (2002) und sodann mittels eines heteroskedastizitäts- und autokorrelationskonsistenten Schätzers von Andrews (1991) standardisiert.
Eine umfangreiche Monte Carlo Simulationsstudie zeigt, dass die Grenzverteilung unseres Tests Standardnormal ist und wir trotz der zahlreichen Robustheitseigenschaften gute Powerergebnisse erzielen. Ferner zeigt ein Vergleich mit anderen Zeitbereich-basierten Lagrange Multiplier Tests aus der Literatur, dass unser Verfahren als Einziges seine Size unter der hier berücksichtigten heteroskedastisch lokal stationären Prozessklasse (siehe Dahlhaus (2000), Palma und Olea (2010), und Cavaliere et al. (2015b)) einhält.
In unserer empirischen Anwendung zeigt sich, dass langes Gedächtnis in verschiedenen ex-post Varianzrisikoprämienreihen zurückbleibt, welche durch eine fraktionelle Kointegrationsbeziehung

zwischen Chicago Board of Options Exchange Volatilitätsindex und realisierter Varianz des Standard und Poors 500 Index repräsentiert werden, zusätzlich zu bedingter und unbedingter Heteroskedastizität.

Im dritten Aufsatz in Kapitel 4 zeigen wir schließlich mittels Simulationsstudien, dass der bekannte Qu (2011) Test auf scheinbares langes Gedächtnis beträchtliche Size-Verzerrungen aufweist, wenn er auf eine Klasse kürzlich entwickelter lokal stationärer Prozesse (siehe Cavaliere et al. (2015b) und Demetrescu und Sibbertsen (2016)) und stochastische Koeffizienten Modelle (siehe Giraitis et al. (2014)) angewendet wird, obgleich diese Modelle eine semiparametrische Darstellung besitzen, die allgemeinhin mit wahrem negativen, kurzen oder langen Gedächtnis assoziiert wird.
Eine umfangreiche Monte Carlo Simulationsstudie zeigt, dass der induzierte liberale Bias leicht zu einer falschen Ablehnung der Nullhypothese führt und auch nicht vernachlässigbar in großen Stichproben bleibt. Dies verdeutlicht, dass nicht nur Prozesse mit Kontaminationen im Niedrigfrequenzbereich mit wahrem langen Gedächtnis verwechselt werden können, sondern dass unter einigen schwächeren Annahmen, wahre negative, kurze oder lange Gedächtnis Prozesse fälschlich zur Klasse der Prozesse mit scheinbarem langen Gedächtnis gezählt werden können. Wir schlagen darüberhinaus ein einfaches Prewhitening-Verfahren vor, das gute Sizeresultate schon in kleinen Stichproben wiederherstellt.
Als empirisches Beispiel berücksichtigen wir einerseits log-Spotpreise, korrespondierende Ein-Perioden Futurespreise, sowie Spreads von Gold, Silber, Platinum und Rohöl als auch andererseits monatliche (saisonbereinigte) Inflationsraten der G7-Staaten. Unsere Ergebnisse zeigen, dass insbesondere die erstgenannten Daten von einer falschinduzierten Ablehnung der Nullhypothese betroffen sind, insbesondere wenn eine empfohlene benutzerspezifische Bandbreitenwahl verwendet wird.

Schlagworte: Langes Gedächtnis · Robuste Statistik · Lokal Stationäre Prozesse.

Short Summary

This doctoral thesis contains three essays on robust long memory inference. Following an introduction in Chapter 1, the first essay in Chapter 2 extends the popular test of Diebold and Mariano (1995) to situations when the forecast error loss differential exhibits long memory. It is shown theoretically that this situation can arise frequently, since long memory can be transmitted from forecasts and the forecast objective to the forecast error loss differential. We show that the nature of this transmission mainly depends on the (un)biasedness of the forecasts and whether the involved series share (un)balanced common long memory. Further theoretical and simulation results show that the conventional test of Diebold and Mariano (1995) is invalidated under these circumstances.

Robust statistics based on a memory and autocorrelation consistent estimator by Robinson (2005) and an extended fixed-bandwidth approach of McElroy and Politis (2012) are considered. The subsequent Monte Carlo study provides a novel comparison of these robust statistics and addresses several practitioners questions like choice of kernel, bandwidth selection and choice of plug-in estimator.

As empirical applications, we conduct forecast comparison tests for the realized volatility of the Standard and Poors 500 index among recent extensions of the heterogeneous autoregressive model (see Corsi (2009), Corsi et al. (2010), and Corsi and Renò (2012)). While 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 in most situations.

In the second essay in Chapter 3 we propose a nonparametric time-domain based Lagrange Multiplier test for long memory with robustness against fixed or switching short-run dynamics and heteroskedasticity of quite general form. Our test is based on the procedure of Harris et al. (2008), utilizing a nonparametric short memory correction, which is subsequently recast in a regression framework in the style of Breitung and Hassler (2002) and standardized using a heteroskedasticity and autocorrelation consistent estimator of Andrews (1991).

An extensive Monte Carlo simulation study shows that its limiting null distribution is standard normal and despite of its several robustness features, we obtain good power results. Moreover, a comparison with other existing time-domain based Lagrange Multiplier testing procedures in the literature shows that it is the only test capable of controlling its size under the heteroskedastic locally stationary process class (c.f. Dahlhaus (2000), Palma and Olea (2010), and Cavaliere et al. (2015b)) considered here.

In our empirical application we find that long memory remains in different types of ex-post Variance Risk Premium series, represented by a fractional cointegration relationship between the Chicago Board of Options Exchange Volatility Index and realized variance of the Standard and Poors 500 index, in addition to conditional and unconditional heteroskedasticity.

Finally in the third essay in Chapter 4 we show by means of simulation study that the well known Qu (2011) test on spurious long memory has substantial size distortions when applied to a recently proposed class of locally stationary processes (see Cavaliere et al. (2015b) and Demetrescu and Sibbertsen (2016)) and random coefficient models (c.f. Giraitis et al. (2014)), although these models obey a semiparametric representation commonly attributed to true negative, short or long memory.
A large scale Monte Carlo simulation study reveals that the induced liberal bias easily leads to a potential spurious rejection of the null and also remains non-negligible in very large samples. This demonstrates that not only processes displaying low-frequency contaminations can be confused with true long memory, but, under some weaker assumptions, true negative, short or long memory processes can be wrongly attributed to the spurious long memory class. Furthermore, we propose a simple prewhitening procedure that recreates good size results and works well even in small samples.
As empirical applications we consider on one hand log spot prices, corresponding one-period futures contract prices, as well as spreads of gold, silver, platinum, and crude oil and on the other hand monthly (seasonally adjusted) inflation rates of G7 countries. We find that in particular the former series are subject to a spurious rejection of the null hypothesis, especially when using recommended user-specific bandwidth choices.

Key words: Long Memory · Robust Statistics · Locally Stationary Processes.

„Von Herzen - Möge es wieder - Zu Herzen gehn!“

– Ludwig van Beethoven, dedication of his “Missa solemnis” op. 123.

Contents

1 Introduction **1**

2 Comparing Predictive Accuracy under Long Memory
- With an Application to Volatility Forecasting - **6**

2.1 Introduction 7
2.2 Diebold-Mariano Test 8
2.2.1 Conventional Approach: HAC 9
2.2.2 Fixed-bandwidth Approach 10
2.3 Long Memory in Forecast Error Loss Differentials 11
2.3.1 Preliminaries 11
2.3.2 Transmission of Long Memory to the Loss Differential 13
2.3.3 Memory Transmission under Common Long Memory 15
2.3.4 Asymptotic and Finite-Sample Behaviour under Long Memory 19
2.4 Long-Run Variance Estimation under Long Memory 20
2.4.1 MAC Estimator 20
2.4.2 Extended Fixed-Bandwidth Approach 22
2.5 Monte Carlo Study 23
2.6 Applications to Realized Volatility Forecasting 29
2.6.1 Predictive Ability of the VIX for Quadratic Variation 32
2.6.2 Separation of Continuous Components and Jump Components 36
2.7 Conclusion 37
2.8 Appendix 39
Proof of Proposition 2.2 39
Proof of Proposition 2.3 41
Proof of Proposition 2.4 42
Proof of Proposition 2.5 43
Proof of Proposition 2.6 46
2.8.1 Supplementary Appendix 47

3 A Robust Lagrange Multiplier Test for Long Memory **65**

3.1 Testing for Long Memory under Locally Stationary Processes 68
3.1.1 Lagrange Multiplier Tests in the Time-Domain 71

3.1.2 A new Robust Lagrange Multiplier Test for Long Memory 75
3.2 Monte Carlo Simulations 76
3.3 Empirical Analysis 87
3.4 Conclusion 92
3.5 Appendix 93

4 A Criticism on Spurious Long Memory 95
4.1 Introduction 96
4.2 True versus Spurious Long Memory 97
4.2.1 Locally Stationary Processes 101
4.2.2 Random Coefficient Models 103
4.3 Monte Carlo Simulations 106
4.4 Empirical Analysis 113
4.4.1 Commodity Application (Locally Stationary Processes) 113
4.4.2 Inflation Rates Application (Random Coefficient Models) 117
4.5 Conclusion 120
4.6 Appendix 122
Proof of Corollary 1 122

Bibliography 130

Chapter 1

Introduction

Introduction

In time series analysis one refers to a discrete-time second-order stationary stochastic process of displaying long memory if its autocorrelations decay very slowly, namely at a hyperbolic rate and if its spectral density is unbounded at the origin. Practically spoken the autocorrelation function $\gamma(k)$ of a long memory process with memory parameter d, evaluated at lag $k \in \mathbb{Z}$, suffices $\gamma(k) \sim Gk^{2d-1}$, as $k \to \infty$ whereas its spectral density $f(\lambda)$ obeys the rate condition $f(\lambda) \simeq G\lambda^{-2d}$, as $\lambda \to 0_+$, with $\lambda \in [-\pi, \pi]$ denoting the frequency and $G \in (0, \infty)$ (e.g. Beran et al. (2013)). These properties remarkably distinguish this process class from short memory processes whose autocorrelations instead decay at most at an exponential rate by simultaneously displaying a bounded spectrum across all frequencies. The phenomenon of long memory in the data can be traced back at least to Hurst (1951) who analyzed the flow regularization properties of the Nile River. First theoretical justification was developed by Mandelbrot and Van Ness (1968) who considered a fractional Brownian motion to model long memory. Today, autoregressive fractionally integrated moving average: $ARFIMA(p,d,q)$ processes originally proposed by Granger and Joyeux (1980) and Hosking (1981) remain the most prominent procedure to model long memory. More concrete, let x_t for $t = 1, \ldots, T$ be given by:

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

where L is the lag operator, $\phi(L) = (1 - \phi_1 L - \ldots - \phi_p L^p)$ and $\theta(L) = (1 + \theta_1 L + \ldots + \theta_q L^q)$ are the usual autoregressive and moving average polynomials, ε_t is a mean-zero martingale difference sequence, and $(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)L^j}{\Gamma(-d)\Gamma(j+1)}$ with $\Gamma(\cdot)$ the gamma function. The $ARFIMA(p,d,q)$ process $\{x_t\}$ has stationary long memory for $d \in (0, 0.5)$. This process class elegantly extends the classic model framework of Box and Jenkins (1976) by introducing the idea of fractionally differencing a time series to arrive at a stationary short memory series.

In the following years numerous applications have been found in Natural science, Finance and Economics. Concerning the latter two: Exchange rates, forward premia, aggregate output and inflation rates (e.g. Cheung (1993), Baillie and Bollerslev (1994), Diebold and Rudebusch (1989) as well as Hassler and Wolters (1995) and Baillie et al. (1996), respectively) serve as prominent examples, alongside the literature strand of volatility forecasting (e.g. Deo et al. (2006), Martens et al. (2009) and Chiriac and Voev (2011)). Complementary, numerous publications where brought forth evolving statistical inference in the presence of long memory time series as well as on long memory properties. Beran et al. (2013) and Giraitis et al. (2012) provide a comprehensive overview of the existing methodology.

Nevertheless, long memory time series in general impose important differences concerning the asymptotic properties of subsequent statistical inference compared to the short memory case. The sole presence of long memory in the data or statistical inference on specific long memory features, especially under locally stationary processes (e.g. Cavaliere et al. (2015b) and Demetrescu and Sibbertsen (2016)), require the development of robust statistics. This thesis contains three essays on robust long memory inference that address these issues. The first essay in Chapter 2 deals with robust statistical inference on the sample mean of a long memory process in the context of the Diebold and Mariano (1995) test. In the second essay in Chapter 3 we consider robust statistical inference on long memory in the presence of locally stationary processes via a nonparametric time-domain based Lagrange Multiplier test. Finally, the third essay in Chapter 4 deals with robust statistical inference on spurious long memory and criticizes the existing paradigm of distinguishing true from spurious long memory solely based on a spectral rate condition.

As is well known the popular test of Diebold and Mariano (1995) (DM) for equal predictive accuracy of two competing forecasts, constructs a forecast error loss differential which is assumed to follow a weakly stationary linear process. The procedure then tests for significance of the sample mean of this series using a simple t-test where the usual transformation stabilizing rate of order $O\left(\sqrt{T}\right)$ holds. In the first essay in Chapter 2 we extend the DM test to situations when the forecast error loss differential exhibits long memory. In this situation the rate condition changes to $O\left(T^{1/2-d}\right)$ with $0 < d < 1/2$ and an appropriate long-run variance estimator for correct studentization of the statistic is required since the autocorrelation function is no longer absolutely summable (e.g. Beran et al. (2013)). It is shown theoretically that this scenario can arise frequently, since long memory can be transmitted from forecasts and the forecast objective to the forecast error loss differential. We show that the nature of this transmission mainly depends on the (un)biasedness of the forecasts and whether the involved series share (un)balanced common long memory. Further theoretical and simulation results show that the conventional DM test is invalidated under these circumstances.
Therefore, robust statistics based on a memory and autocorrelation consistent (MAC) estimator by Robinson (2005) and an extended fixed-bandwidth (fixed-b) approach of McElroy and Politis (2012) are considered. The subsequent Monte Carlo study provides a novel comparison of these robust statistics and addresses several practitioners questions like choice of kernel, bandwidth selection and choice of plug-in estimator. Inter alia we find that the MAC approach delivers superior power results whereas the extended fixed-b approach, especially when employing the modified quadratic spectral kernel, offers best size control.
As empirical applications, we conduct forecast comparison tests for the realized volatility of the Standard and Poors 500 index among recent extensions of the heterogeneous autoregressive model (see Corsi (2009), Corsi et al. (2010), and Corsi and Renò (2012)). While 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 in most situations. Long memory is found in the forecast error loss differentials in the majority of the considered scenarios and our results also remain fairly robust under the QLIKE loss function.

In recent years a growing literature found empirical evidence that macroeconomic and financial time series exhibit various types of unconditional heteroskedasticity (see Loretan and Phillips (1994), McConnell and Perez-Quiros (2000), Sensier and Van Dijk (2004), Stărică and Granger (2005) and Cavaliere et al. (2015b), among others). Complementary, Cavaliere et al. (2015b) show that heteroskedasticity of a quite general form leads to non-pivotal asymptotic null distributions of Lagrange Multiplier tests for the order of fractional integration in an $ARFIMA(p,d,q)$ model. Similar results are found in Kew and Harris (2009) and Demetrescu and Sibbertsen (2016), where the latter study periodogram-based inference.
Therefore, in the second essay in Chapter 3 we propose a nonparametric time-domain based Lagrange Multiplier test for long memory with robustness against fixed or switching short-run dynamics and heteroskedasticity of quite general form. Our test is based on the procedure of Harris et al. (2008), utilizing a nonparametric short memory correction, which is subsequently recast in a regression framework in the style of Breitung and Hassler (2002) and standardized using a heteroskedasticity and autocorrelation consistent estimator of Andrews (1991).
An extensive Monte Carlo simulation study shows that its limiting null distribution is standard normal and despite of its several robustness features, we nevertheless obtain good power results that clearly indicate consistency of our test. Moreover, a comparison with other existing time-domain based Lagrange Multiplier testing procedures in the literature shows that it is the only test capable of controlling its size under the heteroskedastic locally stationary process class (c.f. Dahlhaus (2000), Palma and Olea (2010), and Cavaliere et al. (2015b)) considered here.
In our empirical application we find that long memory remains in different types of ex-post Variance Risk Premium series, represented by a fractional cointegration relationship between the Chicago Board of Options Exchange Volatility Index and realized variance of the Standard and Poors 500 index, in addition to conditional and unconditional heteroskedasticity.

The long memory literature often faces the criticism that a short memory series, contaminated by rare level shifts, displays features like hyperbolically decaying sample autocorrelations and a steep slope of the periodogram for Fourier frequencies near the origin, which are commonly attributed to true long memory (e.g. Lobato and Robinson (1998), Diebold and Inoue (2001), Granger and Hyung (2004) and Perron and Qu (2010)). To address this problem the Qu (2011) test on spurious long memory distinguishes these two model classes based on the differing rate conditions at which the periodogram tapers of at Fourier frequencies near the origin which is much steeper in case of spurious long memory compared to the true long memory case. Moreover, this tests benefits from weak assumptions and superior finite sample performance compared to other existing spurious long memory tests in the literature (c.f. Leccadito et al. (2015)).
In the third essay in Chapter 4 we show by means of simulation study that the Qu (2011) test has substantial size distortions when applied to a recently proposed class of locally stationary processes (see Cavaliere et al. (2015b) and Demetrescu and Sibbertsen (2016)) and random coefficient models (c.f. Giraitis et al. (2014)), although these models obey a semiparametric representation commonly attributed to true negative, short or long memory.

A large scale Monte Carlo simulation study reveals that the induced liberal bias easily leads to a potential spurious rejection of the null and also remains non-negligible in very large samples. This demonstrates that not only processes displaying low-frequency contaminations can be confused with true long memory, but that under some weaker assumptions, true negative, short or long memory processes can be wrongly attributed to the spurious long memory class. Furthermore, we propose a simple prewhitening procedure, which builds on results of Qu (2011), Cavaliere et al. (2017) as well as the algorithm of Bai and Perron (1998, 2003a), that recreates good size results and works well even in small samples.

As empirical applications we consider on one hand log spot prices, corresponding one-period futures contract prices, as well as spreads of gold, silver, platinum, and crude oil and on the other hand monthly (seasonally adjusted) inflation rates of G7 countries. We find that in particular the former series are subject to a spurious rejection of the null hypothesis, especially when using recommended user-specific bandwidth choices.

Chapter 2

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 Christian Leschinski.

Under revision for the Journal of Financial Econometrics.

2.1 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 $g(y_t, \widehat{y}_{it}) \geq 0$ for $i = 1, 2$. The forecast error loss differential is then denoted by $z_t = g(y_t, \widehat{y}_{1t}) - g(y_t, \widehat{y}_{2t})$.

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 Li and Patton (2015)). 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 (both theoretically and via simulations) 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 for similar inference problems (besides the Diebold-Mariano test) 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 2.2 reviews the classic Diebold-Mariano test and presents the fixed-b approach for the short memory case. Section 2.3 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 2.4. Section 2.5 contains our Monte Carlo study and in Section 2.6 we present our empirical results. Conclusions are drawn in Section 2.7. All proofs are contained in the appendix.

2.2 Diebold-Mariano Test

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

$$z_t = \mu_z + \sum_{j=0}^{\infty} \theta_j v_{t-j}\,, \tag{2.1}$$

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

$$t_{DM} = T^{\delta}\frac{\bar{z}}{\sqrt{\widehat{V}}}\,. \tag{2.2}$$

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 2.3.

2.2.1 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,

$$\widehat{V}_{HAC} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{B}\right)\widehat{\gamma}_z(j)\,, \tag{2.3}$$

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 corresponding Diebold-Mariano statistic is given by

$$t_{HAC} = T^{1/2}\frac{\bar{z}}{\sqrt{\widehat{V}_{HAC}}}\,. \tag{2.4}$$

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 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 (2.2) 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.

2.2.2 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

$$t_{FB} = T^{1/2}\frac{\bar{z}}{\sqrt{\widehat{V}(k,b)}} \tag{2.5}$$

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) and Li and Patton (2015), 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.

2.3 Long Memory in Forecast Error Loss Differentials

2.3.1 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 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 2.6 (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 2.1. *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 2.1 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 2.1. 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 2.4. We therefore adopt Definition 2.1 which covers fractional integration as a special case. A similar approach is taken by Dittmann and Granger (2002).[1]

Given Definition 2.1, we now state some assumptions regarding the long memory structure of the forecast objective and the forecasts.

Assumption 2.1 (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 2.1) 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 2.1, 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 2.2 (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 2.2 (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 (2017) 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 2.1 below, for convenience.

Proposition 2.1 (Memory of Products)**.** *Let a_t and b_t be long memory series according to Definition 2.1 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 2.1 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 2.1 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 2.2 and 2.5 for details).

As discussed in Leschinski (2017), Proposition 2.1 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 2.1 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 2.1, 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 $g(y_t, \widehat{y}_{it}) = (y_t - \widehat{y}_{it})^2$ for $i = 1, 2$. The quadratic forecast error loss differential is then given by

$$z_t = (y_t - \widehat{y}_{1t})^2 - (y_t - \widehat{y}_{2t})^2 = \widehat{y}_{1t}^2 - \widehat{y}_{2t}^2 - 2y_t(\widehat{y}_{1t} - \widehat{y}_{2t})\,. \tag{2.6}$$

Note that even though the forecast objective y_t as well as the forecasts $\widehat{y}_{it}$ in (2.6), have time index t, the representation is quite versatile. It allows for forecasts generated from time series models where $\widehat{y}_{it} = \sum_{s=1}^{t-1} \phi_s y_{t-s}$ 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 (2.6), 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.

2.3.2 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 (2.6) in Proposition 2.2. 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 2.2 (Memory Transmission without CLM). *Under Assumptions 2.1 and 2.2, the forecast error loss differential in (2.6) is* $z_t \sim LM(d_z)$*, where*

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

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 2.1 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

$$z_t = \widehat{y}_{1t}^{*2} - \widehat{y}_{2t}^{*2} - 2\left[y_t^*(\mu_1 - \mu_2) + \widehat{y}_{1t}^*(\mu_y - \mu_1) + \widehat{y}_{2t}^*(\mu_y - \mu_2)\right] - 2\left[y_t^*(\widehat{y}_{1t}^* - \widehat{y}_{2t}^*)\right] + const\,. \tag{2.7}$$

From (2.7) 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 2.1, from which the zero mean product series $y_t^*\widehat{y}_{it}^*$ is $LM(\max\{d_y + d_i - 1/2,\ 0\})$ 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 (2.7). Proposition 2.2 then follows from a case-by-case analysis.

Proposition 2.2 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 (2.7). 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.

2.3.3 Memory Transmission under Common Long Memory

The results in Proposition 2.2 are based on Assumption 2.2 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 2.3, below.

Assumption 2.3 (Common Long Memory). *The causal Gaussian process x_t has long memory according to Definition 2.1 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 $\xi_y, \xi_i \neq 0$. Both, η_t and ε_{it} are mean zero causal Gaussian long memory processes with parameters d_η and d_{ε_i} fulfilling $1/2 > d_x > d_\eta, d_{\varepsilon_i} \geq 0$, for $i = 1, 2$.*

Assumption 2.3 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 2.2 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 2.3 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 three important special cases: (i) the forecasts are biased and the ξ_a differ from each other, (ii) the forecasts are biased, but the ξ_a are equal, and (iii) the forecasts are unbiased and $\xi_a = \xi_b$ 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 in the series that are subject to CLM, 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 2.3 that

$$y_t - \widehat{y}_{it} = \beta_y - \beta_i + x_t(\xi_y - \xi_i) + \eta_t - \varepsilon_{it} , \tag{2.8}$$

so that $x_t(\xi_y - \xi_i)$ does not disappear from the linear combination if the scaling parameters ξ_y and ξ_i are different from each other.

We refer to a situation where the $\xi_a = \xi_b$ as “balanced CLM”, whereas CLM with $\xi_a \neq \xi_b$ is referred to as “unbalanced”.

In situation (i) both forecasts are biased and the presence of CLM does not lead to a cancellation of the memory of x_t in the loss differential. Of course this is an extreme case, but it serves to illuminate the mechanisms at work - especially in contrast to the results in Propositions 2.4 and 2.5, below. By substituting the linear relations from Assumption 2.3 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 (2.7). Since analogous terms to those in the first bracket of (2.7) appear in each case, it is possible to focus on the transmission of memory from the forecasts and the forecast objective to the loss differential. We therefore obtain the following result.

Proposition 2.3 (Memory Transmission with Biased Forecasts and Unbalanced CLM)**.** *Let $\xi_i \neq \xi_y$, $\xi_1 \neq \xi_2$, $\mu_i \neq \mu_y$, and $\mu_1 \neq \mu_2$, for $i = 1,2$. Then under Assumptions 2.1 and 2.3, the forecast error loss differential in (2.6) is $z_t \sim LM(d_z)$, where*

$$d_z = \begin{cases} \max\left\{d_y, d_x\right\}, & \text{if } \widehat{y}_{1_t}, \widehat{y}_{2t} \sim CLM(d_x, d_x - b), \text{ except if } \xi_1/\xi_2 = (\mu_y - \mu_2)/(\mu_y - \mu_1) \\ \max\left\{d_2, d_x\right\}, & \text{if } \widehat{y}_{1t}, y_t \sim CLM(d_x, d_x - b), \text{ except if } \xi_1/\xi_y = -(\mu_1 - \mu_2)/(\mu_y - \mu_1) \\ \max\left\{d_1, d_x\right\}, & \text{if } \widehat{y}_{2t}, y_t \sim CLM(d_x, d_x - b), \text{ except if } \xi_2/\xi_y = -(\mu_1 - \mu_2)/(\mu_y - \mu_2) \\ d_x, & \text{if } \widehat{y}_{1t}, \widehat{y}_{2t}, y_t \sim CLM(d_x, d_x - b), \\ & \text{except if } \xi_1(\mu_y - \mu_1) + \xi_y(\mu_1 - \mu_2) = \xi_2(\mu_y - \mu_2)\,. \end{cases}$$

Proof: *See the Appendix.*

In absence of common long memory we observed in Proposition 2.2, that the memory is $\max\left\{d_1, d_2, d_y\right\}$ if the means differ from each other. Now, if two of the series share common long memory, they both have memory of d_x. With this in mind, Proposition 2.3 shows that the transmission mechanism is essentially unchanged and the memory of the loss differential is still dominated by the largest memory parameter. The only exception to this rule is if - by coincidence - the differences in the means and the memory parameters offset each other.

Similar to (i), case (ii) refers to a situation of biasedness, but now with balanced CLM, so that the underlying long memory factor x_t cancels out in the forecast error loss differentials. The memory transmission can than be characterized by the following proposition.

Proposition 2.4 (Memory Transmission with Biased Forecasts and Balanced CLM)**.** *Let $\xi_1 = \xi_2 = \xi_y$. Then under Assumptions 2.1 and 2.3, the forecast error loss differential in (2.6) is $z_t \sim LM(d_z)$, where*

$$d_z = \begin{cases} \max\left\{d_y, d_x\right\}, & \text{if } \widehat{y}_{1_t}, \widehat{y}_{2t} \sim CLM(d_x, d_x - b), \text{ and } \mu_1 \neq \mu_2 \\ \max\{d_2, d_x\}, & \text{if } \widehat{y}_{1t}, y_t \sim CLM(d_x, d_x - b), \text{ and } \mu_y \neq \mu_2 \\ \max\{d_1, d_x\}, & \text{if } \widehat{y}_{2t}, y_t \sim CLM(d_x, d_x - b), \text{ and } \mu_y \neq \mu_1 \\ \tilde{d}, & \text{if } \widehat{y}_{1t}, \widehat{y}_{2t}, y_t \sim CLM(d_x, d_x - b)\ , \end{cases}$$

for some $0 \leq \tilde{d} < d_x$.

Proof: *See the Appendix.*

We refer to the first three cases in Propositions 2.3 and 2.4 as partial CLM, because there is always one of the $\widehat{y}_{it}$ or y_t that is not part of the CLM relationship and the fourth case as full CLM. We can observe that the dominance of the memory of the most persistent series under partial CLM is preserved for both balanced and unbalanced CLM. We therefore conclude that this effect is generated by the interaction with the series that is not involved in the CLM relationship. This can also be seen from equations (2.18) to (2.20) in the proof.
Only in the fourth case with full CLM the memory transmission changes between Propositions 2.3 and 2.4. In this case the memory in the loss differential is reduced to $d_z \leq d_x$.
The third special case (iii) refers to a situation of unbiasedness similar to the last case in Proposition 2.2. In addition to that, it is assumed that there is balanced CLM as in Proposition 2.4, where $\xi_a = \xi_b$, if a_t and b_t are in a common long memory relationship. Compared to the setting of the previous Propositions this is the most ideal situation in terms of forecast quality. Here, we have the following result.

Proposition 2.5 (Memory Transmission with Unbiased Forecasts and Balanced CLM)**.** *Under Assumptions 2.1 and 2.3, and if $\mu_y = \mu_1 = \mu_2$ and $\xi_y = \xi_a = \xi_b$, then $z_t \sim LM(d_z)$, with*

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

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

Proof: *See the Appendix.*

Proposition 2.5 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 2.2 - 2.5 show that long memory can be transmitted from forecasts or the forecast objective to the forecast error loss differentials. Our results also show that the biasedness of the forecasts plays an important role for the transmission of dependence to the loss differentials.

To get further insights into the mechanisms found in Propositions 2.2 - 2.5 consider a situation where two forecasts with different non-zero biases are compared. In absence of CLM, it is obvious from Proposition 2.2 that the memory of the loss differential will be determined by the maximum of the memory orders of the forecasts and the forecast objective. If one of the forecasts has common long memory with the objective, the same holds true - irrespective of the loadings ξ_a on the common factor. As can be seen from Proposition 2.3, even if both forecasts have CLM with the objective, the maximal memory order is transmitted to z_t if the factor loadings ξ_a differ. Only if the factor loadings are equal, the memory will be reduced as stated Proposition 2.4.

If we consider two forecasts that are unbiased in absence of CLM, it can be seen from Proposition 2.2 that the memory of the loss differential is lower than that of the original series. The same holds true in presence of CLM, as covered by Proposition 2.5.

In practical situations, it might be overly restrictive to impose exact unbiasedness (under which memory would be reduced according to Proposition 2.5). 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 2.6).

Biases can also be caused by estimation error. 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.

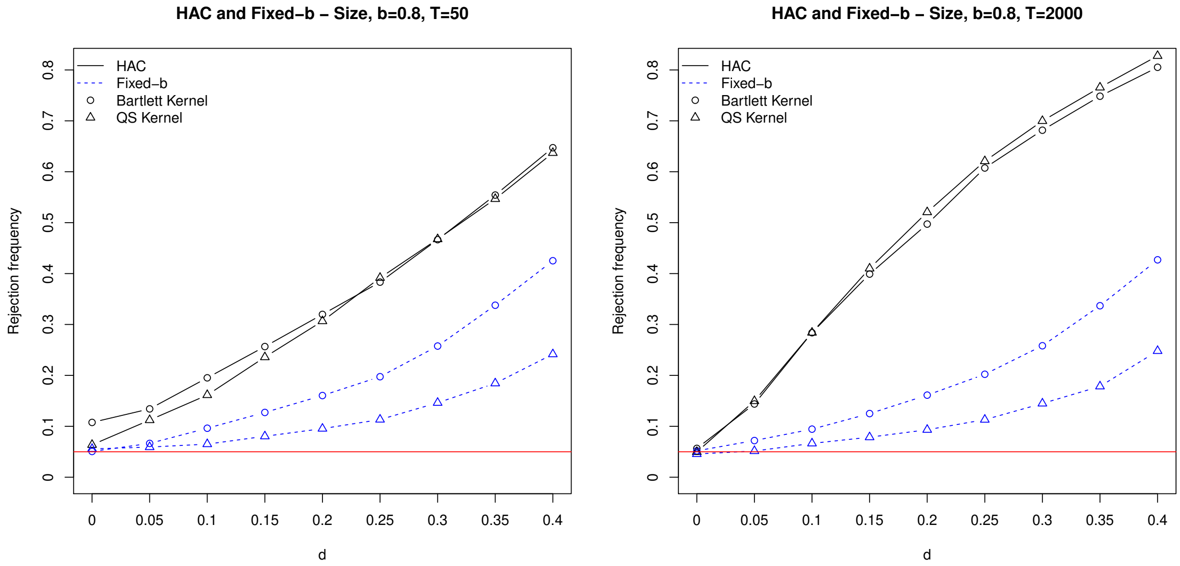


Figure 2.1: Size of the t_{HAC}- and t_{FB}-tests with $T \in \{50, 2000\}$ for different values of the memory parameter d.

2.3.4 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 2.6 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 perform equally in terms of $g(\cdot)$.

Proposition 2.6 (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 2.1 are obtained with $M = 5000$ Monte Carlo repetitions. We simulate samples of $T = 50$ and $T = 2000$ 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]
We start by commenting on the results for the small sample size of $T = 50$ in the left panel of Figure 2.1. As demonstrated by Kiefer and Vogelsang (2005), the fixed-b approach works exceptionally well for the short memory 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. Continuing with the results for $T = 2000$ in the right panel of Figure 2.1, we observe similar findings in general. We note that for the short memory case, size distortions arising from a too small sample size vanish. All tests statistics are well behaved for $d = 0$. On the contrary, size distortions are stronger with an increasing sample size, although the magnitude of additional distortion is moderate. This feature can be attributed to the slow divergence rate (as given in the proof of Proposition 2.6) of the test statistic under long memory.

2.4 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.

2.4.1 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.

[2] The bandwidth parameter of the fixed-b estimator is set to $b = 0.8$, since using a larger fraction of the autocorrelations provides a higher emphasis on size control (c.f. Kiefer and Vogelsang (2005)). Other bandwidth choices lead to similar results.

Robinson (2005) assumes that z_t is linear (in the sense of our equation 2.1, see also Assumption L in Abadir et al. (2009)) 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 $b_0 = \lim_{\lambda \to 0} |\lambda|^{2d} f(\lambda)$.[3] Among others, this assumption covers stationary and invertible ARFIMA processes. A key result for the MAC estimator is that as $T \to \infty$:

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

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. This estimator is given by:

$$\widehat{d} = \arg \min_{d \in (-1/2, 1/2)} U_T(d),$$

where $U_T(d) = \log\left(\frac{1}{m_d}\sum_{j=1}^{m_d} j^{2d} I_T(\lambda_j)\right) - \frac{2d}{m_d}\sum_{j=1}^{m_d} \log j$ (see Robinson (1995a)). Many other estimation approaches (e.g. log-periodogram estimation, etc.) would be a possibility as well.
Next, b_0 can be estimated consistently by:

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

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

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.

2.4.2 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]

2.5 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 2.3 are based on assumptions on the forecasts and the forecast objective, the modified DM tests proposed in Section 2.4 are based

[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 (2.3). Further kernels, including flat-top tapers, are analyzed as well but yield slightly inferior results to those reported here.

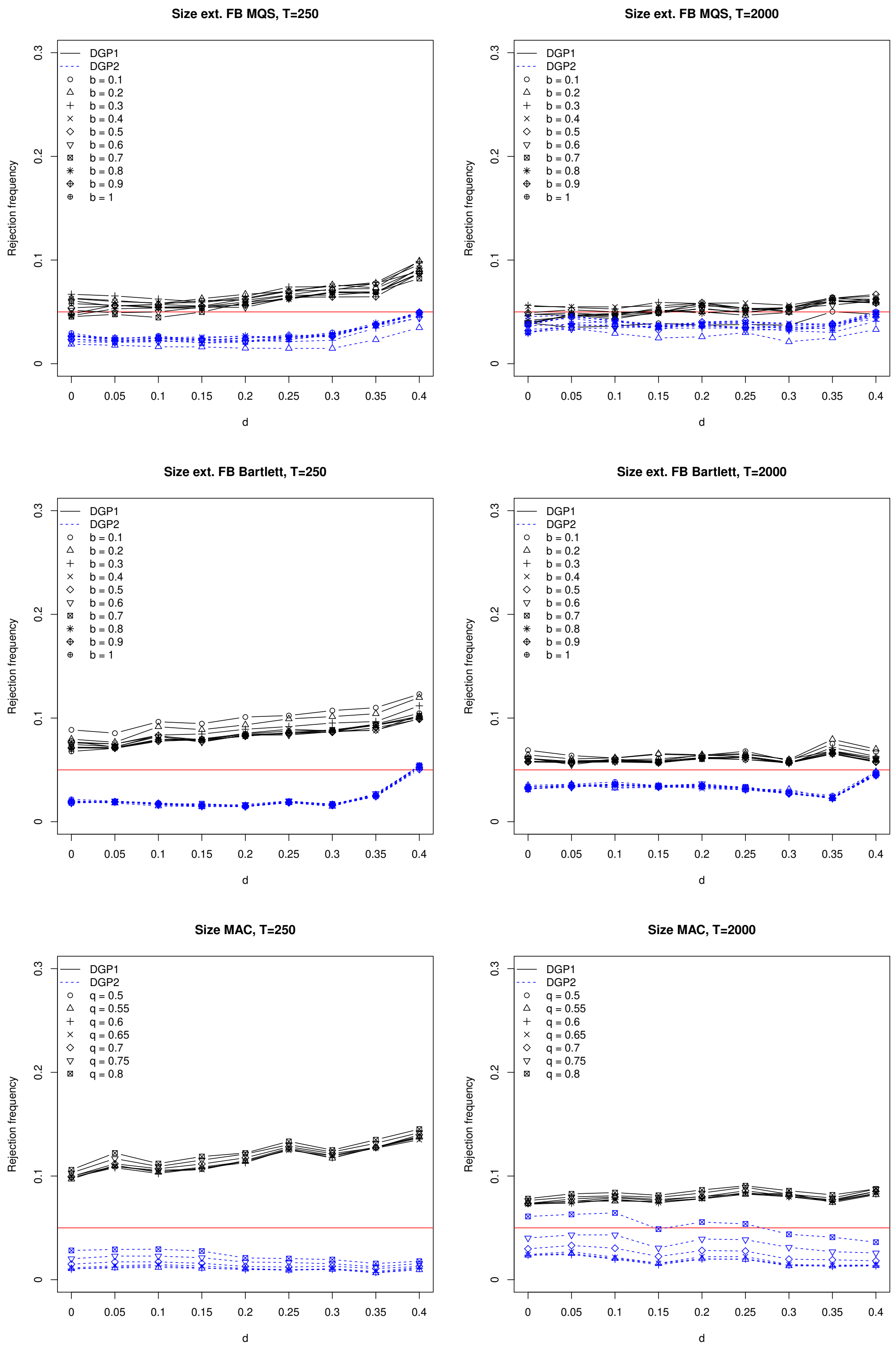


Figure 2.2: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{250, 2000\}$ and bandwidth parameters q and b.

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 $d = \{0, 0.05, 0.1, ..., 0.4\}$, 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.
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 first-order MA dynamics (with MA parameter $\theta = 0.6$) are available in our supplementary appendix (Subsection 2.8.1). In addition to that, the important special case of optimal one-step-ahead forecasts is represented by DGP1 for $d = 0$.
Figure 2.2 shows a comparison of the size of both tests for different degrees of long memory and sample sizes of $T \in \{250, 2000\}$.[6] In the case of DGP1 (black solid lines), 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 $T = 2000$ 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 (blue dashed lines) 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 samples of $T = 250$ observations both

[6] Additional simulation results for $T \in \{50, 1000\}$ are reported in our supplementary appendix (Subsection 2.8.1).

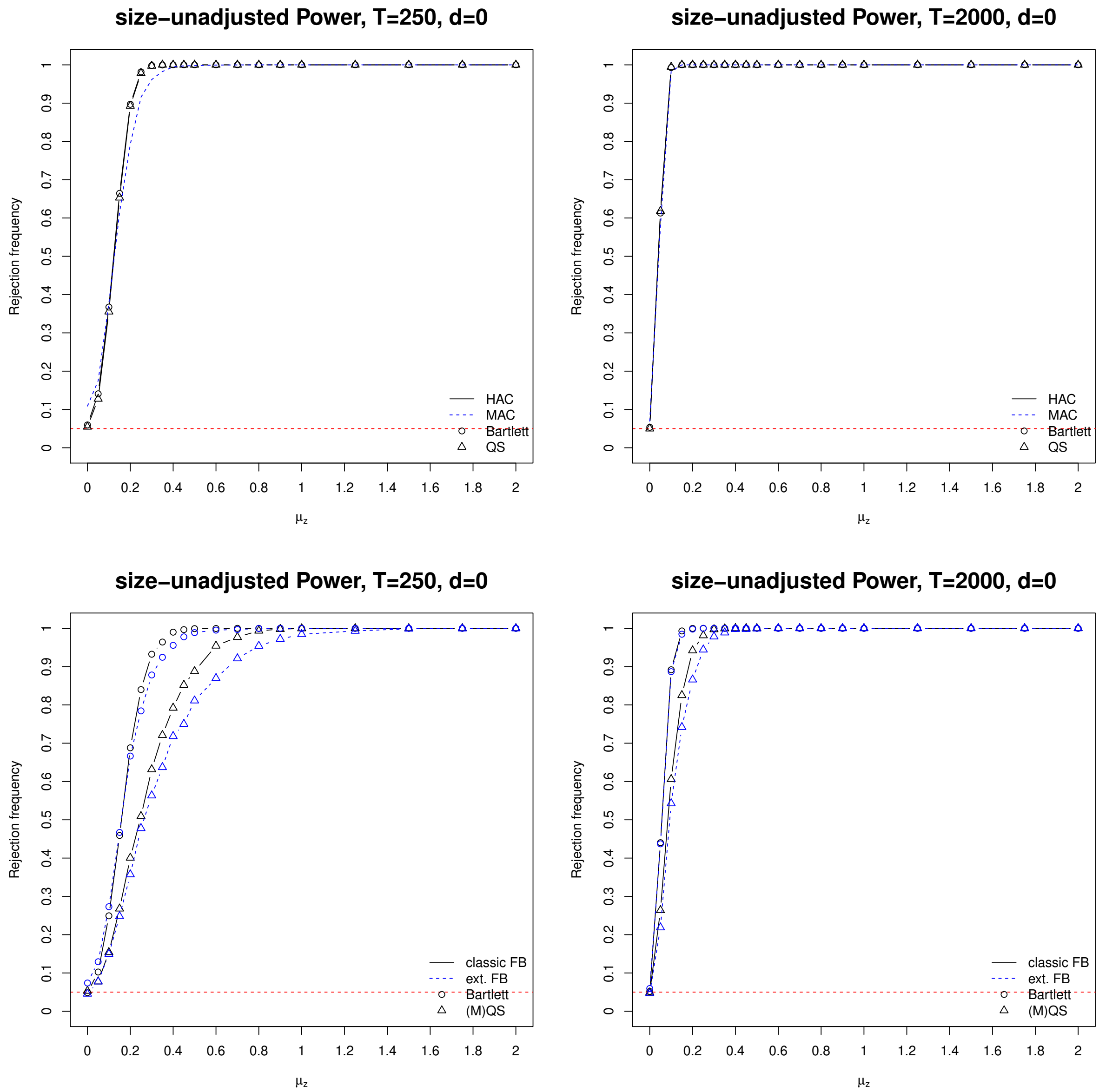


Figure 2.3: Power comparison of the robust statistics t_{EFB} and t_{MAC} with their short memory counterparts when $d = 0$.

tests are conservative. For large values of d, we can observe that the downward size distortions of the t_{EFB}-test vanish. In large samples, all tests show better empirical size properties, as expected. Further simulations considering moving average components are conducted and results are reported in our supplementary appendix (Subsection 2.8.1). 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.

Our simulation study suggests that a 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

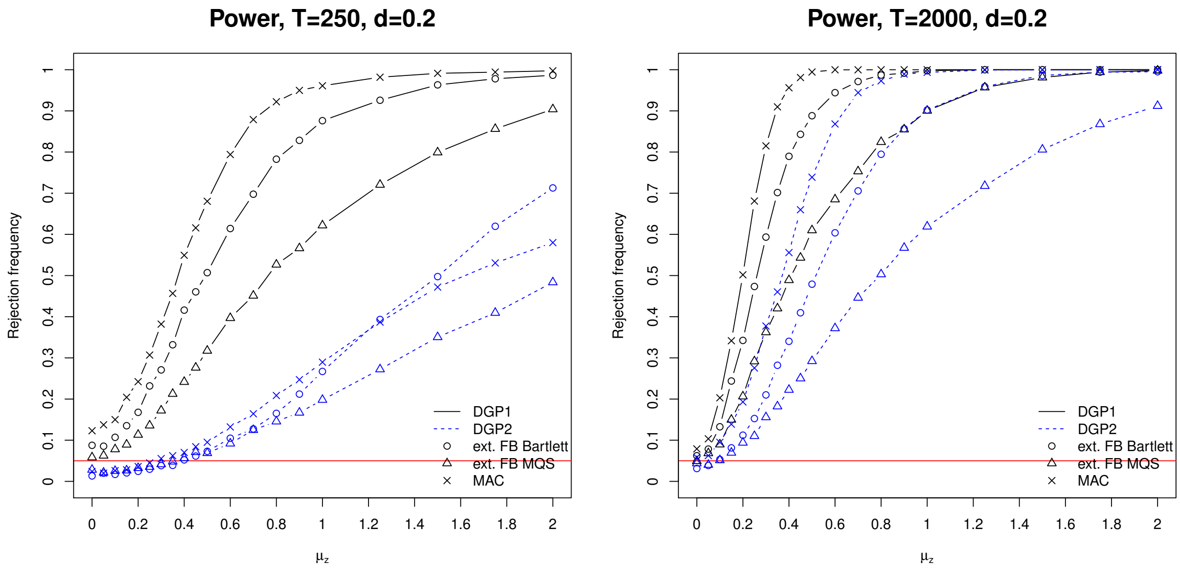


Figure 2.4: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample sizes $T \in \{250, 2000\}$.

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 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 is short memory ($d = 0$). With regard to the previous results, we choose 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). Results are presented in Figure 2.3. 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 2.3 with $T \in \{250, 2000\}$ and DGP1, setting $d = 0$. Although some power loss can be observed as expected, the cost of using the long memory robust procedures is more than acceptable, already 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 (c.f. Section 2.6) closely. We choose the same bandwidth parameters as before to enhance comparability. 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 (μ_z) is added and the respective test is applied. The results are shown in Figure 2.4. 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 2.16 (and 2.23, for MA dynamics) in our supplementary appendix (Subsection 2.8.1). 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.

By comparing the results for DGP1 with those of DGP2 in Figure 2.4, 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.

For robustification of 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 2.12 and 2.13 (and 2.20 and 2.21, for MA dynamics) in our supplementary appendix (Subsection 2.8.1) shows the results of this exercise. In small samples, the size obtained using the ALPW estimator becomes similarly liberal for all procedures and all DGPs. In larger samples of $T = 250$ and beyond, 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 has the highest power for sample sizes of $T \in \{50, 250\}$. For larger samples, however, the former ranking is reestablished suggesting a small sample effect.

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 which might be attributed to the increased variance of the estimator, partly also due to the automatic bandwidth selection approach.

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. In our empirical applications to realized volatility in the next section we consider such comparisons. In general, our conclusions also apply to other inference problems involving the sample mean.

2.6 Applications to Realized Volatility Forecasting

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 (c.f., 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 empirical applications 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 2.4. 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. In total, we have $T = 4883$ observations in our sample. The raw data is obtained from the Thomson Reuters Tick History Database.

Before we turn to the forecast evaluations in Sections 2.6.1 and 2.6.2, 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. Similar to the notation in Bekaert and Hoerova (2014), the model reads

$$\ln RV_t^{(h)} = \alpha + \rho_{22} \ln RV_{t-h}^{(22)} + \rho_5 \ln RV_{t-h}^{(5)} + \rho_1 \ln RV_{t-h}^{(1)} + \varepsilon_t , \tag{2.9}$$

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.[7] 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).

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

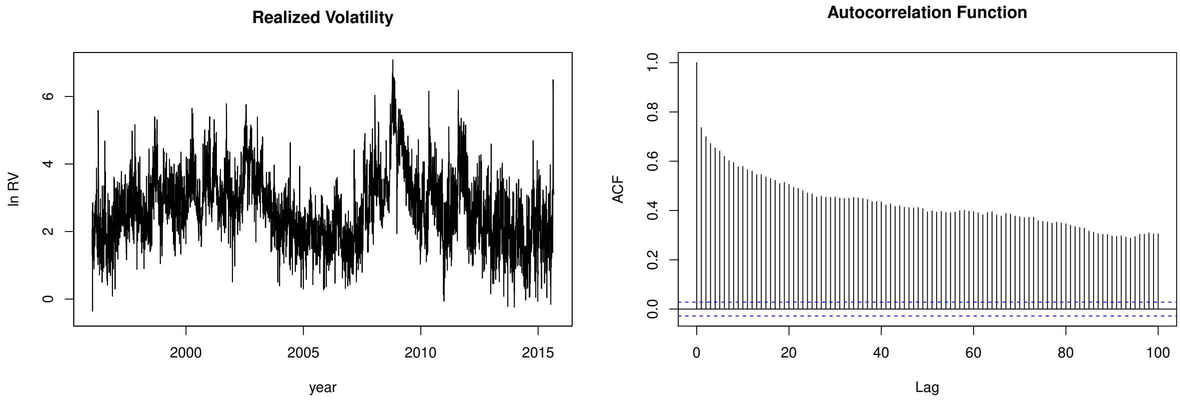


Figure 2.5: Daily log-realized volatility of the S&P500 index and their autocorrelation function.

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):

$$\begin{aligned}\ln RV_t^{(h)} &= \alpha + \rho_{22} \ln C_{t-h}^{(22)} + \rho_5 \ln C_{t-h}^{(5)} + \rho_1 \ln C_{t-h}^{(1)} \\ &\quad + \varpi_{22} \ln\left(1 + J_{t-h}^{(22)}\right) + \varpi_5 \ln\left(1 + J_{t-h}^{(5)}\right) + \varpi_1 \ln\left(1 + J_{t-h}^{(1)}\right) + \varepsilon_t\,. \end{aligned} \tag{2.10}$$

The daily log-realized variance series ($\ln RV_t$) is depicted in Figure 2.5.[8] 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 2.5, 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 2.1. 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

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

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 2.1: 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).

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 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 2.2 - 2.5 can occur.

2.6.1 Predictive Ability of the VIX for Quadratic Variation

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 2.6. 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 2.6. 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

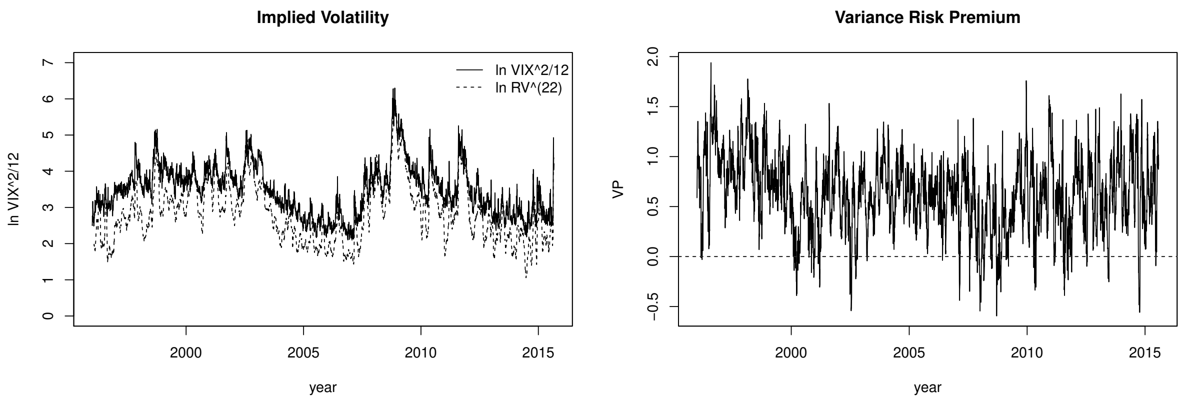


Figure 2.6: Log squared implied volatility and log cumulative realized volatility of the S&P 500 (left panel) and variance risk premium $VP_t = \ln(VIX_t^2/12) - \ln RV_{t+22}^{(22)}$ (right panel).

including Christensen and Nielsen (2006), Nielsen (2007) 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) 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.[9] 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.

[9] 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).

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	2.754	2.985	2.849
												(3.404)	(4.064)	(4.750)	(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	2.503	2.889	2.724
												(2.610)	(3.154)	(3.693)	(4.228)
HAR-RV-TCJ-L	0.082	0.282	0.269	**0.182***	0.163	**1.784**	**1.786**	1.819	0.889	1.016	1.192	1.819	2.153	2.430	2.317
						(1.645)	(1.645)	(2.092)		(1.645)		(3.404)	(4.064)	(4.750)	(5.388)

Table 2.2: Predictive ability of the VIX for future RV (evaluated under MSE loss). 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 are given. 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.

For details on the HAR-RV-TCJ-L model, see Corsi and Renò (2012) and equation (2) in Bekaert and Hoerova (2014).

Table 2.2 reports the results. Models are estimated using a rolling window of $T_w = 1000$ observations.[10] This implies that the forecast window contains 3883 observations.[11] 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 2.2 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 2.2 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 2.6. In regard of the persistence in the loss differential series the improvements are too small to be considered significant.

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

[11] Additional simulation results for a sample size of 4000 observations are available in our supplementary appendix (Subsection 2.8.1). The results are generally in line with those for 2000 observations.

Models	Summary statistics					Short memory inference			Long memory inference						
									t_{MAC}			t_{EFB}			
Model vs. Model+VIX	$\bar{z}/\widehat{\sigma}_z$	$QLIKE1$	$QLIKE2$	$\widehat{d}_{LW}$	$\widehat{d}_{LPWN}$	t_{DM}	t_{HAC}	t_{FB}	0.7	0.75	0.8	0.2	0.4	0.6	0.8
HAR-RV	0.152	2.025	2.023	**0.234***	**0.203**	**3.071**	**3.251**	**2.758**	1.299	1.453	**1.657**	2.758	3.024	3.161	3.068
												(3.404)	(4.064)	(4.750)	(5.388)
HAR-RV-TCJ	0.129	2.025	2.023	**0.188***	0.133	**2.720**	**2.806**	**2.833**	**1.720**	**1.971**	**2.288**	**2.833**	**3.236**	3.446	3.367
												(2.610)	(3.154)	(3.693)	(4.228)
HAR-RV-TCJ-L	0.101	2.024	2.023	**0.186***	0.103	**2.065**	**2.053**	**2.517**	1.574	**1.837**	**2.167**	2.517	2.869	3.073	3.029
						(1.645)	(1.645)	(2.092)		(1.645)		(2.610)	(3.154)	(3.693)	(4.228)

Table 2.3: Predictive ability of the VIX for future RV (evaluated under QLIKE loss). Models excluding the VIX are tested against models including the VIX. See the notes for Table 2.2.

These findings highlight the importance of long memory robust tests for forecast comparisons in practice. As a comparison, we also consider the QLIKE loss function

$$g(y_t, \widehat{y}_t) = \log(\widehat{y}_t) + \frac{y_t}{\widehat{y}_t},$$

in addition to the MSE. The motivation for this is that realized volatility is generally considered to be an unbiased, but perturbed proxy of the underlying latent volatility process. It is shown by Patton (2011) that among the commonly employed loss functions only MSE and QLIKE preserve the true ranking of competing forecasts when being evaluated on a perturbed proxy. We therefore also consider QLIKE, even though our theoretical results in Section 2.3 do not apply to this loss function.

Results are reported in Table 2.3. They suggest that the average standardized forecast error loss differentials are positive and similar in magnitude to the MSE comparison in Table 2.2. Moreover, they have a similar memory structure. From this descriptive viewpoint, results are not sensitive to the choice between the QLIKE and the MSE loss function. When using short memory inference, the null hypothesis of pairwise equal predictive ability amongst the models is rejected in all cases. This is also in line with the previous results.

Turning to long memory-robust statistics, we find somewhat different results, especially for the symmetric HAR-RV-TCJ model including jumps. Here, we mostly observe rejections of equal predictive ability in favor of the inclusion of the VIX. This is likely to be due to the asymmetry of the QLIKE loss function. For other versions of the HAR model, the evidence against the null hypothesis is weaker and thus mainly in line with our previous results.

Models	Summary statistics					Short memory inference			Long memory inference						
									t_{MAC}			t_{EFB}			
HAR-RV vs.	$\overline{z}/\widehat{\sigma}_z$	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**	**4.068**	**4.468**	**4.947**
												(2.610)	(3.154)	(3.693)	(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**	**3.981**	**5.093**	**5.848**
												(2.050)	(2.522)	(2.975)	(3.386)
HAR-RV-TCJ, $h = 22$	0.045	0.292	0.285	**0.359***	**0.343***	0.776	0.912	0.666	0.140	0.152	0.171	0.666	0.925	1.064	1.164
						(1.645)	(1.645)	(2.092)		(1.645)		(4.701)	(5.551)	(6.413)	(7.281)

Table 2.4: Separation of Continuous and Jump Components (evaluated under MSE loss). Reported are the standardized mean ($\overline{z}/\widehat{\sigma}_z$) 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 are given. 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.

2.6.2 Separation of Continuous Components and Jump Components

As a second empirical application, we revisit the question whether the HAR-RV-TCJ model from equation (2.10) leads to a significant improvement in forecast performance compared to the standard HAR-RV-model (2.9) from a purely out-of-sample perspective.

The continuous components and jump components - separated using the approach described above - are shown in Figure 2.7. The occurrence of jumps is often associated with macroeconomic 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.

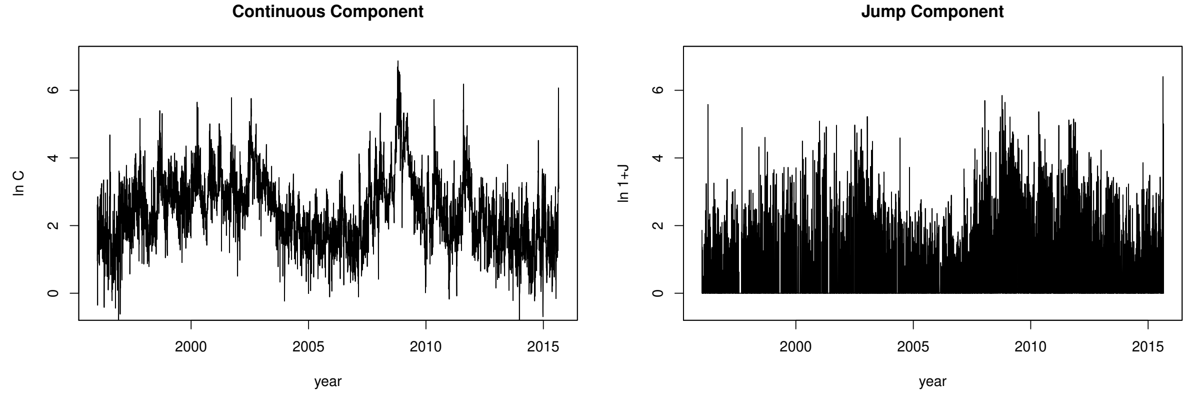


Figure 2.7: Log continuous component $\ln C_t$ and jump component $\ln(1 + J_t)$ of RV_t.

Models	Summary statistics					Short memory inference			Long memory inference						
									t_{MAC}			t_{EFB}			
HAR-RV vs.	$\bar{z}/\widehat{\sigma}_z$	$QLIKE_1$	$QLIKE_2$	$\widehat{d}_{LW}$	$\widehat{d}_{LPWN}$	t_{DM}	t_{HAC}	t_{FB}	0.7	0.75	0.8	0.2	0.4	0.6	0.8
HAR-RV-TCJ, $h=1$	0.066	1.932	1.930	0.044	0.009	**4.409**	**4.080**	**2.834**	**4.193**	**3.944**	**3.823**	**2.834**	**3.012**	**3.368**	**3.731**
												(2.050)	(2.522)	(2.975)	(3.386)
HAR-RV-TCJ, $h=5$	0.035	1.986	1.986	**0.089***	0.016	1.384	1.422	1.288	1.326	1.387	1.517	1.288	1.820	2.577	2.795
												(2.050)	(2.522)	(2.975)	(3.386)
HAR-RV-TCJ, $h=22$	−0.007	2.025	2.025	**0.425***	**0.382***	−0.121	−0.141	−0.113	−0.016	−0.017	−0.019	−0.113	−0.147	−0.161	−0.169
						(1.645)	(1.645)	(2.092)		(1.645)		(7.486)	(8.692)	(9.974)	(11.417)

Table 2.5: Separation of Continuous and Jump Components (evaluated under QLIKE loss). See the notes for Table 2.4.

Table 2.4 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 $\widehat{d}_{LW} = 0.072$ 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 $h = 5$.
When considering the QLIKE loss function as an alternative to the MSE in Table 2.5, we find evidence against the null hypothesis for the case of short-term forecasting, but no for the weekly and monthly horizon. We can therefore confirm that the separation of continuous and jump components indeed improves the forecast performance on daily horizons.

2.7 Conclusion

This paper deals with forecast evaluation under long range dependence. We show in Section 2.3 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 for robustification of DM tests against long memory are discussed in Section 2.4 - 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 absence of short-run dynamics. 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 only find weak 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 at a daily horizon. These examples stress the importance of long memory robust statistics in practice.

2.8 Appendix

Proofs

Proof (Proposition 2.2). *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 (2.6) can be re-expressed as*

$$\begin{aligned}
z_t =& -2y_t(\widehat{y}_{1t} - \widehat{y}_{2t}) + \widehat{y}_{1t}^2 - \widehat{y}_{2t}^2 \\
=& -2(y_t^* + \mu_y)[\widehat{y}_{1t}^* + \mu_1 - \widehat{y}_{2t}^* - \mu_2] + (\widehat{y}_{1t}^* + \mu_1)^2 - (\widehat{y}_{2t}^* + \mu_2)^2 \\
=& -2\{y_t^*\widehat{y}_{1t}^* + \mu_1 y_t^* - y_t^*\widehat{y}_{2t}^* - y_t^*\mu_2 + \mu_y\widehat{y}_{1t}^* + \mu_y\mu_1 - \widehat{y}_{2t}^*\mu_y - \mu_2\mu_y\} \\
& + \widehat{y}_{1t}^{*2} + 2\widehat{y}_{1t}^*\mu_1 + \mu_1^2 - \widehat{y}_{2t}^{*2} - 2\widehat{y}_{2t}^*\mu_2 - \mu_2^2 \\
=& \underbrace{-2[y_t^*(\mu_1 - \mu_2) + \widehat{y}_{1t}^*(\mu_y - \mu_1) - \widehat{y}_{2t}^*(\mu_y - \mu_2)]}_{I} - \underbrace{2[y_t^*(\widehat{y}_{1t}^* - \widehat{y}_{2t}^*)]}_{II} + \underbrace{\widehat{y}_{1t}^{*2} - \widehat{y}_{2t}^{*2}}_{III} + const\,.
\end{aligned} \tag{2.11}$$

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 (2017), 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 (2.11). 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 1 from Leschinski (2017). We thus have for $i \in \{1,2\}$

$$y_t^*\widehat{y}_{it}^* \sim \begin{cases} LM(\max\left\{d_y + d_i - 1/2, 0\right\}), & \text{if } S_{y,\widehat{y}_i} \neq 0 \\ LM(d_y + d_i - 1/2), & \text{if } S_{y,\widehat{y}_i} = 0 \end{cases} \tag{2.12}$$

$$\text{and} \qquad \widehat{y}_{it}^{*2} \sim LM(\max\{2d_i - 1/2, 0\})\,. \tag{2.13}$$

Further note that

$$d_y > d_y + d_i - 1/2 \;\text{ and }\; d_i > d_y + d_i - 1/2 \tag{2.14}$$

and

$$d_i > 2d_i - 1/2\,, \tag{2.15}$$

since $0 \leq d_a < 1/2$ for $a \in \{y, 1, 2\}$.
Using these properties, we can determine the memory d_z in (2.11) 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 (2.14) and (2.15) and we obtain $d_z = \max\left\{d_y,\ d_1,\ d_2\right\}$.*

2. *Second, if* $\mu_1 = \mu_2 \neq \mu_y$*, then* y_t^* *drops out from (2.11), but the two forecasts* $\widehat{y}_{1t}$ *and* $\widehat{y}_{2t}$ *remain. From (2.14) and (2.15), 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 (2.14) and (2.15), 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 (2.11). 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 (2.12) and (2.13). Furthermore, the memory of the squared series given in (2.13) is always non-negative from Corollary 1 in Leschinski (2017) 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*

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

Proof (Proposition 2.3). *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 2.3, we have

$$\mu_i = \beta_i + \xi_i \mu_x \tag{2.16}$$

and

$$\mu_y = \beta_y + \xi_y \mu_x \,. \tag{2.17}$$

We can now re-express the forecast error loss differential z_t in (2.11) 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*

$$\begin{aligned} z_t = & -2\{y_t^*(\mu_1-\mu_2) + x_t^*[\xi_1(\mu_y-\mu_1) - \xi_2(\mu_y-\mu_2)] + x_t^* y_t^*(\xi_1-\xi_2) - x_t^*(\xi_1\varepsilon_{1t} - \xi_2\varepsilon_{2t}) \\ & + \varepsilon_{1t}(\mu_y-\mu_1) - \varepsilon_{2t}(\mu_y-\mu_2) + \mu_x(\varepsilon_{1t}\xi_1 - \varepsilon_{2t}\xi_2) + y_t^*(\varepsilon_{1t}-\varepsilon_{2t})\} \\ & + x_t^{*^2}(\xi_1^2-\xi_2^2) + \varepsilon_{1t}^2 - \varepsilon_{2t}^2 + 2\mu_x(\varepsilon_{1t}\xi_1 - \varepsilon_{2t}\xi_2) + const\,. \end{aligned} \tag{2.18}$$

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

$$\begin{aligned} z_t = & -2\{x_t^*[(\mu_y-\mu_1)\xi_1 + \xi_y(\mu_1-\mu_2)] - \widehat{y}_{2t}^*[\mu_y-\mu_2] - \xi_y x_t^* \widehat{y}_{2t}^* + x_t^*[\varepsilon_{1t}(\xi_y-\xi_1) + \xi_1\eta_t] \\ & + \varepsilon_{1t}(\xi_y\mu_x-\mu_1) + \eta_t(\mu_1-\mu_2) + \varepsilon_{1t}\eta_t - \widehat{y}_{2t}^*\eta_t\} \\ & - (2\xi_1\xi_y - \xi_1^2)x_t^{*^2} + \varepsilon_{1t}^2 - \widehat{y}_{2t}^{*^2} - 2\beta_y\varepsilon_{1t} + const\,. \end{aligned} \tag{2.19}$$

   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:*

$$\begin{aligned} z_t = & -2\{x_t^*[\xi_1(\mu_y-\mu_1) - \xi_2(\mu_y-\mu_2) + \xi_y(\mu_1-\mu_2)] \\ & + x_t^*[(\xi_y-\xi_1)\varepsilon_{1t} - (\xi_y-\xi_2)\varepsilon_{2t} + (\xi_1-\xi_2)\eta_t] \\ & + x_t^{*2}[\xi_y(\xi_1-\xi_2) - \frac{1}{2}(\xi_1^2-\xi_2^2)] \\ & + \varepsilon_{1t}(\mu_y-\mu_1) - \varepsilon_{2t}(\mu_y+\mu_2) + \mu_x(\xi_1\varepsilon_{1t} + \xi_2\varepsilon_{2t}) + \eta_t(\varepsilon_{1t}-\varepsilon_{2t}) + \eta_t[\mu_1-\mu_2]\} \\ & + \varepsilon_{1t}^2 - \varepsilon_{2t}^2 + 2\mu_x(\xi_1\varepsilon_{1t} - \xi_2\varepsilon_{2t}) + const\,. \end{aligned} \tag{2.20}$$

As in the proof of Proposition 2.2, we can now determine the memory orders of z_t in (2.18), *(2.19) and (2.20) 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*

$$y_t^*(\mu_1-\mu_2)+x_t^*[\xi_1(\mu_y-\mu_1)-\xi_2(\mu_y-\mu_2)] \text{ in (2.18)},$$

$$x_t^*[(\mu_y-\mu_1)\xi_1+\xi_y(\mu_1-\mu_2)]-\widehat{y}_{2t}^*(\mu_y-\mu_2) \text{ in (2.19)}$$

and

$$x_t^*[\xi_1(\mu_y-\mu_1)-\xi_2(\mu_y-\mu_2)+\xi_y(\mu_1-\mu_2)] \text{ in (2.20)}$$

have the same structure as

$$y_t^*(\mu_1-\mu_2)+\widehat{y}_{1t}^*(\mu_y-\mu_1)-\widehat{y}_{2t}^*(\mu_y-\mu_2) \text{ in (2.11)}$$

and that all of the other non-constant terms in (2.18), (2.19) and (2.20) are either squares or products of demeaned series, so that their memory is reduced according to Proposition 1 from Leschinski (2017). From Assumption 2.3, 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 (2.14) and (2.15), 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.
To obtain the results in Proposition 2.3, set the terms in square brackets in equations (2.18), (2.19) and (2.20) equal to zero and solve for the quotient of the factor loadings. This determines the transmission of the memory of x_t^*. For the effect of the series that is not involved in the CLM relationship, we impose the restrictions $\mu_1 \neq \mu_2$ and $\mu_i \neq \mu_y$, as stated in the proposition. □

Proof (Proposition 2.4). *The results in Proposition 2.4 follow directly from equations (2.18), (2.19) and (2.20), above. For (2.18) the terms in square brackets can be re-expressed as*

$$[\xi_1(\mu_y-\mu_1)-\xi_2(\mu_y-\mu_2)] = (\xi_1-\xi_2)\mu_y+\xi_2\mu_2-\xi_1\mu_1 \,.$$

Obviously, for $\xi_1=\xi_2$, this is reduced to $\xi_2\mu_2-\xi_1\mu_1$, which does not vanish, since $\mu_1 \neq \mu_2$. The other cases are treated entirely analogous. For (2.19) we have

$$[(\mu_y-\mu_1)\xi_1+\xi_y(\mu_1-\mu_2)] = \xi_1\mu_y-\xi_y\mu_2 \,,$$

and in (2.20)

$$[\xi_1(\mu_y-\mu_1)-\xi_2(\mu_y-\mu_2)+\xi_y(\mu_1-\mu_2)] = (\xi_1-\xi_2)\mu_y-(\xi_1-\xi_y)\mu_1+(\xi_2-\xi_y)\mu_2 = 0 \,,$$

so that x_t^* drops out and the memory is reduced. □

Proof (Proposition 2.5). *First note that under the assumptions of Proposition 2.3, (2.18) is reduced to*

$$
\begin{aligned}
z_t &= -2\{-x_t^*(\xi_1\varepsilon_{1t} - \xi_2\varepsilon_{2t}) + y_t^*(\varepsilon_{1t} - \varepsilon_{2t})\} + \varepsilon_{1t}^2 - \varepsilon_{2t}^2 + const\,, \\
&= -2\{-\underbrace{\xi_1 x_t^*\varepsilon_{1t}}_{I} + \underbrace{\xi_2 x_t^*\varepsilon_{2t}}_{II} + \underbrace{y_t^*\varepsilon_{1t}}_{III} - \underbrace{y_t^*\varepsilon_{2t}}_{IV}\} + \underbrace{\varepsilon_{1t}^2}_{V} - \underbrace{\varepsilon_{2t}^2}_{VI} + const\,,
\end{aligned}
\tag{2.21}
$$

(2.19) becomes

$$
\begin{aligned}
z_t &= -2\{-x_t^*(\xi_y\widehat{y}_{2t}^* - \xi_1\eta_t) + (\varepsilon_{1t} - \widehat{y}_{2t}^*)\eta_t + \varepsilon_{1t}(\xi_y\mu_x - \mu_1)\} + \varepsilon_{1t}^2 - \widehat{y}_{2t}^{*^2} - 2\beta_y\varepsilon_{1t} - \xi_1\xi_y x_t^{*^2} + const\,, \\
&= -2\{-\underbrace{\xi_y x_t^*\widehat{y}_{2t}^*}_{I} + \underbrace{\xi_1 x_t^*\eta_t}_{II} + \underbrace{\varepsilon_{1t}\eta_t}_{III} - \underbrace{\widehat{y}_{2t}^*\eta_t}_{IV} + \underbrace{\varepsilon_{1t}(\xi_y\mu_x - \mu_1)}_{V}\} + \underbrace{\varepsilon_{1t}^2}_{VI} - \underbrace{\widehat{y}_{2t}^{*^2}}_{VII} - \underbrace{2\beta_y\varepsilon_{1t}}_{VIII} - \underbrace{\xi_1\xi_y x_t^{*^2}}_{IX} + const\,,
\end{aligned}
\tag{2.22}
$$

and finally (2.20) is

$$
\begin{aligned}
z_t &= -2(\varepsilon_{1t} - \varepsilon_{2t})\eta_t + \varepsilon_{1t}^2 - \varepsilon_{2t}^2 + const, \\
&= -2\underbrace{\varepsilon_{1t}\eta_t}_{I} + \underbrace{2\varepsilon_{2t}\eta_t}_{II} + \underbrace{\varepsilon_{1t}^2}_{III} - \underbrace{\varepsilon_{2t}^2}_{IV} + const\,.
\end{aligned}
\tag{2.23}
$$

We can now proceed as in the proof of Proposition 2.2 and infer the memory orders of each term in the respective linear combination from Proposition 2.1 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 (2.21), we have

$$
I_{2.21} \sim \begin{cases} LM(\max\{d_x + d_{\varepsilon_1} - 1/2,\ 0\}), & if\ S_{x,\varepsilon_1} \neq 0 \\ LM(d_x + d_{\varepsilon_1} - 1/2), & if\ S_{x,\varepsilon_1} = 0 \end{cases}
$$

$$
II_{2.21} \sim \begin{cases} LM(\max\{d_x + d_{\varepsilon_2} - 1/2,\ 0\}), & if\ S_{x,\varepsilon_2} \neq 0 \\ LM(d_x + d_{\varepsilon_2} - 1/2), & if\ S_{x,\varepsilon_2} = 0 \end{cases}
$$

$$
III_{2.21} \sim \begin{cases} LM(\max\left\{d_y + d_{\varepsilon_1} - 1/2,\ 0\right\}), & if\ S_{y,\varepsilon_1} \neq 0 \\ LM(d_y + d_{\varepsilon_1} - 1/2), & if\ S_{y,\varepsilon_1} = 0 \end{cases}
$$

$$
IV_{2.21} \sim \begin{cases} LM(\max\left\{d_y + d_{\varepsilon_2} - 1/2,\ 0\right\}), & if\ S_{y,\varepsilon_2} \neq 0 \\ LM(d_y + d_{\varepsilon_2} - 1/2), & if\ S_{y,\varepsilon_2} = 0 \end{cases}
$$

$$
V_{2.21} \sim LM(\max\{2d_{\varepsilon_1} - 1/2,\ 0\})
$$

$$
and \qquad VI_{2.21} \sim LM(\max\{2d_{\varepsilon_2} - 1/2\,,\ 0\}).
$$

Since by definition $d_x > d_{\varepsilon_i}$, the memory of $V_{2.21}$ and $VI_{2.21}$ is always of a lower order than that of $I_{2.21}$ and $II_{2.21}$. As in the proof of Proposition 2.2, the squares in terms $V_{2.21}$ and $VI_{2.21}$ establish *zero as the lower bound of* d_z. *Therefore, we have*

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

Similarly, in (2.22), we have

$$\begin{aligned}
I_{2.22} &\sim \begin{cases} LM(\max\{d_x + d_2 - 1/2,\ 0\}), & \text{if } S_{x,\widehat{y_2}} \neq 0 \\ LM(d_x + d_2 - 1/2), & \text{if } S_{x,\widehat{y_2}} = 0 \end{cases} \\
II_{2.22} &\sim \begin{cases} LM(\max\left\{d_x + d_\eta - 1/2,\ 0\right\}), & \text{if } S_{x,\eta} \neq 0 \\ LM(d_x + d_\eta - 1/2), & \text{if } S_{x,\eta} = 0 \end{cases} \\
III_{2.22} &\sim \begin{cases} LM(\max\left\{d_{\varepsilon_1} + d_\eta - 1/2,\ 0\right\}), & \text{if } S_{\varepsilon_1,\eta} \neq 0 \\ LM(d_{\varepsilon_1} + d_\eta - 1/2), & \text{if } S_{\varepsilon_1,\eta} = 0 \end{cases} \\
IV_{2.22} &\sim \begin{cases} LM(\max\left\{d_2 + d_\eta - 1/2,\ 0\right\}), & \text{if } S_{\widehat{y_2},\eta} \neq 0 \\ LM(d_2 + d_\eta - 1/2), & \text{if } S_{\widehat{y_2},\eta} = 0 \end{cases} \\
V_{2.22} &\sim LM(d_{\varepsilon_1}) \\
VI_{2.22} &\sim LM(\max\{2d_{\varepsilon_1} - 1/2,\ 0\}) \\
VII_{2.22} &\sim LM(\max\{2d_2 - 1/2,\ 0\}) \\
VIII_{2.22} &\sim LM(d_{\varepsilon_1}) \\
\text{and} \qquad IX_{2.22} &\sim LM(\max\{2d_x - 1/2, 0\}).
\end{aligned}$$

Here, $V_{2.22}$ *can be disregarded since it is of the same order as* $VIII_{2.22}$. $VIII_{2.22}$ *dominates* $VI_{2.22}$, *because* $d_{\varepsilon_1} < 1/2$. *Finally, as* $d_{\varepsilon_1} < d_x$ *holds by assumption,* $III_{2.22}$ *is dominated by* $II_{2.22}$ *and* $d_\eta < d_x$, *so that* $IX_{2.22}$ *dominates* $II_{2.22}$. *Therefore,*

$$d_z = \max\left\{d_2 + \max\left\{d_x, d_\eta\right\} - 1/2,\ 2\max\{d_x, d_2\} - 1/2,\ d_{\varepsilon_1}\right\}.$$

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 (2.23), we have

$$\begin{aligned}
I_{2.23} &\sim \begin{cases} LM(\max\left\{d_\eta + d_{\varepsilon_1} - 1/2,\ 0\right\}), & \text{if } S_{\eta,\varepsilon_1} \neq 0 \\ LM(d_\eta + d_{\varepsilon_1} - 1/2), & \text{if } S_{\eta,\varepsilon_1} = 0 \end{cases} \\
II_{2.23} &\sim \begin{cases} LM(\max\left\{d_\eta + d_{\varepsilon_2} - 1/2,\ 0\right\}), & \text{if } S_{\eta,\varepsilon_1} \neq 0 \\ LM(d_\eta + d_{\varepsilon_2} - 1/2), & \text{if } S_{\eta,\varepsilon_2} = 0 \end{cases}
\end{aligned}$$

$$III_{2.23} \sim LM(\max\{2d_{\varepsilon_1} - 1/2,\ 0\})$$
$$IV_{2.23} \sim LM(\max\{2d_{\varepsilon_2} - 1/2,\ 0\})\,.$$

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 $III_{2.23}$ and $IV_{2.23}$. □

Proof (Proposition 2.6). *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. □

2.8.1 Supplementary Appendix

This supplementary appendix which is also available online, contains additional simulation results regarding additional sample sizes and moving average dynamics. Several size and power experiments are conducted. An overview can be found in the list of contents located at the beginning of each subsection.

Contents:

- Section A: Simulation Results for other Sample Sizes
- Section B: Simulation Results for Moving Average Dynamics

A: Simulation Results for other Sample Sizes

- Figure 2.8: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{50, 1000\}$ and bandwidth parameters q and b.
- Figure 2.9: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample sizes $T \in \{50, 1000\}$.
- Figure 2.10: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample size $T = 4000$.
- Figure 2.11: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample size $T = 4000$, when testing right-sided.
- Figure 2.12: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{50, 250, 1000, 2000\}$ and bandwidth parameters q and b, if the ALPW estimator is used for the plug-in estimation of d.
- Figure 2.13: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$, sample sizes $T \in \{50, 250, 1000, 2000\}$, if the ALPW estimator is used for the plug-in estimation of d.
- Figure 2.14: Power comparison of the robust statistics t_{EFB} and t_{MAC} with their short memory counterparts, when $d = 0$ under DGP1. Samples sizes are $T \in \{50, 250, 1000, 2000\}$.
- Figure 2.15: Power comparison of the robust statistics t_{EFB} and t_{MAC} with their short memory counterparts, when $d = 0$ under DGP2. Samples sizes are $T \in \{50, 250, 1000, 2000\}$.
- Figure 2.16: Power comparison of the t_{MAC}- and t_{EFB}-statistics, for $d = 0.2$ and sample sizes $T \in \{50, 250, 1000, 2000\}$, adjusted for size and with known memory parameter.

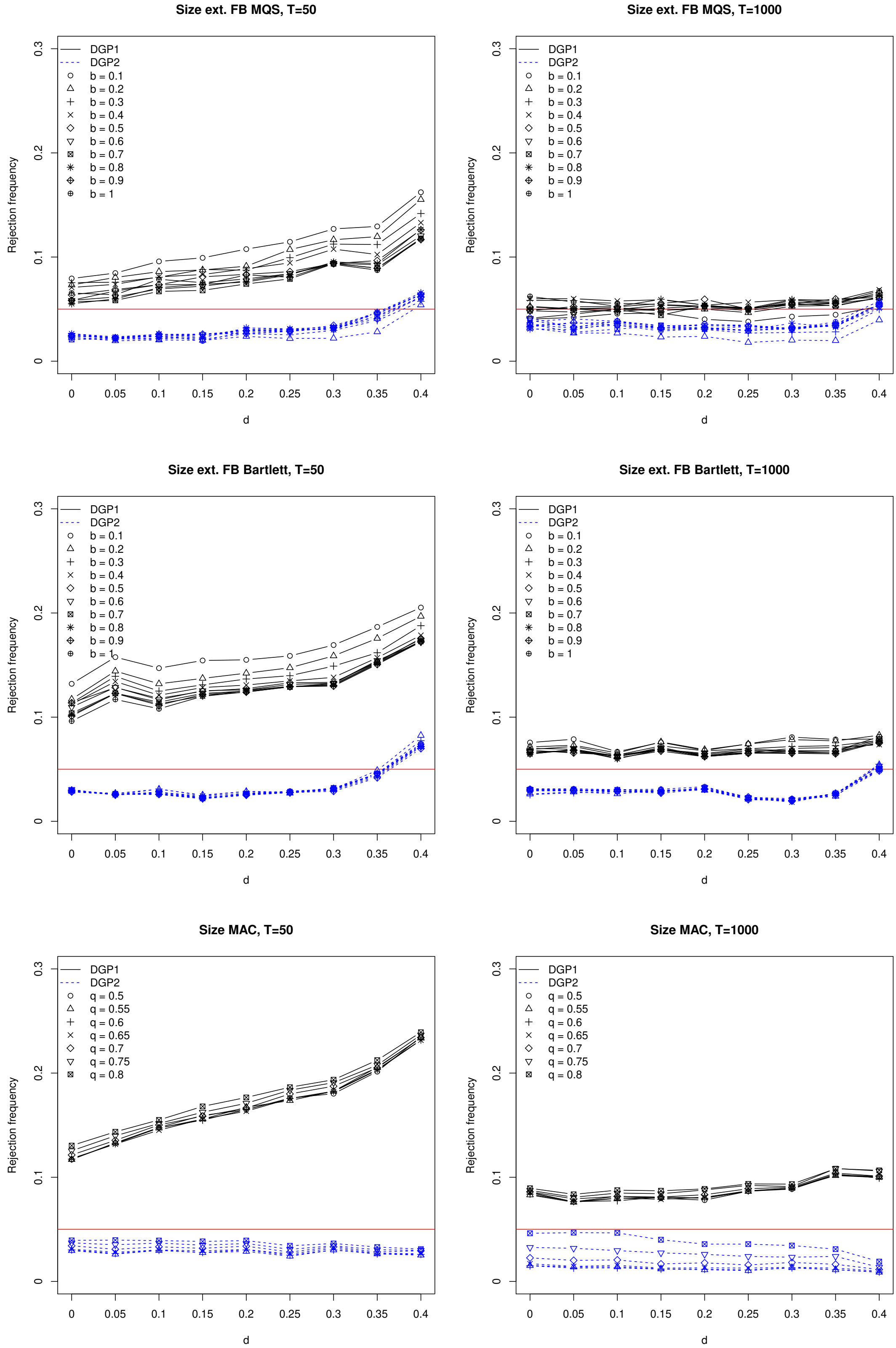


Figure 2.8: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{50, 1000\}$ and bandwidth parameters q and b.

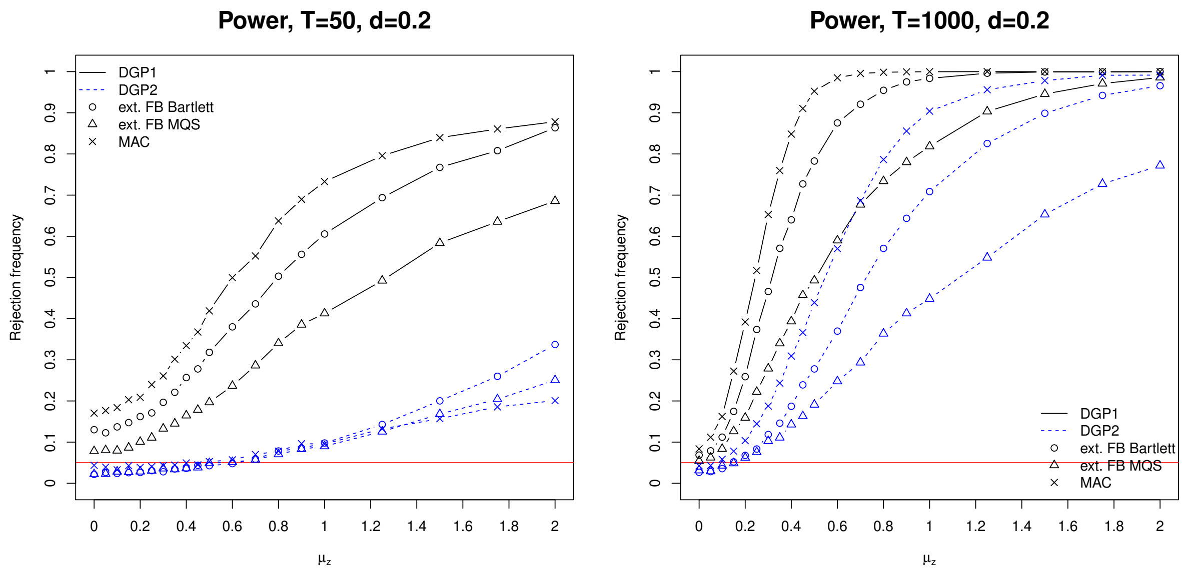


Figure 2.9: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample sizes $T \in \{50, 1000\}$.

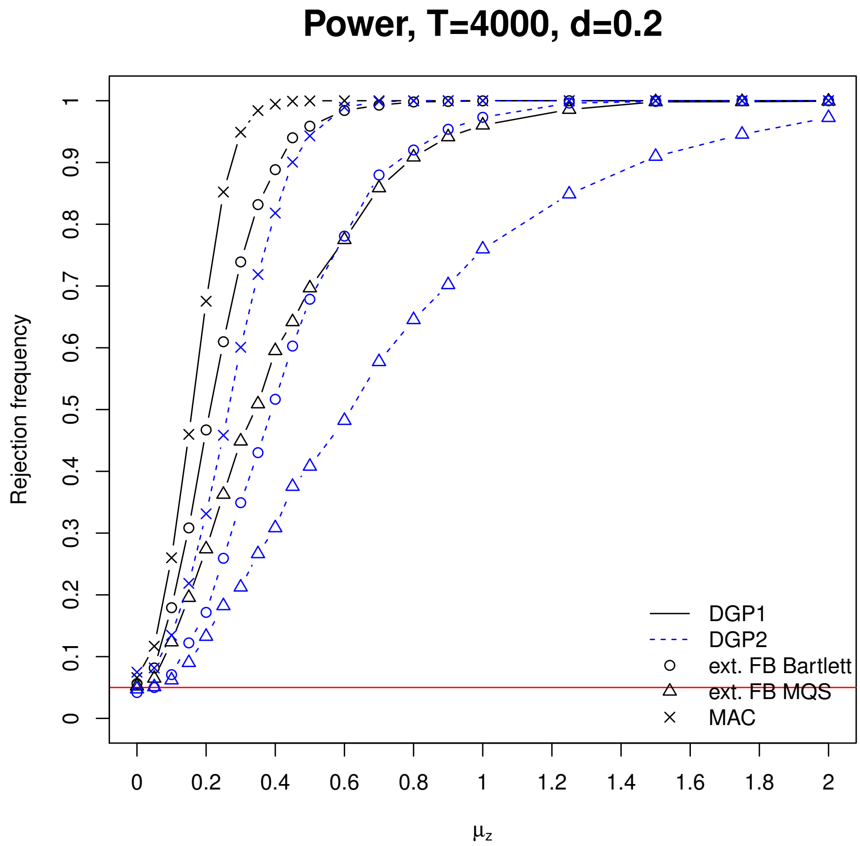


Figure 2.10: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample size $T = 4000$.

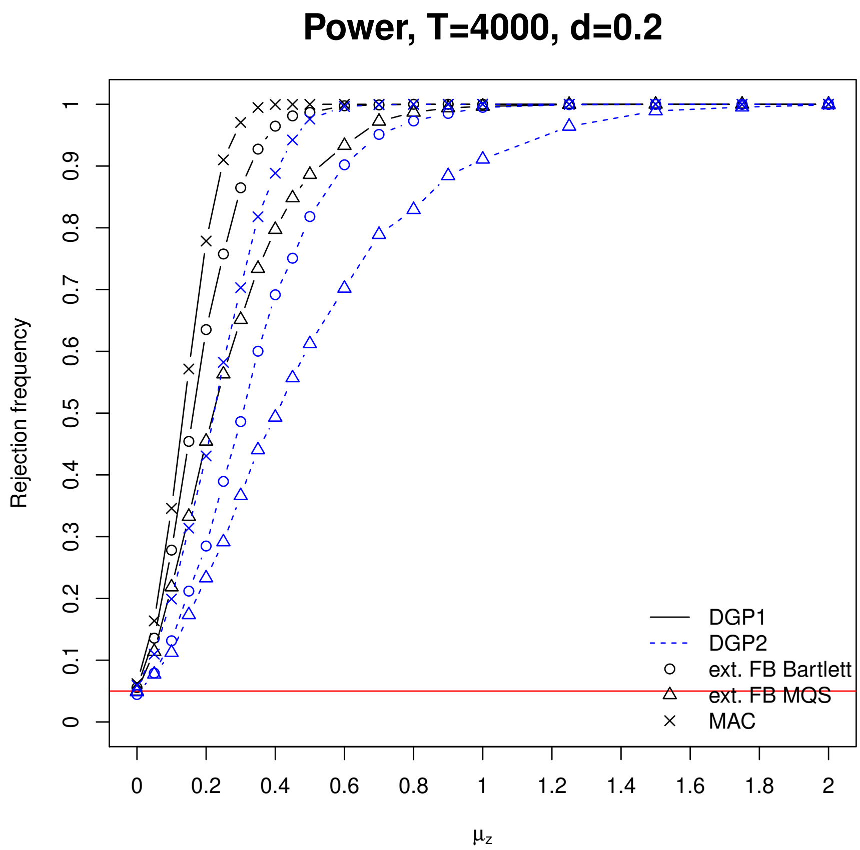


Figure 2.11: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample size $T = 4000$, when testing right-sided.

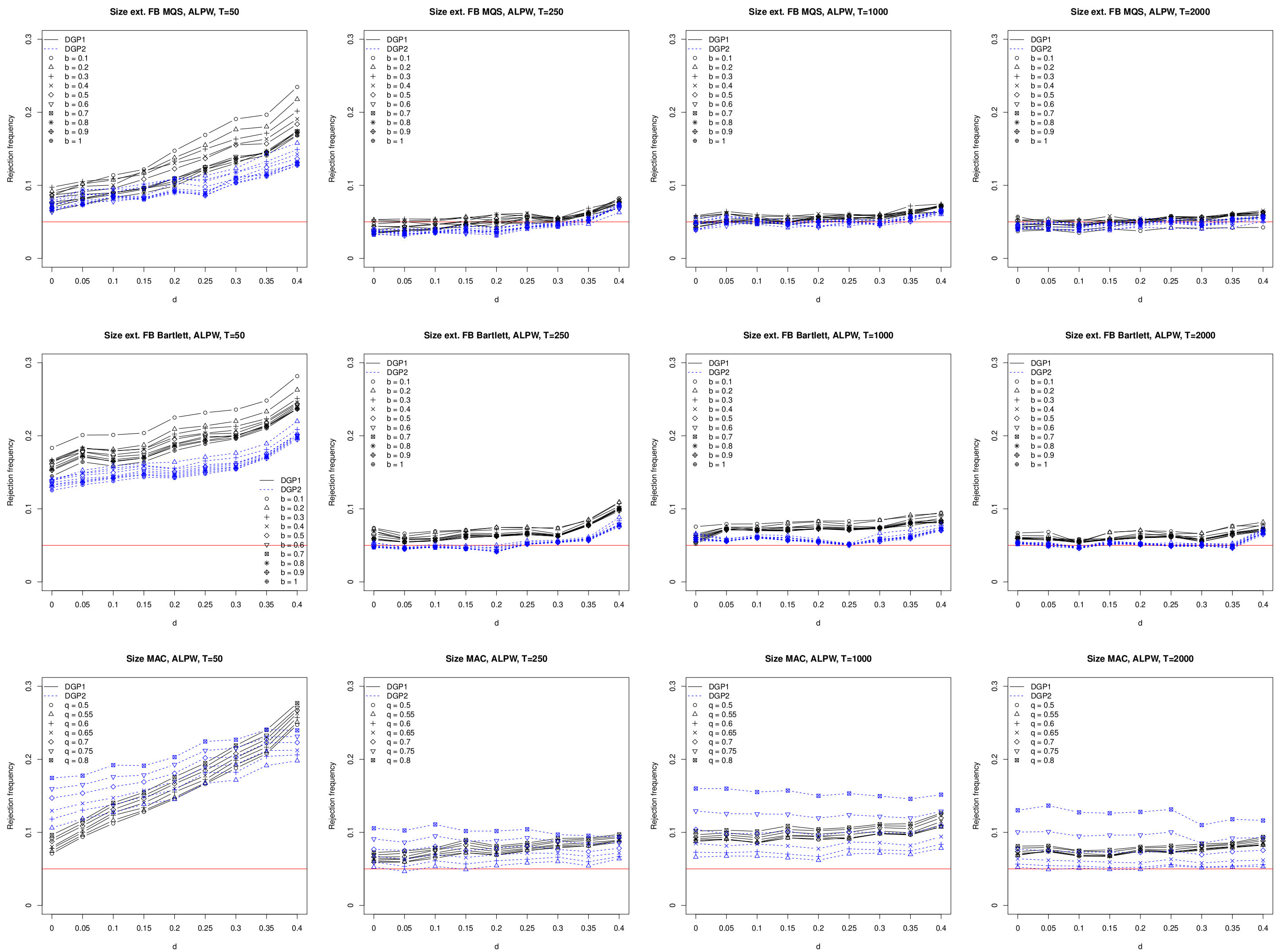


Figure 2.12: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{50, 250, 1000, 2000\}$ and bandwidth parameters q and b, if the ALPW estimator is used for the plug-in estimation of d.

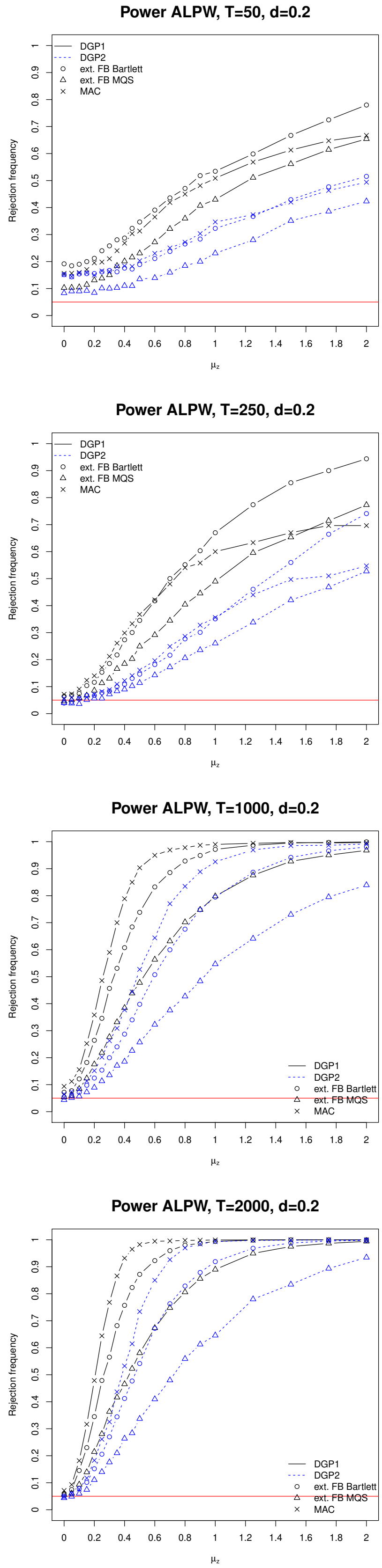


Figure 2.13: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$, sample sizes $T \in \{50, 250, 1000, 2000\}$, if the ALPW estimator is used for the plug-in estimation of d.

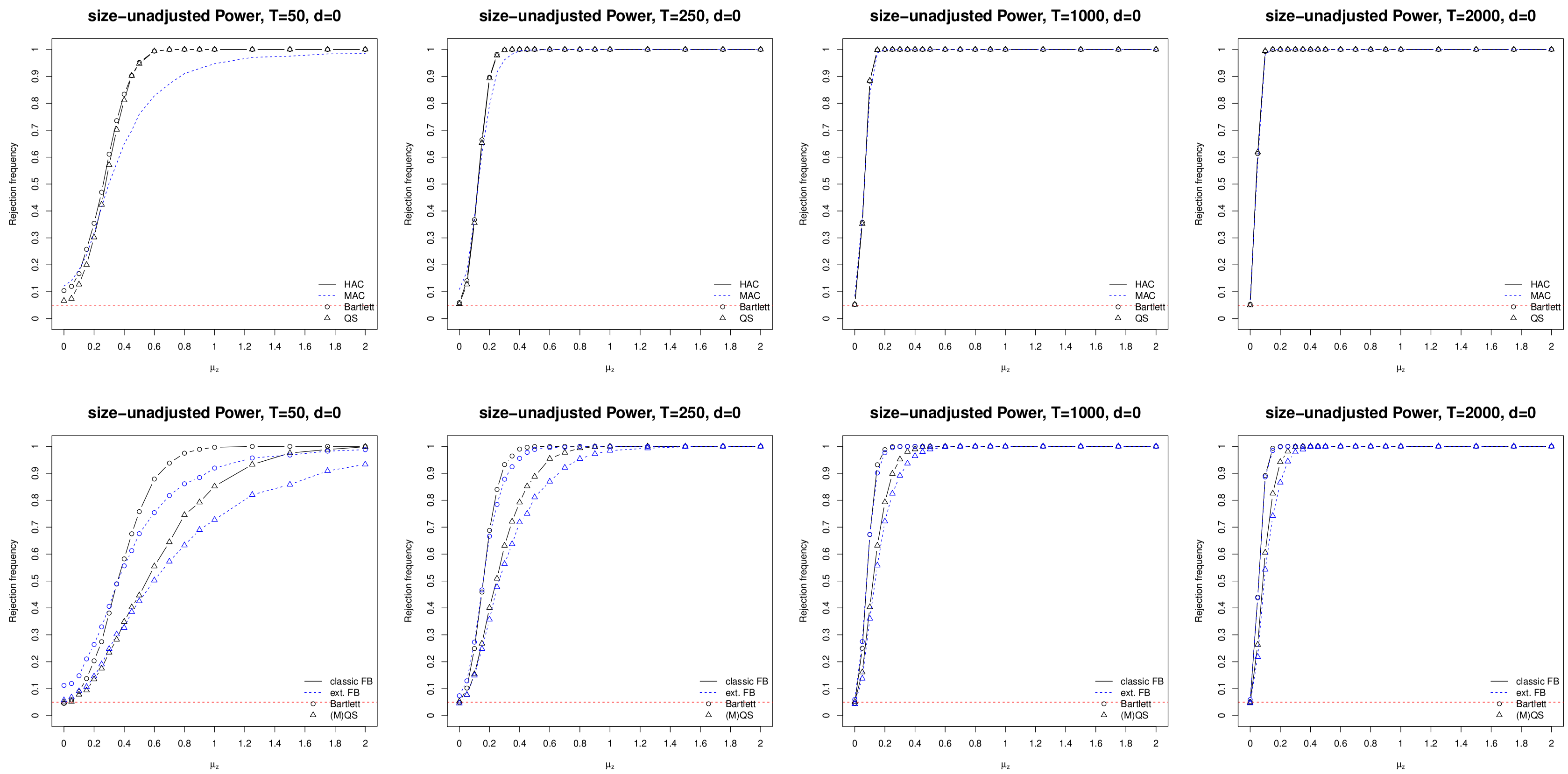


Figure 2.14: Power comparison of the robust statistics t_{EFB} and t_{MAC} with their short memory counterparts, when $d = 0$ under DGP1. Samples sizes are $T \in \{50, 250, 1000, 2000\}$.

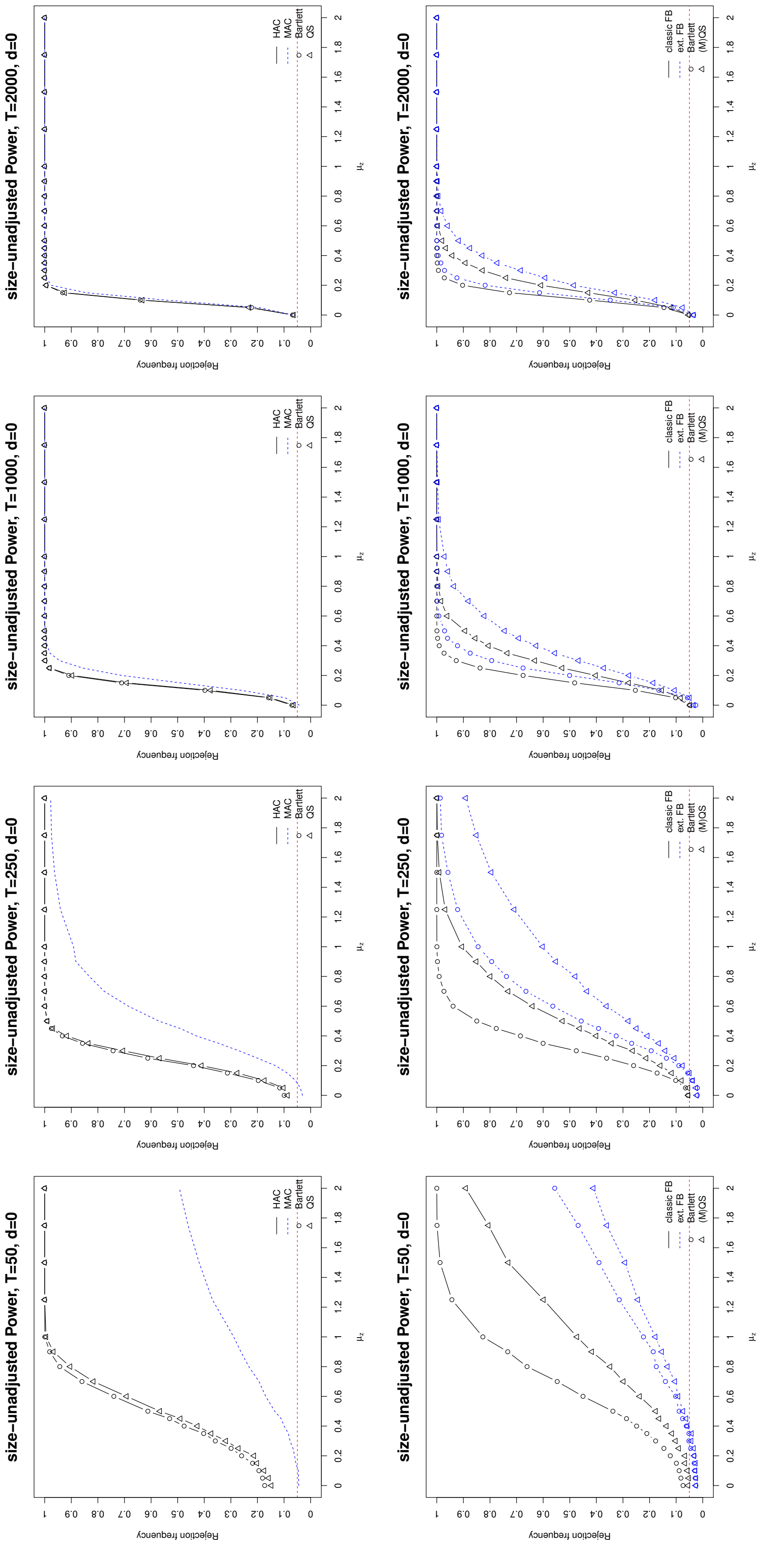


Figure 2.15: Power comparison of the robust statistics t_{EFB} and t_{MAC} with their short memory counterparts, when $d = 0$ under DGP2. Samples sizes are $T \in \{50, 250, 1000, 2000\}$.

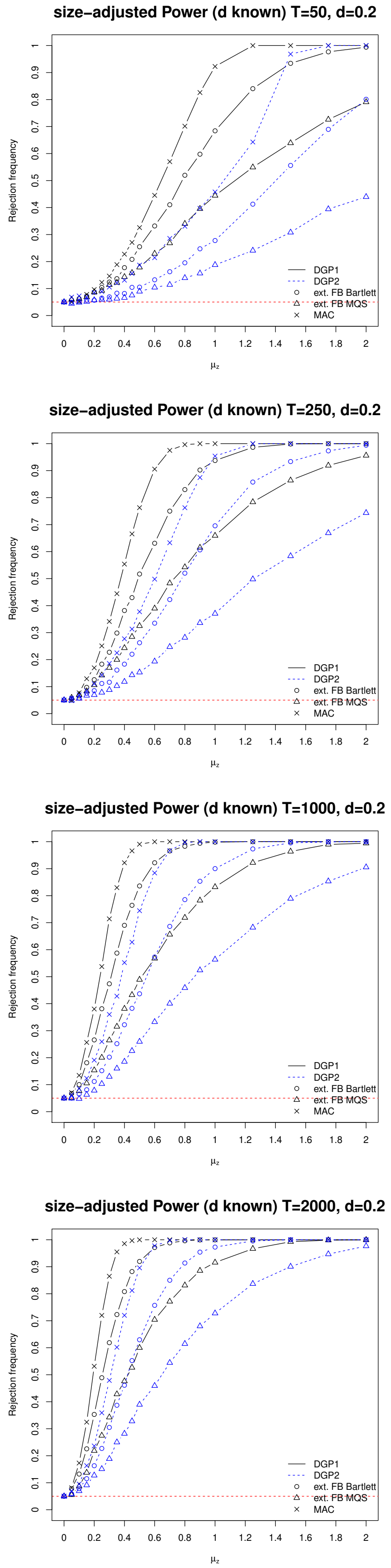


Figure 2.16: Power comparison of the t_{MAC}- and t_{EFB}-statistics, for $d = 0.2$ and sample sizes $T \in \{50, 250, 1000, 2000\}$, adjusted for size and with known memory parameter.

B: Simulation Results for Moving Average Dynamics

This section contains additional simulation results for DGP3, a first-order MA process with MA parameter $\theta = 0.6$.

- Figure 2.17: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{50, 250, 1000, 2000\}$ and bandwidth parameters q and b.
- Figure 2.18: Power comparison of the t_{MAC}- and t_{EFB}-statistics, for $d = 0.2$ and sample sizes $T \in \{50, 250, 1000, 2000\}$.
- Figure 2.19: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample size $T = 4000$, when testing right-sided.
- Figure 2.20: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{50, 250, 1000, 2000\}$ and bandwidth parameters q and b, if the ALPW estimator is used for the plug-in estimation of d.
- Figure 2.21: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$, sample sizes $T \in \{50, 250, 1000, 2000\}$, if the ALPW estimator is used for the plug-in estimation of d.
- Figure 2.22: Power comparison of the robust statistics t_{EFB} and t_{MAC} with their short memory counterparts, when $d = 0$. Sample sizes are $T \in \{50, 250, 1000, 2000\}$.
- Figure 2.23: Power comparison of the t_{MAC}- and t_{EFB}-statistics, for $d = 0.2$ and sample sizes $T \in \{50, 250, 1000, 2000\}$, adjusted for size and with known memory parameter.

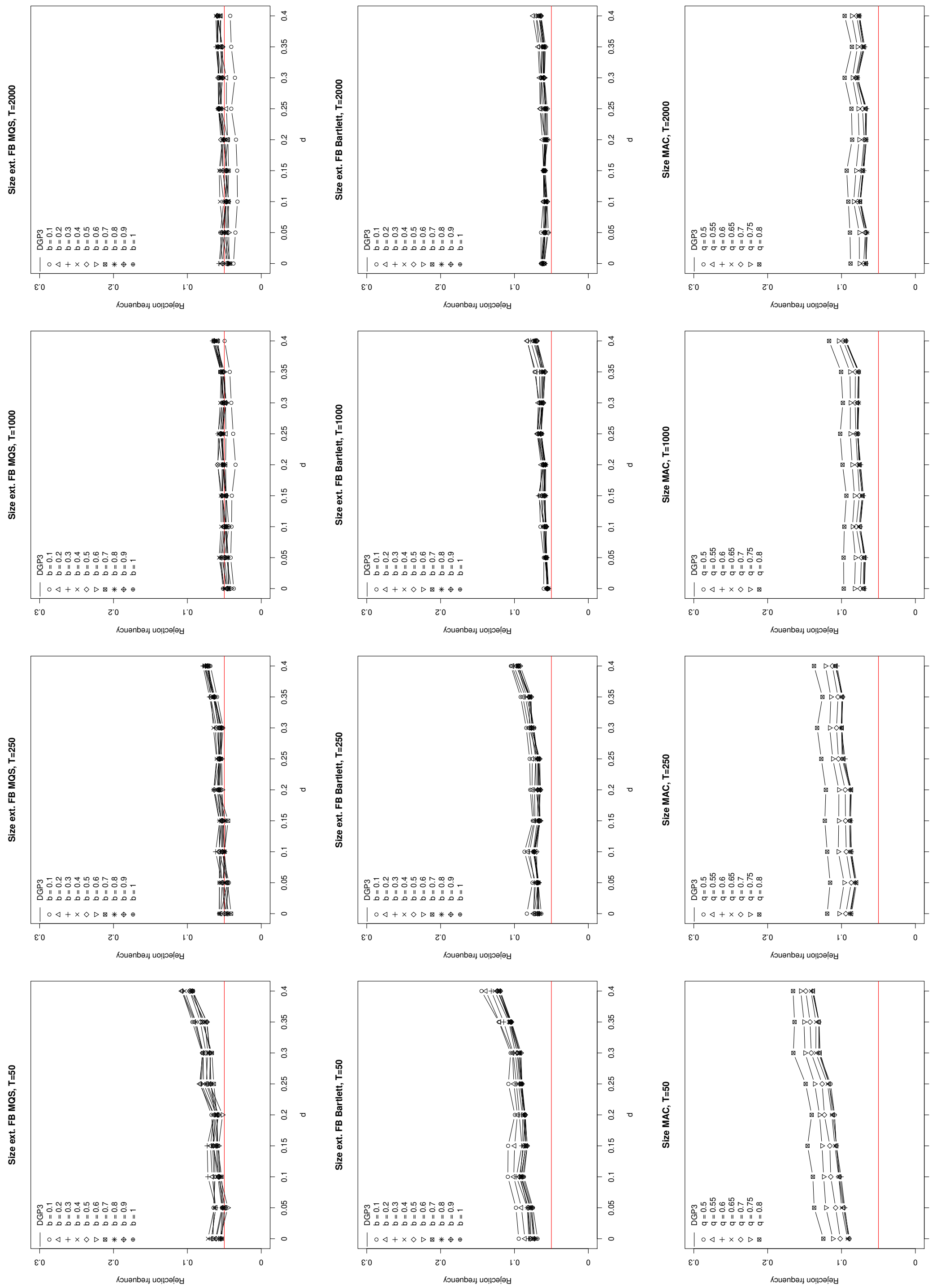


Figure 2.17: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{50, 250, 1000, 2000\}$ and bandwidth parameters q and b.

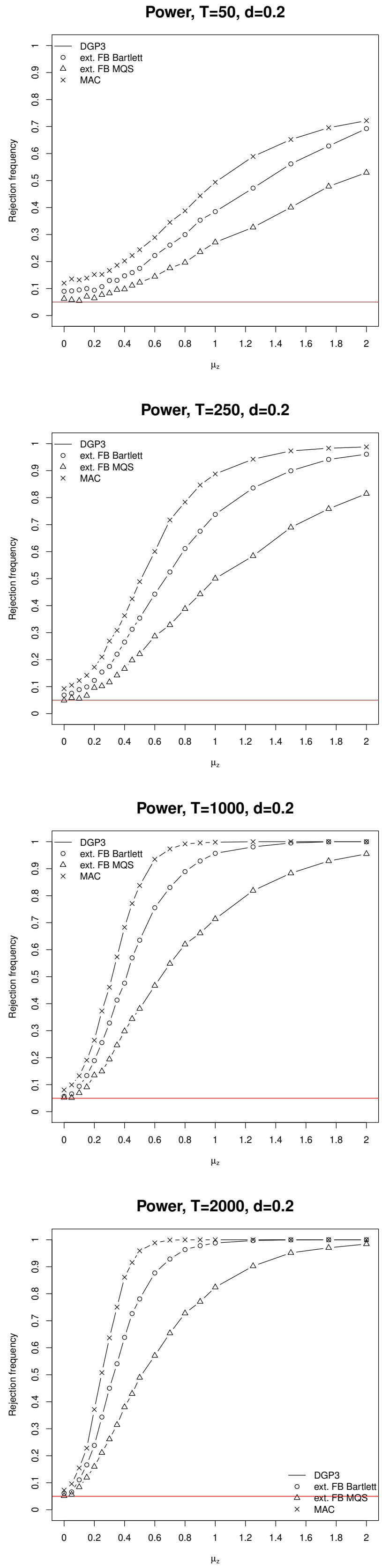


Figure 2.18: Power comparison of the t_{MAC}- and t_{EFB}-statistics, for $d = 0.2$ and sample sizes $T \in \{50, 250, 1000, 2000\}$.

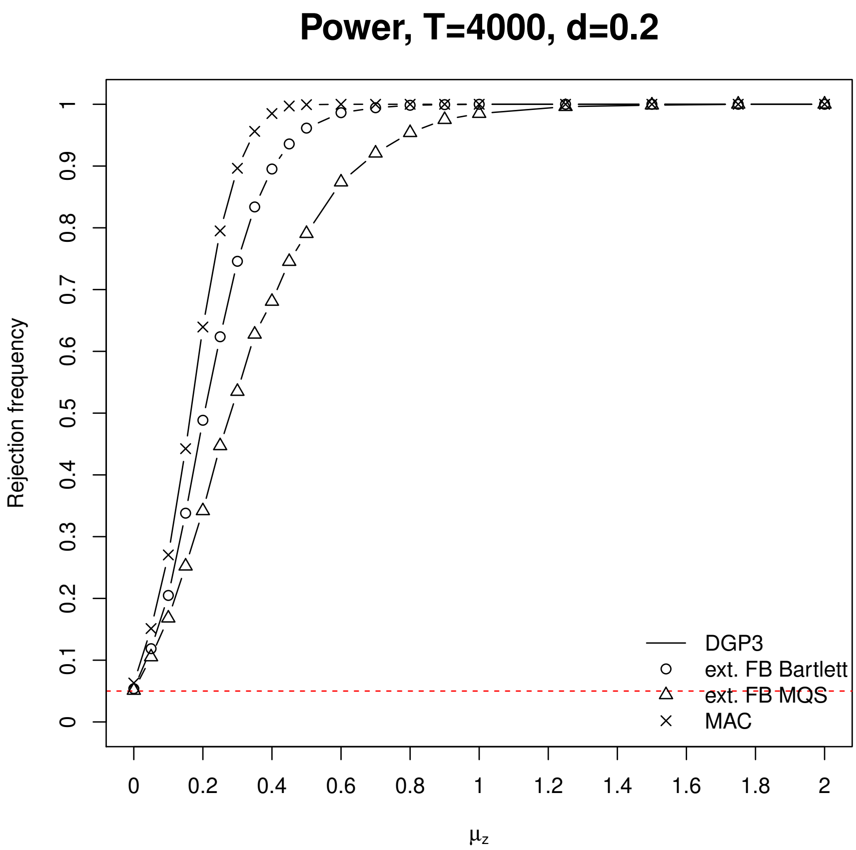


Figure 2.19: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$ and sample size $T = 4000$, when testing right-sided.

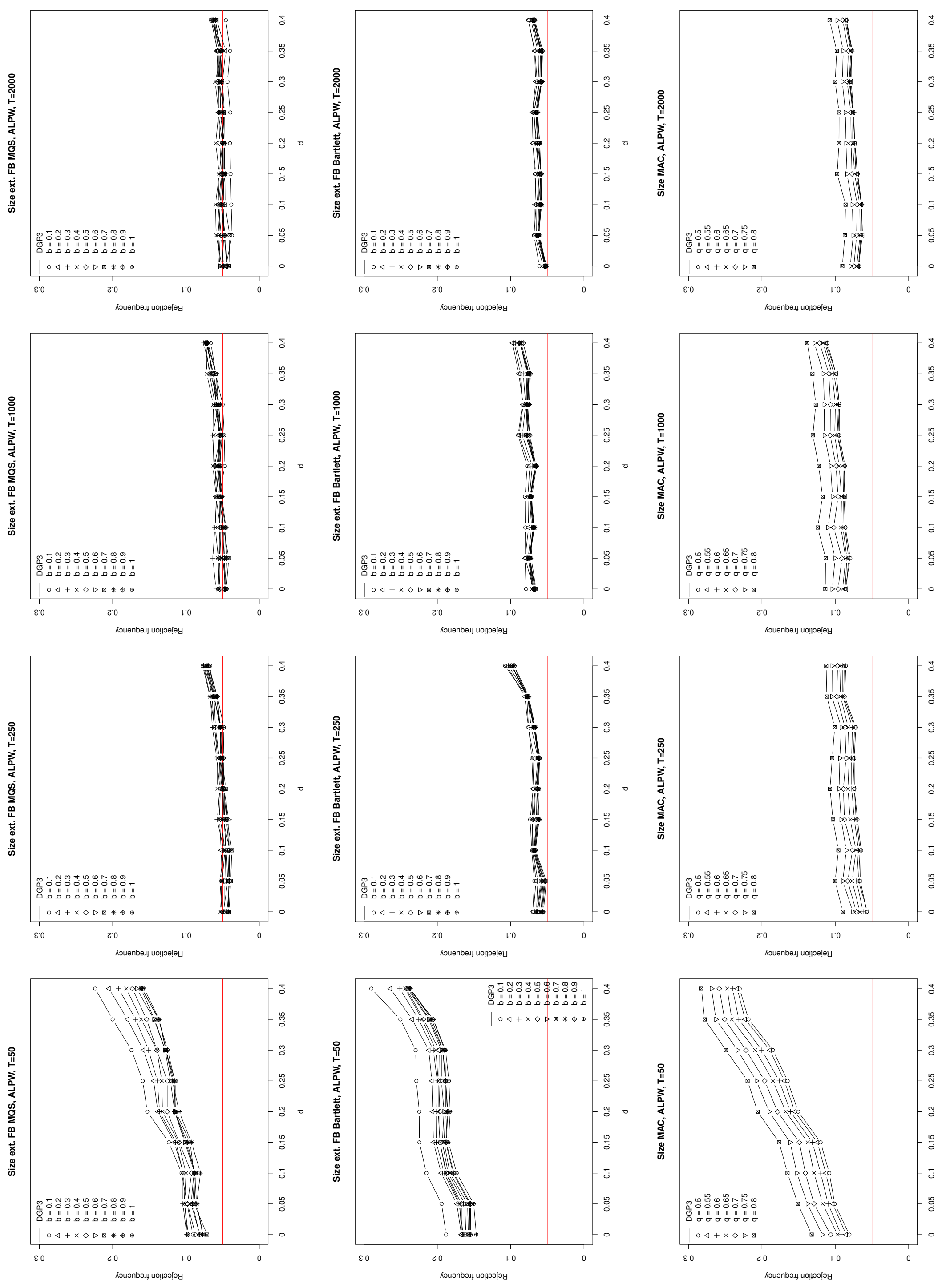


Figure 2.20: Size of the t_{MAC}- and t_{EFB}-statistics for different degrees of long memory d, sample sizes $T \in \{50, 250, 1000, 2000\}$ and bandwidth parameters q and b, if the ALPW estimator is used for the plug-in estimation of d.

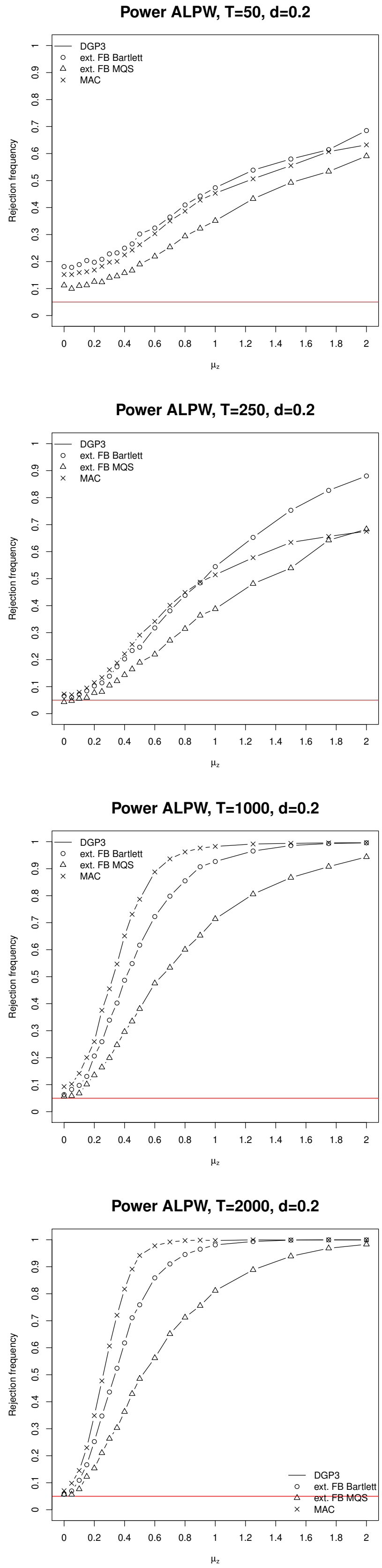


Figure 2.21: Power comparison of the t_{MAC}- and t_{EFB}-statistics for $d = 0.2$, sample sizes $T \in \{50, 250, 1000, 2000\}$, if the ALPW estimator is used for the plug-in estimation of d.

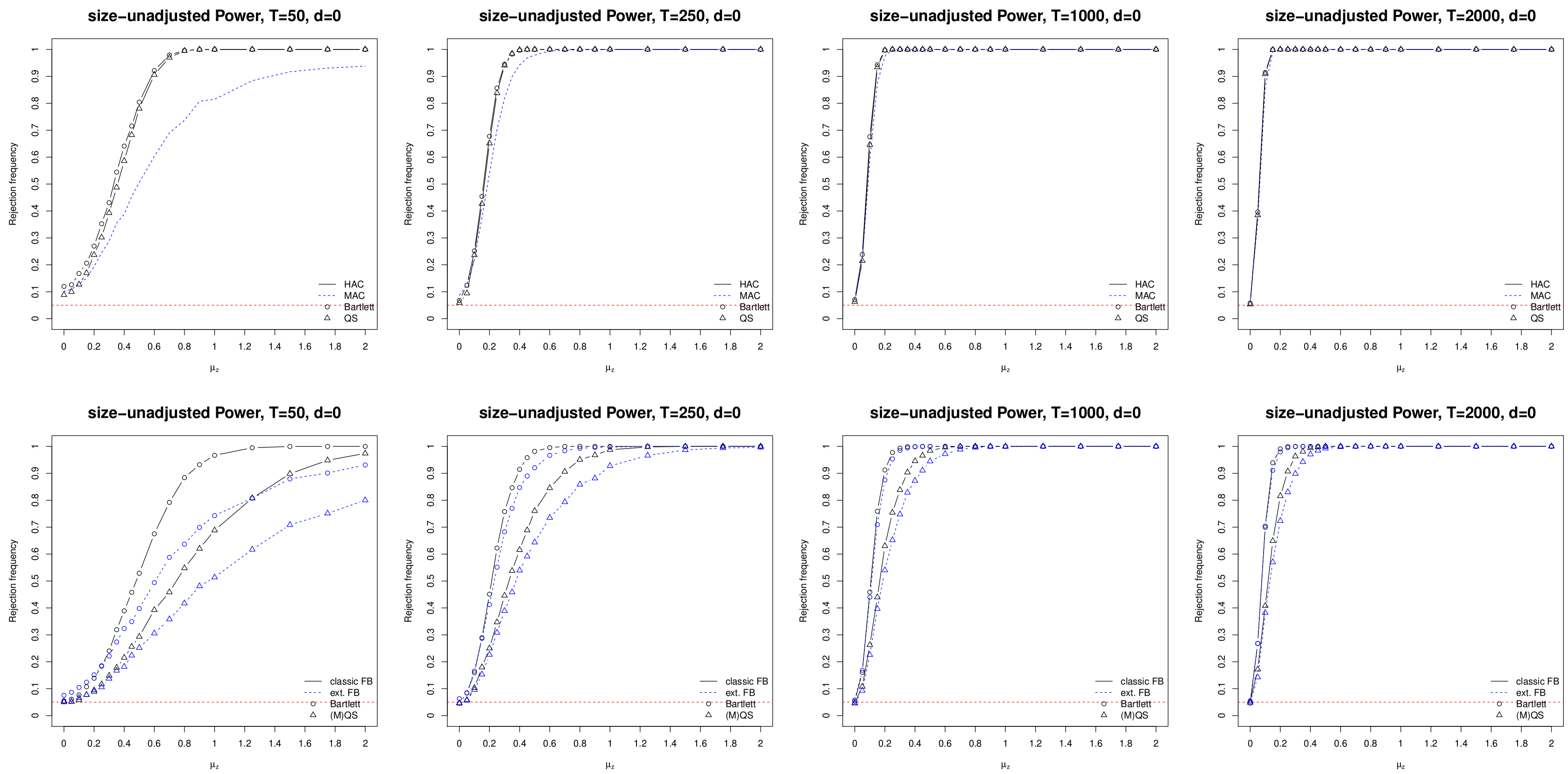


Figure 2.22: Power comparison of the robust statistics t_{EFB} and t_{MAC} with their short memory counterparts, when $d = 0$. Sample sizes are $T \in \{50, 250, 1000, 2000\}$.

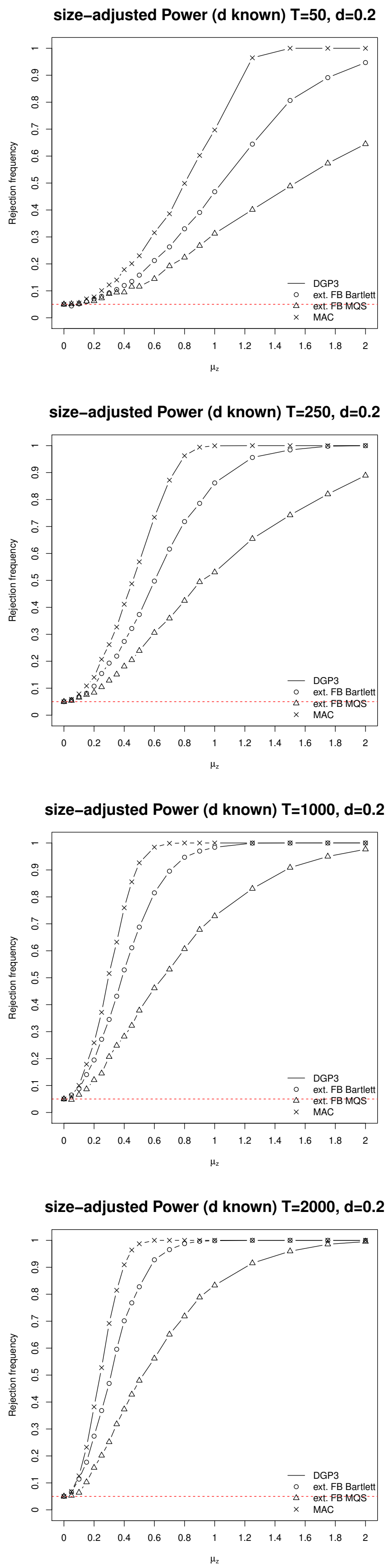


Figure 2.23: Power comparison of the t_{MAC}- and t_{EFB}-statistics, for $d = 0.2$ and sample sizes $T \in \{50, 250, 1000, 2000\}$, adjusted for size and with known memory parameter.

Chapter 3

A Robust Lagrange Multiplier Test for Long Memory

A Robust Lagrange Multiplier Test for Long Memory

Co-authored with Matei Demetrescu and Philipp Sibbertsen.

Distinguishing long from short memory is an important task of applied time series analysis due to the different asymptotic behavior of subsequent statistical inference. As a motivating example one can consider testing for significance of the mean of a given series by a simple t-test. Under short memory it is well known that the usual transformation stabilizing rate is of order $O(\sqrt{T})$, whereas long memory leads to $O(T^{1/2-d})$ with $0 < d < 1/2$. Moreover, the selection of an appropriate long-run variance estimator for correct studentization of the statistic crucially hinges on the memory degree present in the given series. In this sense tests based on the Lagrange Multiplier (LM) principle are especially appealing. Apart from only requiring an estimate of the restricted likelihood under the null they are well known to be the locally best invariant tests when conducting inference on the fractional integration parameter in an $ARFIMA(p,d,q)$ model (c.f. Tanaka (1999)). Therefore, Robinson (1991, 1994), Agiakloglou and Newbold (1994), Tanaka (1999), Breitung and Hassler (2002), Nielsen (2005) and Harris et al. (2008) among others, propose various LM testing procedures. Nevertheless, all these contributions require the assumption of at least unconditional homoskedasticity in the error terms of the data generating process. In recent years however, a growing body of research found empirical evidence that macroeconomic and financial time series exhibit various types of unconditional heteroskedasticity (see Loretan and Phillips (1994), McConnell and Perez-Quiros (2000), Sensier and Van Dijk (2004), Stărică and Granger (2005) and Cavaliere et al. (2015b), among others). Complementary research advanced analyzing the severe effects such time-varying conditional and unconditional volatility has on standard inference procedures like unit root or cointegration tests (see Cavaliere and Taylor (2007, 2008, 2009), Demetrescu and Hanck (2012a,b), Cavaliere et al. (2015a) and c.f. abundant references in Harris and Kew (2017)). In particular Cavaliere et al. (2015b) show that heteroskedasticity of a quite general form leads to non-pivotal asymptotic null distributions of LM tests for the order of fractional integration in an $ARFIMA(p,d,q)$ model. Similar results are found in Kew and Harris (2009) and Demetrescu and Sibbertsen (2016), where the latter study periodogram-based inference.

It is straightforward that regression-based tests can easily be robustified by the use of heteroskedasticity consistent (HAC) standard errors in the style of Andrews (1991). However, these will not account for time-varying dynamics in the short-run components of locally stationary (LS) processes as proposed by Dahlhaus (2000) and Palma and Olea (2010). In particular smooth or abrupt changes in the short memory indirectly induce heteroskedasticity and distort limiting null distributions of long memory tests due to switching in the main and off-diagonal elements of the variance covariance matrix of the process. A HAC correction for the regression-based tests without prewhitening is futile since the test will reject consistently under unaccounted short memory dynamics under the null (c.f. Harris et al. (2008)).

The same reasoning applies to the regression-based tests with HAC standardization and lag augmentation, based on a stationarity assumption like in Demetrescu et al. (2008). Complementary, in a parametric setting when using a prewhitening approach like in Robinson (1994), Tanaka (1999) or Breitung and Hassler (2002), the procedure should be flexible enough to allow for changes in the short-run dynamics. However, without strong restrictions on the parameter vector, the complexity of the prewhitening step is likely to be uncontrolable. In light of these shortcomings, nonparametric short memory corrections seem a more promising approach. In the frequency domain one can consider the local quasi likelihood approximation employing the localized periodogram (see Dahlhaus (2000), Beran (2009) or Palma and Olea (2010)). However, this is plausibly not very reliable in finite samples due to the use of two bandwidth parameters. In the time-domain, Harris et al. (2008) discuss a nonparametric correction for short-run dynamics of the LM test in the version of Tanaka (1999). By leaving out an increasing number of low-order autocovariances, this LM test is solely based on the standardized weighted sum of higher-order autocovariances, whose behavior is controlable under the null. However, under the long memory alternative, this quantity diverges. Nevertheless, these authors still require second-order stationarity and conditional homoskedasticity in their assumptions.

Therefore, we propose a nonparametric time-domain based LM test for long memory with robustness against fixed or switching short-run dynamics and heteroskedasticity of a quite general form. To this end, we recast the test of Harris et al. (2008) in a regression framework in the style of Breitung and Hassler (2002), which makes the use of HAC standard errors straightforward. An extensive Monte Carlo simulation study shows that the limiting null distribution of our test is standard normal and despite of its several robustness features, we still obtain good power results. Moreover, a comparism with other existing time-domain based Lagrange Multiplier testing procedures in the literature shows that it is the only test capable of controlling its size under the heteroskedastic locally stationary process class considered here. In our empirical application we find that long memory remains in different types of ex-post Variance Risk Premium series, represented by a fractional cointegration relationship between CBOE Volatility Index and realized variance of the S&P500 index, in addition to conditional and unconditional heteroskedasticity.

The rest of this paper is structured as follows: Section 3.1 defines the heteroskedastic locally stationary process, gives an overview over the existing methodology as well as various LM tests in the time-domain and presents our new test. Section 3.2 contains the result of our Monte Carlo (MC) simulations. An empirical application on the ex-post Variance Risk Premium series is considered in Section 3.3 and Section 3.4 concludes. Additional simulation results are gathered in the appendix.

3.1 Testing for Long Memory under Locally Stationary Processes

Following Dahlhaus (2000), Palma and Olea (2010) as well as Cavaliere et al. (2015b), let the heteroskedastic LS process $\{y_t\}$ for $t = 1, \ldots, T$, be defined by:

$$\underbrace{(1-L)^d y_t}_{u_t} = \underbrace{\frac{\theta(L,\tilde{\tau})}{\phi(L,\tilde{\tau})}}_{c(L,\tilde{\tau},\psi)^{-1}} \varepsilon_t\ , \tag{3.1}$$

where d is the memory parameter with $d \in [0, 0.5)$, $(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)L^j}{\Gamma(-d)\Gamma(j+1)}$ and $\phi(L,\tilde{\tau})$ as well as $\theta(L,\tilde{\tau})$ denote autoregressive and moving average polynomials of order p and q, respectively. The short-run polynomials are allowed to exhibit sudden or smooth deterministic switches in which case they depend on the relative sample size $\tilde{\tau} = \lfloor \frac{t}{T} \rfloor \in (0,1)$. One can easily observe that this model nests the classic $ARFIMA(p,d,q)$ process originally proposed by Granger and Joyeux (1980) and Hosking (1981) if all parameters remain fixed, and therefore independent of $\tilde{\tau}$, and one assumes $\{\varepsilon_t\}$ to be white noise. Concerning the extended Wold representation of $\{u_t\}$, we impose:

Assumption 3.1 (Extended Wold Representation). *Let* $u_t = \sum_{j=0}^{\infty} b_{j,t}\varepsilon_{t-j}$*, where:*

$$b_{j,t} = b_j c_j(\tilde{\tau})\ ,$$

with b_j *being s-summable and* $c_j(\tilde{\tau})$*, piecewise Lipschitz functions, such that* $\sup_j \max_{\tilde{\tau}} \left|c_j(\tilde{\tau})\right| < \infty$ *holds. For simplicity let any discontinuities of* $c_j(\cdot)$ *occur at the same relative time* $\tilde{\tau}_i \in (0,1)$ *for all* j*.*

Note that $\{u_t\}$ is globally non-stationary since the dependence of the short-run polynomials on $\tilde{\tau}$ almost surely generates at least one deterministic switch in the main and off-diagonal elements of the autocovariance matrix. However, the process is second-order stationary in each regime with all roots of the local short-run polynomials assumed to lie strictly outside the unit circle which is in turn implied by the summability condition of the local Wold coefficients. More practically spoken, this for example, allows to model sudden breaks as well as linear, logistic, or trigonometric trends in $c(L,\tilde{\tau},\psi)^{-1}$ of $\{y_t\}$ which almost surely leads to indirectly induced unconditional heteroskedasticity. Concerning the error terms $\{\varepsilon_t\}$, we directly follow Cavaliere et al. (2015b) and adopt:

Assumption 3.2 (Heteroskedasticity). *Let* $\varepsilon_t = \sigma_t z_t$ *satisfy the following conditions, respectively:*

a) $\{\sigma_t\}_{t\in\mathbb{Z}}$ *is non-stochastic and uniformly bounded, and satisfies* $\sigma_t := \sigma(t/T) > 0$ *for all* $t = 1, \ldots, T$*, where* $\sigma(\cdot) \in D[0,1]$*, the space of càdlàg functions on* $[0,1]$*.*

b) $\{z_t\}$ *is a martingale difference sequence (MDS) with respect to the natural filtration* $\mathfrak{F}_t$*, the sigma-field generated by* $\{z_s\}_{s\le t}$*, such that* $\mathfrak{F}_{t-1} \subseteq \mathfrak{F}_t$ *for* $t = \ldots, -1, 0, 1, 2, \ldots$ *which satisfies:*

(*i*) $E(z_t^2) = 1$,

(*ii*) $\tau_{r,s} := E(z_t^2 z_{t-r} z_{t-s})$ *is uniformly bounded for all* $t \geq 1$, $r \geq 0$, $s \geq 0$, *where also* $\tau_{r,s} > 0$ *for all* $r \geq 0$,

(*iii*) *For all integers* q *such that* $3 \leq q \leq 8$ *and for all integers* $r_1, \ldots, r_{q-2} \geq 1$, *the q'th order cumulants* $\kappa_q(t,t,t-r_1,\ldots,t-r_{q-2})$ *of* $(z_t, z_t, z_{t-r_1}, \ldots, z_{t-r_{q-2}})$ *satisfy the requirement that* $\sup_t \sum_{r_1,\ldots,r_{q-2}=1}^{\infty} \left|\kappa_q(t,t,t-r_1,\ldots,t-r_{q-2})\right| < \infty$.

Therefore, we also study the possibility of directly induced heteroskedasticity via a scaling of the noise term in addition to a large variety of, potentially asymmetric, conditionally heteroskedastic models. More concrete, part *a*) of Assumption 3.2 controls the extent of unconditional heteroskedasticity present in $\{\varepsilon_t\}$. In particular $\sigma(\cdot)$ is so flexible that it allows for one, multiple, or even periodic switching patterns in the unconditional variance as long as the number of these occurring switches remains bounded. Furthermore, piecewise defined linear, smooth, and even broken trends are permitted as well. Complementary part *b*) of Assumption 2 controls the kind of conditional heteroskedasticity present in the error terms. Specifically (*ii*) and (*iii*) of part *b*) is standard in this literature (e.g. Demetrescu et al. (2008), Hassler et al. (2009), or Kew and Harris (2009)) and is required to allow for higher-order dependence in $\{z_t\}$ which permits for example the usage of asymmetric generalized autoregressive conditional heteroskedastic (GARCH) type models. Lastly, if one instead assumes fixed parameters in the short-run polynomials in (3.1) and under Assumption 3.2, the process $\{y_t\}$ is identical to the heteroskedastic $ARFIMA(p,d,q)$ model of Cavaliere et al. (2015b) with a Wold representation in Assumption 3.1 displaying fixed coefficients.

In the remainder of this paper we are interested in testing the following pair of hypothesis:

$$H_0 : d = \bar{d} = 0 \qquad vs. \qquad H_1 : d > 0\,. \tag{3.2}$$

Under the null, assuming fixed short-run dynamics and $\{\varepsilon_t\}$ to be white noise, it is well known that the autocorrelation function (acf) of (3.1) decays at most exponentially, whereas under the alternative this quantity decays hyperbolically. If one additionally allows for deterministic switching in the short-run dynamics and/or $\{\varepsilon_t\}$ to obey Assumption 3.2a) the acf is clearly not defined. However, due to our Assumption 3.1 the sample autocorrelations will nevertheless behave in a similar fashion. Moreover, note that we are focussing on the case of a constant memory parameter. However, this can easily be extended to more general setups where $d \equiv d(\tilde{\tau})$, as long as $d(\tilde{\tau})$ is completely specified under the null such that $u_t := (1-L)^{d(\tilde{\tau})} y_t$ remains integrated of order zero.[1]

By assuming fixed parameters in (3.1) and conditionally homoskedastic Gaussian innovations, one can define the residuals as $\hat{\varepsilon}_t(\gamma) := \hat{\varepsilon}_t(d,\psi) := c(L,\psi)(1-L)^d y_t$. It is well known that the

[1] Note that allowing $d \equiv d(\tilde{\tau})$ and assuming $\{z_t\}$ in Assumption 3.2 to be white noise, results in (3.1) to coincide with the LS-ARFIMA model of Palma and Olea (2010).

concentrated log-likelihood is then given by:[2]

$$\ell(d,\psi) := -\frac{T}{2}\log\left(\hat{\sigma}^2(d,\psi)\right) = -\frac{T}{2}\log\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t(d,\psi)^2\right),$$

which leads to the restricted Quasi-Maximum Likelihood (QML) or equivalently the Conditional Sum-of-Squares (CSS) estimator:

$$\tilde{\psi} := \arg\max_{\psi\in\Psi} \ell\left(\bar{d},\psi\right) = \arg\min_{\psi\in\Psi} \hat{\sigma}^2(\bar{d},\psi). \tag{3.3}$$

Let $\tilde{\gamma} = (\bar{d},\tilde{\psi}')'$ denote the QML estimator of γ under the null which is shown to be consistent by Cavaliere et al. (2017) if $\{y_t\}$ displays fixed parameters and heteroskedasticity of the form given in Assumption 3.2. Consider the score vector given by $D_T(\gamma) := \partial\ell(\gamma)/\partial\gamma$ and by further defining $H_T(\gamma) := \partial^2\ell(\gamma)/\partial\gamma\partial\gamma'$ as the Hessian matrix of the concentrated log-likelihood, it is well known from the literature (e.g. Rao (1973)) that:

$$S_{1T} := D_T(\tilde{\gamma})_1\sqrt{-H_T^{-1}(\tilde{\gamma})_{11}} \quad \overset{d}{\to} \quad N(0,1), \tag{3.4}$$

holds asymptotically under the null. This defines the LM test for the d parameter as the first element of the score vector divided by the square root of the Fisher Information evaluated at the $(1,1)$ element of the Hessian matrix.[3]

If fixed short-run dynamics are present one can define $\xi(z,\gamma) := \partial\log\left((1-z)^d c(z,\psi)\right)/\partial\gamma|_{\gamma=\gamma_0}$ and the geometric expansion $\xi(z,\gamma_0)' := \sum_{j=1}^{\infty}\xi_j z^j$ with $\xi_j = \left(-j^{-1},c_j'\right)'$ denoting the respective vector of coefficients of the considered expansion. Moreover, define:

$$\Xi := \sum_{j=1}^{\infty}\xi_j\xi_j' = \begin{bmatrix}\pi^2/6 & \kappa' \\ \kappa & \Phi\end{bmatrix},$$

with $\kappa := -\sum_{j=1}^{\infty} j^{-1}c_j$ and $\Phi := \sum_{j=1}^{\infty} c_j c_j'$, where Φ gives the Fisher information for ψ (c.f. Tanaka (1999) or Cavaliere et al. (2015b)). This allows one to define $\omega^2 := \left(\Xi^{-1}\right)_{1,1} = \left(\pi^2/6 - \kappa'\Phi^{-1}\kappa\right)^{-1}$ as the asymptotic variance of the restricted QML estimator of the first element of the score vector under conditional homoskedasticity (see also Box and Pierce (1970)). Therefore, the studentization of (3.4) crucially hinges on the correct specification of the order of the short-run polynomials of (3.1) already in a fixed parameter setup. We will analyze the finite sample impact a potential misspecification has for various LM tests in our Monte Carlo analysis presented in Section 3.2. Furthermore, following Cavaliere et al. (2015b), let $\varpi^2 := (\Xi^{-1}(\sum_{j,k=1}^{\infty}\xi_j\xi_k'\tau_{j,k})\Xi^{-1})_{1,1}\lambda$ denote the asymptotic variance of the restricted QML estimator of the memory parameter under Assump-

[2] In the remainder of this Subsection we adopt the notation of Cavaliere et al. (2015b) for reasons of convenience.
[3] One can alternatively consider the square of above statistic which is then of course $\chi^2(1)$ distributed in the limit.

tion 3.2 which ultimately leads to:

$$S_{1T} \xrightarrow{d} \left(\lambda \frac{\varpi^2}{\omega^2}\right)^{1/2} N(0,1), \quad \text{with} \quad \lambda := \left(\int_0^1 \sigma^4(s)ds\right) / \left(\int_0^1 \sigma^2(s)ds\right)^2 , \tag{3.5}$$

asymptotically under the null if $\{y_t\}$ displays fixed coefficients. Note that (3.5) is a key result of Cavaliere et al. (2015b) since it shows that heteroskedasticity of the kind given in Assumption 3.2 leads to an asymptotically non-pivotal null distribution. The coefficient λ can hereby be interpreted as a measure for the extent of unconditional heteroskedasticity present in the error terms. Note that $\lambda = 1$ occurs under conditional homoskedasticity ($\lambda \approx 1$ under conditional heteroskedasticity) and $\lambda > 1$ holds in case unconditional heteroskedasticity is present, ensured through the Cauchy-Schwarz inequality (see (3.5)). In the latter case the ratio $\frac{\varpi^2}{\omega^2} = 1$ which results in an asymptotically biased liberal test if quantiles of the standard normal distribution are used. If pure conditional heteroskedasticity, or in combination with unconditional heteroskedasticity is present, the bias is ambiguous and we refer to the simulation results in Section 3.2 concerning the exact effect in finite samples. For the sake of completeness, it shall be mentioned that in case of conditional homoskedasticity $\lambda \frac{\varpi^2}{\omega^2} = 1$ holds which breaks down to the classic result of (3.4).

3.1.1 Lagrange Multiplier Tests in the Time-Domain

Before presenting our new test we will give an overview over various existing LM tests in the time-domain, which also appear in our Monte Carlo analysis, in a chronological order.[4] In the following we omit any procedures based on a frequency-domain approach (e.g. Robinson (1994) or Lobato and Robinson (1998)) since on one hand it is well known that short-run dynamics induce high-frequency leakage already in a fixed parameter setup and on the other hand the use of a standard periodogram under unconditional heteroskedasticity instead of its localized counterpart as proposed in Dahlhaus (2000) clearly leads to invalid subsequent statistical inference for any test building on this spectral estimator (e.g. Demetrescu and Sibbertsen (2016)).
By assuming fixed parameters in (3.1) and Gaussian white noise innovations, Tanaka (1999), extending results of Robinson (1994) who proposed the equivalent frequency-domain approach, shows that:

$$S_{1T} \approx \frac{\sqrt{T}\sum_{k=1}^{T-1} \frac{1}{k}\hat{\rho}_k}{\hat{\omega}} =: Ta \quad \xrightarrow{d} \quad N(0,1)\,, \tag{3.6}$$

holds asymptotically under the null, where $\hat{\rho}_k = \frac{\sum_{j=k+1}^{T} \hat{\varepsilon}_{j-k}\hat{\varepsilon}_j}{\sum_{j=1}^{T} \hat{\varepsilon}_j^2}$ denote the estimated autocorrelation of the residuals at lag k. Note that (3.6) displays much similarity to the well known Box and

[4]Note that the tests of Tanaka (1999), Breitung and Hassler (2002), Demetrescu et al. (2008) and Cavaliere et al. (2015b) are in fact tests for the fractional integration parameter in a $ARFIMA(p,d,q)$ model. Therefore testing for long memory as considered in this paper is just a subset of their applicability since they also nest a unit root, stationarity or $d = \frac{1}{2}$ hypothesis.

Pierce (1970) statistic. The test is easily motivated by the fact that $\{y_t\}$ has short memory under the null which results in a sum of autocorrelations that are, if correctly standardized, well controllable. However, under the long memory alternative, the autocorrelations decay hyperbolically resulting in a diverging numerator in (3.6) and an asymptotically consistent test. Nevertheless, a drawback of this procedure is that Tanaka (1999) assumes omniscience of the exact kind and order of short-run components present in the process and calculates the residuals via a full parametric approach. To this end ω^2 is not estimated via a HAC estimator, but provided in a closed-form expression by Tanaka (1999) only if $\{u_t\}$ follows an AR(1) or MA(1) process, where in the former case one receives $\omega^2 = \frac{\pi^2}{6} - \frac{1-\phi^2}{\phi^2}(\log(1-\phi))^2$. Therefore, apart from requiring fixed parameters, and conditionally homoskedastic Gaussian innovations this procedure is sensitive in particular to a potential underspecification of the model.
Breitung and Hassler (2002) recast the test of Tanaka (1999) in a regression framework which notably improves its finite sample performance. Assuming conditional homoskedastic innovations and no short-run dynamics in (3.1) the test is carried out on $\hat{\varrho}$ in the following classic Dickey Fuller style regression setup:

$$u_t = \varrho u^*_{t-1} + e_t, \quad \text{with} \quad u^*_{t-1} = \sum_{j=1}^{t-1} j^{-1} u_{t-j}\,.$$

This results in a right-sided squared t-test which displays a limiting null distribution given as follows:

$$BH := \frac{(\sum_{t=2}^{T} u_t u^*_{t-1})^2}{\hat{\sigma}_e^2 \sum_{t=2}^{T} u^{*2}_{t-1}} \quad \xrightarrow{d} \quad \chi^2(1)\,. \tag{3.7}$$

Moreover, an outer product of gradient estimator is considered for estimating the Fisher Information in (3.7). In case fixed short-run dynamics are present, Breitung and Hassler (2002) assume that $\{u_t\}$ follows a stationary AR(p) process and propose to carry out the test in a two-step procedure. First, one specifies some reasonable order p and computes the corresponding residuals $\{\hat{\varepsilon}_t\}$ in a prewhitening step. Secondly, the test is then analogously carried out on $\hat{\zeta}$ in the following regression:

$$\hat{\varepsilon}_t = \zeta \hat{\varepsilon}^*_{t-1} + \sum_{i=1}^{p} \tilde{\varphi}_i u_{t-i} + \tilde{e}_t, \quad \text{with} \quad \hat{\varepsilon}^*_{t-1} = \sum_{j=1}^{t-1} j^{-1} \hat{\varepsilon}_{t-j}\,.$$

Note that the prewhitening step does not alter the limiting null distribution of the test given in (3.7). Apart from requiring fixed parameters and conditionally homoskedastic innovations in (3.1), the general approach of Breitung and Hassler (2002) for dealing with fixed short-run dynamics can be viewed as being to restrictive. The specification of the model order is solely based on a user-specific choice and not carried out by information criteria or a deterministic lag

length selection.[5] Therefore, choosing p too low results in an underspecification of the model, whereas choosing p too large leads to a loss in efficiency.
Denote in the following the numerator of the Tanaka (1999) test statistic given in (3.6) by $\tilde{N}$. By assuming fixed parameters in (3.1) and $\varepsilon_t \overset{iid}{\sim} (0,1)$, Harris et al. (2008) show that:

$$|E(\tilde{N})| = \sqrt{T}\left|\sum_{j=1}^{T-1} \frac{T-j}{jT}\gamma_j\right| \quad \to \quad \infty\,,$$

holds asymptotically under the null in case $\{u_t\}$ remains autocorrelated. Therefore, the test of Tanaka (1999) and Breitung and Hassler (2002) will spuriously reject in case an underspecification of the model occurs. To circumvent this problem they instead propose a modified version of the test which builds on a nonparametric correction and retains a standard normal limiting null distribution:

$$HCL := \frac{\hat{N}_k}{\hat{\omega}_l} = \frac{(T-k)^{1/2}\sum_{j=k}^{T-1}\frac{1}{j-k+1}\hat{\gamma}_j}{\sqrt{\sum_{j=-l}^{l} h_j \sum_{i=-l}^{l}\hat{\gamma}_i\hat{\gamma}_{i+j}}} \quad \overset{d}{\to} \quad N(0,1)\,. \tag{3.8}$$

Here, $h_0 = \pi^2/6$, $h_j = H_{|j|}/|j|$ for $j = \pm1, \pm2, \ldots$, $H_j = \sum_{i=1}^{j} i^{-1}$, $l = \lfloor(\frac{2}{3})T^{12/25}\rfloor$ and $k = (cT)^{1/2}$ for some constant $c > 0$. Therefore, instead of applying the procedure on the residuals computed via a full parametric approach, as done in the previous tests, the key intuition is to leave out the first k sample autocovariances $\hat{\gamma}_j = T^{-1}\sum_{t=1}^{T-j} u_{t-j}u_t$, when computing the numerator in (3.8). With this nonparametric correction, Harris et al. (2008) achieve robustness against the influence of fixed short-run dynamics when testing for long memory. Moreover, we conjecture that this feature will prevail to some extent in case the short-run polynomials in (3.1) exhibit some form of deterministic switching as discussed in our Assumption 3.1 and provide respective simulation results in Section 3.2. Lastly, note that in line with the previous introduced procedures, Harris et al. (2008) do not allow for any kind of conditional or unconditional heteroskedasticity as discussed in Assumption 3.2.
Demetrescu et al. (2008) extend the approach of Breitung and Hassler (2002) in several ways. Instead of restricting $\{u_t\}$ to follow an AR(p) process they assume $u_t = \sum_{j=0}^{\infty} b_j\varepsilon_{t-j}$ to follow a general linear process with s-summable fixed Wold coefficients in the sense of our Assumption 3.1 and with innovations $\{\varepsilon_t\}$ forming a MDS with absolutely summable eighth-order cumulants. Therefore, they allow for various kinds of fixed short-run dynamics and even, possibly asymmetric, conditional heteroskedasticity.[2] Contrary to Breitung and Hassler (2002) who advocate a two-step procedure encompassing a fixed autoregressive order p and prewhitening setup, they

[5] Complementary, Demetrescu et al. (2008) argue that applying model selection criteria influences the finite sample distribution of $\hat{\zeta}$ and that of the corresponding squared t-test (see also Leeb and Pötscher (2005)).

[2] Note that the assumptions of Demetrescu et al. (2008) in fact do not cover the full extent of models considered in Assumption 3.2b) (see also Cavaliere et al. (2015b))

show that:

$$u_t = \varsigma u_{t-1}^* + \sum_{i=1}^{p} a_i u_{t-i} + \tilde{\varepsilon}_t, \quad \text{with} \quad t = p+1, \ldots, T \ ,$$

$$\hat{s}(\hat{\varsigma}) = \left[\left(\sum_{t=p+1}^{T} V_{tp} V_{tp}' \right)^{-1} \left(\sum_{t=p+1}^{T} V_{tp} V_{tp}' \hat{\varepsilon}_t^2 \right) \left(\sum_{t=p+1}^{T} V_{tp} V_{tp}' \right)^{-1} \right]_{11}^{1/2} ,$$

$$ALM := \frac{\hat{\varsigma}}{\hat{s}(\hat{\varsigma})} \ \xrightarrow{d} \ N(0,1) \ , \tag{3.9}$$

holds asymptotically under the null, with $V_{tp} = (u_{t-1}^*, u_{t-1}, \ldots, u_{t-p})'$, $p = O(T^{\kappa})$ for $p \to \infty$ as $T \to \infty$ and $\kappa \in (\frac{1}{2s}, \frac{1}{4})$ with $s > 2$. Therefore, their Dickey Fuller style regression contains a sample size dependent lag augmentation that automatically corrects for the influence of any kind of fixed short-run dynamics. Since p diverges at a slower speed than the transformation stabilizing rate of the test statistic any hereby introduced bias will vanish asymptotically. Lastly, White standard errors (c.f. White (1980)) are employed to compute $\hat{s}(\hat{\varsigma})$ which provides robustness against conditional and unconditional heteroskedasticity.[6] Nevertheless the assumptions of Demetrescu et al. (2008) do not allow for any deterministic switching of the short-run dynamics in (3.1) as discussed in our Assumption 3.1.

Assuming fixed parameters in (3.1) with innovations obeying Assumption 3.2, Cavaliere et al. (2015b) propose a Wild Bootstrap procedure to bootstrap the non-pivotal null distribution of S_{1T} given in (3.5). The algorithm proceeds as follows:[7]

1.) Estimate model (3.1) assuming fixed parameters for a given order p and q using the restricted QML estimator (3.3), yielding $\tilde{\gamma} = (\bar{d}, \tilde{\psi}')'$, together with the corresponding residuals, $\tilde{\varepsilon}_t := \hat{\varepsilon}_t(\bar{d}, \tilde{\psi})$.

2.) Compute the centered residuals $\tilde{\varepsilon}_{c,t} := \tilde{\varepsilon}_t - T^{-1} \sum_{i=1}^{T} \tilde{\varepsilon}_i$ and construct the bootstrap errors $\varepsilon_t^* := \tilde{\varepsilon}_{c,t} w_t$, where w_t for $t = 1, \ldots, T$ is an iid sequence, with $E(w_t) = 0$, $E(w_t^2) = 1$ and $E(w_t^4) < \infty$.

3.) Construct the bootstrap sample $\{y_t^*\}$ from:

$$y_t^* = (1-L)^{-\bar{d}} u_t^*, \quad u_t^* = c(L, \tilde{\psi})^{-1} \varepsilon_t^*, \quad t = 1, \ldots, T \ .$$

4.) Using $\{y_t^*\}$, compute the bootstrap test statistic S_{1T}^* and define the corresponding p-value as $P_T^* := 1 - G_{1T}^*(S_{1T})$, where $G_{1T}^*(\cdot)$ denotes the conditional (on the original data) cumulated

[6] The test of Demetrescu et al. (2008) can be viewed as a particular example of various time-domain based tests employing White standard errors to robustify statistical inference against conditional and unconditional heteroskedasticity in this context as thoroughly discussed in Kew and Harris (2009).

[7] Note that Cavaliere et al. (2015b) propose a second Wild Bootstrap procedure building on the square of (3.5). However, since the authors explicitly propose the above described bootstrap in case of a one-sided alternative like in (3.2), we focus in the following on this version.

distribution function of S^*_{1T} which is approximated through numerical simulation.

5.) The Wild Bootstrap test for H_0 versus H_1 given in (3.2), rejects at the nominal significance level α if $P^*_T \leq \alpha$.

Cavaliere et al. (2015b) show under some regularity conditions that the Wild Bootstrap procedure leads to a LM test with an asymptotically pivotal null distribution even in the presence of conditional and unconditional heteroskedasticity of the form considered in Assumption 3.2. Nevertheless, apart from requiring fixed short-run dynamics in (3.1) and the computational burden of this procedure, knowledge of the exact model order is assumed which is a very restrictive assumption in practice. Although Cavaliere et al. (2015b) explicitly refer to Sin and White (1996) arguing that the usual Bayesian Information Criterium (BIC) used for model specification is consistent in this context, only sparse simulation evidence is reported on this matter.

3.1.2 A new Robust Lagrange Multiplier Test for Long Memory

As becomes clear from the above discussion, none of the existing LM tests should be able to control its size well in the presence of the heteroskedastic LS process (3.1), fully satisfying Assumption 3.1 and 3.2. Moreover, even if a practitioner identifies unconditional heteroskedasticity in the data using the four existing testing procedures which build on the Residual Variance Profile (RVP) proposed by Cavaliere and Taylor (2007) (see Section 3.3), an open question remains if the identified unconditional heteroskedasticity is indirectly induced through deterministic parameter switching as in our Assumption 3.1 or directly induced by a scaling of the noise term as in Assumption 3.2a). In both cases the RVP tests should be affected equally and to the best of our knowledge no other more sophisticated procedure exists in this context. Furthermore, one could think of scenarios covered by Assumption 3.1 where only the side diagonals of the autocovariance matrix of the process switch deterministicly, whereas the main diagonal remains unaffected. Reconsidering (3.5), this is likely to have an equally disastrous effect on statistical inference.

In view of these shortcomings we propose a new LM test in the time-domain that displays robustness against all these non-stationarities as well as all forms of conditional heteroskedasticity covered by Assumptions 3.1 and 3.2, respectively, when testing the pair of hypothesis (3.2) or equivalently about $\{u_t\}$ being integrated of order zero in model (3.1). Recall that $\{u_t\}$ is short memory which implies s-summable Wold coefficients locally at any time t, except for the break points $\tilde{\tau}$. To control for, possibly deterministicly switching, short-run dynamics we adopt the idea of Harris et al. (2008) who employ a nonparametric correction (see (3.8)) and successively recast their test in a regression framework in the spirit of Breitung and Hassler (2002) to inherit a superior finite sample performance (c.f. (3.7)). Therefore, concerning our test let:

$$u^*_{t-q-1} = \sum_{j=q+1}^{t-1} \frac{1}{j-q} u_{t-j} \, ,$$

so simply skip $u_{t-1},\dots,u_{t-q}$ which are correlated with u_t, even with correlation depending on $\tilde{\tau}$ and compute:

$$\hat{S}_q^* := \frac{1}{\hat{\omega}\sqrt{T}} \sum_{t=q+2}^{T} u_t u_{t-q-1}^* , \tag{3.10}$$

$$\hat{\omega} = \left(\sum_{j=-q_T}^{q_T} \tilde{\kappa}(x)\hat{\gamma}_j^* \right)^{1/2} .$$

For MA(q) with $q < \infty$ processes, the main idea is that $u_t u_{t-q-1}^*$ has zero mean under the null, so a central limit theorem for dependent, heterogeneous processes may apply to obtain limiting normality of the corresponding sample averages. Under Assumption 3.1 with an MA(∞) process displaying local Wold coefficients, this is only approximately the case. To allow for pivotal asymptotics, we shall let $q = q_T$ go to infinity with suitable rates (see Harris et al. (2008)). Robustness is then obtained by ensuring that the standard error estimate $\hat{\omega}$ converges to the correct limit. Harris et al. (2008) provide a closed-form expression of this entity, but rely on conditional homoskedastic innovations and fixed short-run dynamics. However, in order to allow for the full extent of heteroskedasticity covered by Assumption 3.2, we use a standard HAC estimator in the style of Andrews (1991) to estimate ω in (3.10) where any kernel $\tilde{\kappa}(x)$ with $x \in \mathbb{R} \to [-1, 1]$, fulfilling the regularity conditions of Andrews (1991) may be applied and where $\hat{\gamma}_j^*$ denotes the jth-order sample autocovariance of $u_t u_{t-q-1}^*$. The procedure is then equivalent to regressing u_t on u_{t-q-1}^* and using robust standard errors to compute t-ratios, so it stays a regression-based test. Therefore, let $\{y_t\}$ satisfy Assumption 3.1 as well as 3.2 and, provided that $q = q_T = CT^{\kappa^*}$ for some $\frac{1}{s-1} < \kappa^* < \frac{1}{2}$ with $C > 0$ holds, one should asymptotically obtain under the null:

$$\hat{S}_q^* \xrightarrow{d} N(0, 1) . \tag{3.11}$$

We investigate the finite sample performance of (3.10) compared to the other introduced LM tests via an extensive Monte Carlo simulation study in Section 3.2 and obtain results that clearly justify (3.11). Moreover, we discuss the impact the choice of kernel and bandwidth selection has on the size and power of our test and give some recommendations concerning its practical implementation.

3.2 Monte Carlo Simulations

In order to analyze the finite sample performance of our new robust LM test ($\hat{S}_q^*$) in comparison to the other tests introduced in Subsection 3.1.1, we conduct an extensive Monte Carlo simulation study. Throughout this section we report results based on $M = 10{,}000$ replications, consider sample sizes of $T \in \{250, 500, 1000, 2000\}$ typically found in Finance data and choose a nominal significance level of $\alpha = 5\%$.

Let the data generating process (DGP) $\{y_t\}$ for $t = 1, \ldots, T$, be given by:

$$(1-L)^d y_t = \frac{(1+\theta(\tilde{\tau})L)}{(1-\phi(\tilde{\tau})L)}\varepsilon_t, \quad \text{where} \quad \varepsilon_t = \sigma_t z_t \quad \text{and} \quad z_t \overset{iid}{\sim} N(0,1)\,, \tag{3.12}$$
$$\sigma_t^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{I}(t \geq \lfloor \tau T \rfloor)\,,$$
$$\tau \in \{1/4, 3/4\}, \quad \text{with} \quad \nu := \sigma_1/\sigma_0 \quad \text{and} \quad \nu \in \{1/3, 1, 3\}\,.$$

As becomes clear the DGP follows the heteroskedastic LS process (3.1) with model order $(p,q) = (1,1)$ and satisfies Assumption 3.1 and 3.2a). We first focus on the performance of the tests under the null and therefore set $d = 0$ which results in $y_t = u_t \sim I(0)$ and, concerning the first set of simulations, choose the remaining parameters in (3.12) as follows: $\tilde{\tau} = 0$ which refers to fixed short-run dynamics with ϕ and $\theta \in \{-0.5, 0, 0.5\}$. Note that this setting results in (3.12) coinciding with the heteroskedastic $ARFIMA(1, d = 0, 1)$ model of Cavaliere et al. (2015b) governing a potential single break in the unconditional variance occurring at relative sample size τ. The parameter values for ν and τ are inspired by real data observed during the so-called "Great Moderation" at the beginning of the 1990's as well as during the Great Financial Crisis (2007-2008) and are directly taken from these authors.[8]

We implement the test of Tanaka (1999) from (3.6) in its plain version (Ta) which assumes no short-run dynamics present in the DGP, as well as the AR(1) specification for which a closed form expression exists (Ta_{AR}). Next we consider the test by Breitung and Hassler (2002) from (3.7) again in a plain version (BH) as well as the AR(1) version (BH_{AR}). The test of Harris et al. (2008) (HCL) from (3.8) is implemented by choosing a trimming bandwith $k = \lfloor (cT)^{1/2} \rfloor$ with $c = 1$ which closely follows the recommendations of these authors. To further compare the performance of our test with other procedures displaying robustness to heteroskedasticity, we first consider the augmented LM test by Demetrescu et al. (2008) (ALM) from (3.9) where the truncation lag $p_K = \lfloor K(T/100)^{1/4} \rfloor$ with $K = 4$ (see Schwert (2002)) is chosen accordingly.

In case of the Wild Bootstrap procedure of Cavaliere et al. (2015b) (WB^{LM}), we use the algorithm described in Subsection 3.1.1 utilizing 499 Bootstrap replications and closely follow the recommendations of these authors by constructing the bootstrap errors using a simple two-point distribution: $P(w_t = -1) = P(w_t = 1) = 0.5$. At this point we would like to point out that Cavaliere et al. (2015b) in most part of their paper assume omniscience of the order of short-run dynamics governing the DGP which is naturally an unrealistic assumption in applied work. Therefore, we impose the additional restriction that the order of the DGP has to be specified by the BIC prior to running the Bootstrap algorithm. We choose the order among all permutations of $p, q \in \{0, 1\}$. Recall that Cavaliere et al. (2015b) explicitly refer to Sin and White (1996) stating that the usual BIC model specification procedure is consistent in the context of a heteroskedastic $ARFIMA(p,d,q)$ model which is considered here in the following simulations. Finally, concerning our test ($\hat{S}_q^*$), we choose $\kappa^* = 1/4$ and $C = \frac{s}{100^{1/4}}$ with $s \in \{4, 8\}$ (see Schwert (2002)) to additionally

[8] This type of process can also be found in the works of Kim et al. (2002), Cavaliere and Taylor (2007), Kew and Harris (2009), Harris and Kew (2017) and Demetrescu and Sibbertsen (2016), among others.

																$s=8$				$s=4$			
		Ta		Ta_{AR}		BH		BH_{AR}		HCL		ALM		WB^{LM}		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	0	0	2.15	0	100	100	5.34	91.52	5.04	5.84	2.79	2.32	4.74	4.73	5.34	4.29	5.44	4.28	7.86	4.44	8.09	4.96
	500	0	0	2.47	0	100	100	5.11	99.62	6.17	5.77	3.61	2.23	4.43	4.5	5.68	4.43	5.46	4.61	9.77	4.83	9.29	4.81
	1000	0	0	3.34	0	100	100	5.12	100	6.14	6.13	3.46	3.67	4.9	5.1	6.03	4.59	5.43	4.60	7.01	4.68	6.63	4.39
	2000	0	0	3.73	0	100	100	4.81	100	6.06	5.68	4.20	3.72	5.25	5.29	5.24	4.63	5.58	4.72	4.89	4.60	5.04	5.07
0	250	2.71		1.63		4.63		4.64		2.85		3.02		4.32		4.34		4.27		4.39		4.49	
	500	3.36		2.09		4.78		4.57		3.09		3.52		4.57		4.49		4.50		4.59		4.95	
	1000	3.73		2.91		4.81		5.23		3.63		3.65		4.51		4.58		4.62		4.84		4.73	
	2000	3.92		3.08		4.88		5.16		3.84		3.94		4.94		4.82		4.76		4.69		5.03	
0.5	250	100	100	0.64	0.14	100	99.21	4.06	12.75	3.28	3.85	3.34	2.79	5.2	3.54	4.81	4.65	5.28	4.67	5.69	4.58	6.42	4.72
	500	100	100	0.99	0.07	100	100	4.59	18.06	3.82	4.40	3.52	3.02	5.1	4.5	4.96	4.89	5.63	4.82	6.16	4.70	6.11	4.67
	1000	100	100	1.65	0.05	100	100	4.67	29.36	4.62	4.81	3.63	3.70	4.75	4.6	5.02	4.60	4.97	4.94	5.31	4.55	5.28	4.88
	2000	100	100	2.18	0.01	100	100	4.69	46.97	4.70	5.09	3.91	4.11	4.91	4.81	4.74	4.81	5.15	5.49	5.14	4.84	5.25	5.26

Table 3.1: Reported are the rejection frequencies (Size) for testing $d = 0$ in (3.12) for various LM tests under conditional homoskedasticity ($\nu = 1$) for $\alpha = 5\%$. The DGP is generated with fixed parameters ($\tilde{\tau} = 0$), various sample sizes T and short-run dynamics ϕ and θ present solely in the autoregressive (ar) or the moving average (ma) polynomial, respectively.

analyze the effect of the bandwidth selection and concerning $\hat{\omega}$, we implement a HAC estimator with Quadratic Spectral ($\hat{S}^*_{q,QS}$) and Bartlett ($\hat{S}^*_{q,B}$) kernel (c.f. Andrews (1991)) since these are the most frequently used kernels in practice.

We start our analysis by setting $\nu = 1$ in (3.12), resulting in a classic $ARMA(p,q)$ model with conditionally homoskedastic Gaussian innovations. On one hand this creates a situation where all considered testing procedures are expected to show a good size control, provided a correct parametric specification is carried out. On the other hand it offers a direct comparison of our test with the HCL test of Harris et al. (2008) in a scenario where the additional robustness features against heteroskedasticity are in fact not needed. The results are reported in Table 3.1. We observe that the Ta (Ta_{AR}) test controls its size if the model order is correctly specified. However, if the order is under- or misspecified, its size deteriorates, whereas an overspecification results in a loss in efficiency. The BH (BH_{AR}) test shows similar results, but with a superior size control in the correct specification cases compared to the test of Tanaka (1999). This demonstrates the benefits of the tests regression style approach as pointed out by Breitung and Hassler (2002), but also shows the drawback a wrong model specification has on parametric tests already in a fixed parameter setup. On the contrary the HCL test shows apart from minor deviations, a good size control across all scenarios clearly advocating the nonparametric correction idea of Harris et al. (2008). The ALM test controls its size, albeit conservative, throughout all scenarios, whereas the WB^{LM} procedure gives globally a very good performance at the cost of an extremely high computation time due to its bootstrap nature. Lastly, our test displays a very good size control in all setups already in small sample sizes independently of the choice of kernel and bandwidth trimming. It is at least on par with the WB^{LM} procedure and has a much lower computation

$\tau = \frac{1}{4}$ and $\nu = \frac{1}{3}$

																$s=8$				$s=4$			
		Ta		Ta_{AR}		BH		BH_{AR}		HCL		ALM		WB^{LM}		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	0	0	7.13	0	99.82	99.99	17.30	84.51	12.29	13.72	2.67	2.32	4.79	4.24	4.71	4.33	4.69	4.20	6.81	4.18	6.46	4.25
	500	0	0	9.25	0	100	100	18.52	96.74	14.30	13.92	2.84	1.92	4.01	4.91	5.08	4.43	5.33	4.46	8.17	4.70	8.12	5.06
	1000	0	0	10.03	0	100	100	19.07	99.86	14.93	14.65	2.88	2.68	4.38	4.83	5.47	4.98	5.23	4.41	6.68	4.62	6.27	4.70
	2000	0	0	11.40	0	100	100	20.02	100	15.28	14.75	3.47	3.20	3.88	5.19	5.37	4.70	5.49	4.65	5.47	4.79	5.13	4.62
0	250	9.47		5.48		18.17		16.61		8.72		2.78		2.06		4.05		3.69		3.96		3.90	
	500	10.41		7.31		19.16		17.98		9.62		2.87		2.43		4.28		4.18		4.42		4.33	
	1000	11.74		8.70		19.29		18.66		10.70		3.07		2.72		4.67		4.47		4.71		4.74	
	2000	12.31		10.32		19.84		19.06		11.75		3.18		2.79		4.59		4.88		4.75		4.77	
0.5	250	99.94	96.21	1.40	0.52	99.94	95.21	13.01	22.47	8.70	9.97	3.07	2.73	4.07	2.21	4.49	4.02	4.55	4.18	4.84	4.40	5.24	4.68
	500	100	99.94	3.08	1.00	100	99.86	15.29	29.95	10.79	11.88	2.97	2.84	5.11	3.07	4.64	4.28	5.10	4.71	4.96	4.50	4.59	4.30
	1000	100	100	5.37	0.86	100	100	16.35	38.58	12.01	13.63	2.69	2.89	4.84	3.38	4.62	4.70	5.34	5.00	4.50	5.08	5.15	4.80
	2000	100	100	7.57	0.51	100	100	17.93	51.04	13.09	13.61	3.11	3.32	5.00	3.69	4.67	4.73	4.51	4.75	4.95	4.97	4.59	4.57

$\tau = \frac{3}{4}$ and $\nu = \frac{1}{3}$

																$s=8$				$s=4$			
		Ta		Ta_{AR}		BH		BH_{AR}		HCL		ALM		WB^{LM}		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	0	0	2.92	0	99.99	100	7.03	89.67	6.81	7.79	2.91	2.26	4.86	4.91	5.12	4.21	5.28	4.62	8.11	4.45	7.41	4.65
	500	0	0	3.91	0	100	100	7.67	99.18	8.72	8.15	3.20	2.37	4.59	4.76	5.34	4.35	5.37	4.71	9.20	4.99	9.12	4.77
	1000	0	0	4.67	0	100	100	7.35	100	7.73	8.26	3.13	2.85	5.1	4.62	5.49	4.91	5.49	4.85	6.59	4.92	6.80	4.31
	2000	0	0	5.12	0	100	100	7.57	100	8.72	8.21	3.40	3.46	4.55	4.86	5.42	4.78	5.24	4.70	4.81	4.69	5.21	4.74
0	250	3.80		1.94		7.74		7.21		3.92		2.98		3.71		4.30		4.28		4.50		4.53	
	500	4.76		3.03		7.68		7.43		4.34		3.09		4.42		4.56		4.67		4.77		4.60	
	1000	5.6		3.83		8.05		7.64		4.96		3.24		4.55		5.20		4.80		4.59		4.78	
	2000	5.83		4.77		8.15		7.61		5.68		3.56		4.53		4.69		4.85		4.74		4.76	
0.5	250	100	98.97	0.53	0.21	100	98.61	6.45	15.41	4.20	5.57	2.98	2.87	5.06	3.19	4.72	4.32	5.03	4.91	5.53	4.90	5.60	4.54
	500	100	100	1.46	0.21	100	99.97	6.47	20.54	4.88	5.93	3.04	2.78	5.04	4.05	5.28	4.24	5.18	4.89	6.14	4.68	5.89	4.86
	1000	100	100	2.52	0.14	100	100	6.97	30.43	6.23	6.92	3.27	3.14	5.16	4.1	4.88	5.09	5.20	4.72	5.37	4.61	5.35	4.98
	2000	100	100	3.14	0.02	100	100	8.02	47.44	7.25	6.76	3.82	3.58	4.65	4.5	4.71	4.68	4.84	5.03	5.12	4.81	5.12	4.97

$\tau = \frac{1}{4}$ and $\nu = 3$

																$s=8$				$s=4$			
		Ta		Ta_{AR}		BH		BH_{AR}		HCL		ALM		WB^{LM}		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	0	0	2.75	0	100	100	7.11	89.26	7.32	7.74	3.13	2.21	4.53	4.6	5.20	4.72	5.14	3.78	7.84	4.87	7.42	4.75
	500	0	0	3.77	0	100	100	8.05	99.26	8.35	8.15	3.56	2.15	4.2	4.65	5.64	4.66	5.56	4.34	8.84	4.93	9.02	4.87
	1000	0	0	4.46	0	100	100	7.38	100	8.61	8.75	3.45	3.17	4.39	5.13	6.17	4.68	5.87	4.81	6.83	4.84	6.68	4.76
	2000	0	0	5.06	0	100	100	7.71	100	8.41	8.10	3.83	3.29	4.5	4.96	5.73	4.83	5.53	4.30	5.06	4.80	4.73	4.81
0	250	3.95		2.21		6.92		7.40		3.86		2.88		3.87		4.09		4.18		4.24		4.23	
	500	4.39		3.08		7.22		7.83		4.46		3.20		3.77		4.45		4.36		4.63		4.42	
	1000	5.07		4.21		7.72		8.22		5.50		3.35		4.57		4.64		4.56		4.78		4.86	
	2000	5.59		4.65		7.85		7.68		5.37		3.66		4.66		4.79		4.96		4.82		4.60	
0.5	250	99.99	98.86	0.81	0.21	99.98	98.51	6.97	15.74	4.54	5.46	3.03	2.44	5.35	2.74	4.71	4.27	5.20	4.27	5.02	4.59	5.54	4.86
	500	100	99.99	1.82	0.19	100	99.99	7.04	22.05	5.52	6.22	3.40	2.75	4.69	4.01	5.38	4.82	5.33	5.01	5.58	5.21	6.16	5.14
	1000	100	100	2.97	0.08	100	100	7.00	32.16	5.83	6.44	3.28	3.02	4.61	4.2	4.93	4.50	5.02	4.66	5.45	5.32	5.60	5.20
	2000	100	100	3.89	0.02	100	100	7.65	47.72	6.85	7.29	3.34	3.67	5.19	4.44	5.27	5.00	5.25	5.13	4.64	4.74	5.31	5.22

$\tau = \frac{3}{4}$ and $\nu = 3$

																$s=8$				$s=4$			
		Ta		Ta_{AR}		BH		BH_{AR}		HCL		ALM		WB^{LM}		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	0	0	7.45	0.01	99.81	100	17.55	84.37	11.93	12.96	3.14	2.54	4.38	4.5	4.67	4.05	4.64	4.16	6.52	4.75	6.28	4.31
	500	0	0	8.52	0	100	100	18.85	96.91	14.37	14.24	3.10	2.09	4.39	4.89	5.41	4.72	5.11	4.26	7.67	4.65	7.77	4.71
	1000	0	0	10.49	0	100	100	19.06	99.89	14.91	14.57	2.95	2.46	3.96	5.1	5.20	4.67	5.34	4.65	6.53	4.82	6.08	4.66
	2000	0	0	11.18	0	100	100	19.47	100	15.75	15.03	3.18	3.15	3.74	5.06	5.52	4.80	5.65	4.94	5.42	4.64	5.14	4.96
0	250	9.22		5.22		18.97		17.88		8.71		3.21		2.35		4.10		3.67		4.18		3.91	
	500	10.54		7.83		19.56		18.63		10.07		2.97		2.64		4.27		4.11		4.23		4.38	
	1000	11.47		9.26		19.79		19.39		11.18		3.05		2.65		4.47		4.28		4.59		4.53	
	2000	12.37		10.64		19.69		20.16		11.68		3.21		3.23		4.58		4.67		4.85		4.49	
0.5	250	99.95	96.09	1.80	0.42	99.98	95.23	15.78	26.61	8.62	10.76	3.38	3.22	4.68	2.26	4.99	4.32	5.01	4.12	4.40	4.17	5.10	4.14
	500	100	99.92	3.37	1.03	100	99.90	16.95	31.10	10.79	12.47	3.13	2.65	4.92	2.9	4.74	4.16	5.27	4.51	4.59	4.23	4.80	4.42
	1000	100	100	6.41	1.04	100	100	17.37	40.41	12.16	12.43	3.09	2.70	4.93	3.92	5.18	4.44	5.16	4.97	4.26	4.44	4.66	4.62
	2000	100	100	7.68	0.68	100	100	18.17	52.03	13.07	13.68	3.35	3.18	4.78	3.78	4.79	4.48	5.02	4.87	4.34	4.64	4.54	4.68

Table 3.2: Reported are the rejection frequencies (Size) for testing $d = 0$ in (3.12) for various LM tests under directly induced unconditional heterokedasticity. The DGP is generated with fixed parameters ($\tilde{\tau} = 0$), various sample sizes T, $\tau \in \{1/4, 3/4\}$, $\nu \in \{1/3, 3\}$ and short-run dynamics ϕ and θ present solely in the autoregressive (ar) or the moving average (ma) polynomial, respectively.

time. Moreover, we observe that our test shows a better size control than the HCL test of Harris et al. (2008) which is an interesting additional result in this classic setup and emphasizes the benefits of our idea of recasting the HCL test into a regression framework in the style of Breitung and Hassler (2002).

Next we ceteris paribus set $\tau \in \{1/4, 3/4\}$ and $\nu \in \{1/3, 3\}$ in (3.12) which results in the DGP displaying directly induced unconditional heteroskedasticity and report our findings in Table 3.2. We observe throughout that the Ta (Ta_{AR}), BH (BH_{AR}) and HCL test's size control deteriorates already in small sample sizes, leading in the majority of the cases to a liberal asymptotic bias to an extent that no correct statistical inference is possible anymore. This finding is supported by the theoretical results of Cavaliere et al. (2015b) (see (3.5)). The effect is hereby more pronounced in case the length of the regime with the large variance is small compared to the length of the regime with the small variance. As expected the robust tests of Demetrescu et al. (2008) (ALM), Cavaliere et al. (2015b) (WB^{LM}) and our $\hat{S}_q^*$ test still control their size. Nevertheless, the ALM test remains conservative and the WB^{LM} procedure shows similar behavior in particular in those cases where no short-run dynamics are present. We conjecture that this problem is an artifact of the model selection step conducted prior to running the bootstrap algorithm. Since Cavaliere et al. (2015b) do not discuss any further options in this context, this problem remains a drawback of the WB^{LM} procedure and is left for future research. However, among these robust procedures, our test displays the best performance throughout.[3]

To additionally analyze the effect of conditional heteroskedasticity, we consider the following set of models which ceteris paribus replace $\{\varepsilon_t\}$ in equation (3.12), respectively:

Model A: $\varepsilon_t = z_t = h_t^{1/2} e_t, \quad h_t = 0.1 + 0.5 z_{t-1}^2,\ e_t \sim N(0,1)\,.$

Model B: $\varepsilon_t = z_t = h_t^{1/2} e_t, \quad h_t = 0.1 + 0.5 z_{t-1}^2,\ e_t \sim (3/5)^{1/2} t_5\,.$

Model C: $\varepsilon_t = z_t = h_t^{1/2} e_t, \quad h_t = 0.1 + 0.2 z_{t-1}^2 + 0.79 h_{t-1},\ e_t \sim N(0,1)\,.$

Model D: $\varepsilon_t = z_t = h_t^{1/2} e_t, \quad h_t = 0.1 + 0.2 z_{t-1}^2 + 0.79 h_{t-1},\ e_t \sim (3/5)^{1/2} t_5\,.$

Model E: $\varepsilon_t = z_t = h_t^{1/2} e_t, \quad \log(h_t) = -0.23 + 0.9 \log(h_{t-1}) + 0.25\left(e_{t-1}^2 - 0.3 e_{t-1}\right),\ e_t \sim N(0,1)\,.$

Model F: $\varepsilon_t = z_t = h_t^{1/2} e_t, \quad h_t = 0.0216 + 0.6896 h_{t-1} + 0.3174 (z_{t-1} - 0.1108)^2,\ e_t \sim N(0,1)\,.$

Model G: $\varepsilon_t = z_t = h_t^{1/2} e_t, \quad h_t = 0.005 + 0.7 h_{t-1} + 0.28(|z_{t-1}| - 0.23 z_{t-1})^2,\ e_t \sim N(0,1)\,.$

Model H: $\varepsilon_t = z_t = e_t \exp(h_t), \quad h_t = 0.936 h_{t-1} + 0.5 v_t,\ (v_t, e_t) \sim N(0, diag(\sigma_v^2, 1)),\ \sigma_v = 0.424\,.$

Model I: $\varepsilon_t = \sigma_t z_t, \quad \sigma_t = 1 + 3\mathbb{I}(t \geq \lfloor 0.75 \rfloor T),\ z_t = h_t^{1/2} e_t,\ h_t = 0.1 + 0.5 z_{t-1}^2,\ e_t \sim N(0,1)\,.$

Models A-H display parameter specifications based on applied work (c.f. Gonçalves and Kilian (2004)) and are also considered in Cavaliere et al. (2015b). Model A and B (C and D) are standard stationary ARCH(1) (GARCH(1,1)) models, displaying Gaussian and t_5-distributed shocks with unit variance, respectively. Model E is the EGARCH(1,1) model of Nelson (1991), Model F is the AGARCH(1,1) model of Engle (1990) and Model G is the GJR-GARCH(1,1) model of Glosten et al. (1993). These three models contribute asymmetric conditional heteroskedasticity to our study. Model H is a first-order autoregressive stochastic volatility model, whereas

[3]Note that in additional unreported results we also analyzed the effect of a break occurring in the middle of the sample ($\tau = 1/2$), different break magnitudes, switching patterns and the variance displaying a piecewise linear trend. All these settings, satisfying Assumption 3.2a), yielded qualitatively similar results.

	T	Ta	Ta_{AR}	BH	BH_{AR}	HCL	ALM	WB^{LM}	$s=8$: $\hat{S}^*_{q,QS}$	$s=8$: $\hat{S}^*_{q,B}$	$s=4$: $\hat{S}^*_{q,QS}$	$s=4$: $\hat{S}^*_{q,B}$
Model A	250	8.82	1.98	17.19	8.29	2.39	2.85	2.58	3.91	4.21	4.15	4.16
	500	10.55	3.63	18.56	9.35	2.87	3.14	2.28	4.48	4.43	4.33	4.54
	1000	12.37	4.57	19.71	9.97	3.53	3.7	2.92	4.52	4.73	4.91	4.85
	2000	13.51	5.96	21.82	11.56	3.48	3.83	3.21	4.66	4.99	5.01	4.98
Model B	250	10.65	1.93	22.18	9.04	2.35	2.9	2.23	3.77	4.07	4.25	4.22
	500	13.34	3.5	26.11	10.91	2.67	3.42	2.2	4.63	4.48	4.3	4.33
	1000	16.11	5.33	30.04	12.5	3.21	3.57	2.56	4.5	4.52	4.48	4.56
	2000	18.11	7.27	34.04	14.63	3.19	3.7	2.72	4.62	4.67	4.79	4.93
Model C	250	8.05	4.14	15.14	13.61	5.25	3.15	2.57	4.36	4.09	4.28	4.16
	500	10.94	6.95	19.46	18.35	6.87	3.08	2.39	4.26	4.25	4.45	4.13
	1000	14.29	10.56	25.42	23.38	7.83	3.4	3.01	4.48	4.55	4.39	4.74
	2000	16.75	14.47	31.64	29.26	9.01	3.36	3.2	4.78	4.46	5.08	4.85
Model D	250	8.52	5.02	17.77	15.3	5.39	3.02	2.56	3.99	3.91	4.06	4.14
	500	12.03	8.22	23.74	21.81	6.67	3.4	2.86	4.43	4.1	4.36	4.29
	1000	16.18	12.51	31.04	27.92	7.82	3.46	3.37	4.23	4.38	4.53	4.22
	2000	19.74	15.96	38.84	34.94	8.46	3.81	3.53	4.75	4.53	4.42	4.25
Model E	250	10.1	4.84	21.28	15.77	3.44	2.48	1.9	3.69	3.89	4.11	3.94
	500	13.14	7.03	24.94	18.36	3.36	3.4	1.64	4.4	4.25	4.04	4.27
	1000	16.17	9.96	29.18	22.76	3.51	3.36	3.02	4.26	4.43	4.59	4.58
	2000	17.2	12.96	32.66	25.76	3.39	3.81	3.24	4.74	4.5	4.66	4.75
Model F	250	10.58	5.57	21.73	19.25	6.39	3.08	2.28	3.87	3.95	3.82	3.98
	500	14.8	9.92	28.93	25.8	7.69	3.19	2.89	3.94	3.85	4.3	4.39
	1000	19.08	14.55	37.54	34.35	9.32	3.36	3.01	4.45	4.45	4.37	4.3
	2000	24.22	19.47	48.06	43.92	9.93	3.47	2.86	4.85	4.55	4.29	4.32
Model G	250	8.9	4.73	19.47	17.05	5.83	2.44	2.35	3.85	3.88	3.53	3.9
	500	12.4	8.32	26.8	23.2	6.97	2.83	2.13	3.98	4.03	3.64	3.56
	1000	15.29	11.67	33.63	30.71	7.31	2.84	2.25	4.11	3.73	3.73	3.67
	2000	19.93	15.58	42.68	38.87	7.77	2.57	2.19	4.07	3.81	3.41	3.38
Model H	250	9.23	4.74	19.52	15.17	4.16	2.94	2.96	3.82	3.75	4.4	4.41
	500	11.72	7.63	22.21	18.23	4.32	3.05	3.17	4.31	4.63	4.33	4.19
	1000	13.51	9.66	24.46	21.03	4.11	3.4	2.87	4.48	4.54	4.56	4.69
	2000	14.85	11.2	25.94	21.71	3.92	3.51	2.89	4.82	4.85	4.77	4.78
Model I	250	16.02	5.84	34.11	24.59	9.8	3.14	1.55	3.73	3.47	3.98	3.79
	500	18.63	9.17	37.34	27.62	11.31	3.1	2.29	4.06	4.25	3.93	4.11
	1000	20.99	12.55	40.36	28.49	12.72	3.12	2.23	4.36	4.29	4.38	4.48
	2000	22.3	14.87	42.61	30.72	13.99	3.36	2.35	4.68	4.54	4.37	4.3

Table 3.3: Reported are the rejection frequencies (Size) for testing $d = 0$ in (3.12) for various LM tests under conditional heterokedasticity (Models A-I). The DGP is generated for various sample sizes T and no short-run dynamics.

Model I combines conditional and directly induced unconditional heteroskedasticity (see Cavaliere et al. (2015b)). Note that in the following study we do not consider the effect of fixed short-run dynamics in order to sharpen the analysis with respect to the influence of pure conditional heteroskedasticity of the form covered in Assumption 3.2b). Therefore, the DGP in (3.12) boils down to $\{y_t\} = \{u_t\} = \{\varepsilon_t\}$ for $t = 1, \ldots, T$. All results are reported in Table 3.3. Again we globally observe that the Ta (Ta_{AR}), BH (BH_{AR}) and HCL tests size control deteriorates already in small sample sizes completely invalidating correct statistical inference. The ALM test as well as the WB^{LM} procedure control their size, but nevertheless remain conservative under all considered models. Our $\hat{S}_q^*$ test, however, displays best size control already in small sample sizes and across all scenarios. So far, we solely concentrated in our simulations on directly induced unconditional and conditional heteroskedasticity, employing models primarily studied i.a. by Cavaliere et al. (2015b, 2017) which obey Assumption 3.2. In a next step, we turn to analyze the impact deterministic switching in the short-run dynamics has on the size control of all the considered LM tests. Therefore, we let the DGP in (3.12) now ceteris paribus display conditionally homoskedastic Gaussian innovations $\varepsilon_t \overset{iid}{\sim} N(0,1)$ for $t = 1, \ldots, T$ and define the following switching patterns for the short-run dynamics:

$$\phi(\tilde{\tau}) \text{ or } \theta(\tilde{\tau}) = -0.5 + 1\mathbb{I}(t \geq \lfloor \tilde{\tau} T \rfloor) \quad \text{with} \quad \tilde{\tau} \in \{0.25, 0.5, 0.75\}\,, \tag{3.13}$$

$$\phi(\tilde{\tau}) = \begin{cases} -0.5 + \tilde{\tau} \\ \frac{0.5}{1+\exp(-10+20\tilde{\tau})} \\ 0.5\sin(\pi\tilde{\tau}) \end{cases} \quad \text{with} \quad \tilde{\tau} \in (0,1]\,, \tag{3.14}$$

$$\phi(\tilde{\tau}) \equiv \phi_t = 0.9 \frac{a_t}{\max\limits_{0 \leq j \leq t} |a_j|}, \quad a_t - a_{t-1} := (1 - \phi^* L)^{-1} e_t \quad \text{with} \quad e_t \overset{iid}{\sim} N(0,1)\,. \tag{3.15}$$

As one can observe, (3.13) generates a single switch in the autoregressive or moving average coefficient which only affects the off-diagonal elements of the autocovariance matrix of the process. In contrast (3.14), which is based on the work of Kapetanios and Yates (2014), generates a continuum of switches in the main and off-diagonal elements and leads to indirectly induced heteroskedasticity. More concrete the autoregressive coefficient follows a linear, logistic, or trigonometric trend, respectively. Note that under (3.13) - (3.14) the DGP fulfills our Assumption 3.1 since the short-run dynamics switch in a deterministic fashion. In compliance with the literature, we therefore refer to these models as deterministic coefficient (DC) models. Lastly, (3.15) is the random coefficient (RC) model recently introduced by Giraitis et al. (2014), which lets the autoregressive coefficient process $\{\phi_t\}$ switch randomly by simultaneously bounding it to the stationary region. Since (3.15) does not fulfill our Assumption 3.1 due to its stochastic nature, we consider this specification only as an additional robustness check for our test and explicitly refer to Giraitis et al. (2014) for further technical details on this model type as well as its related statistical inference. Since any model specification based on the BIC criterium is not expected to work under such a DGP, we impose omniscience of the model order when

																s = 8				s = 4			
		Ta		Ta_{AR}		BH		BH_{AR}		HCL		ALM		WB^{LM}		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
$\phi(\tilde{\tau})/\theta(\tilde{\tau})$	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
DC Models																							
$-0.5+\mathbb{I}(t \geq \lfloor 0.25T \rfloor)$	250	96.44	63.01	17.62	0.7	95.51	57.93	22.37	10.23	3.74	3.76	4.03	3.91	38.47	5.89	5.07	4.6	5.26	4.75	5.89	4.43	5.92	4.89
	500	99.93	89.02	39.92	0.96	99.91	86.18	42.38	11.43	4.46	4.79	4.45	5.13	57.39	6.6	5.02	4.54	5.27	5	6.25	4.91	6.23	4.47
	1000	100	99.46	72.62	0.98	100	99.01	71.27	13.99	5.39	5.58	4.47	5.01	80.6	6.52	5.07	4.92	5.76	5.39	5.48	4.74	5.86	5.16
	2000	100	100	95.02	0.78	100	99.99	94.36	18.56	6.3	6.04	5.44	5.19	96.87	6.29	4.99	4.83	5.2	5.48	5.26	4.87	5.22	4.94
$-0.5+\mathbb{I}(t \geq \lfloor 0.5T \rfloor)$	250	39.52	5.21	56.4	2.6	36.9	11.23	58.47	10.06	4.25	4.86	5.21	5.93	75.13	7.97	4.8	4.64	5.48	4.67	6.37	4.8	6.35	5
	500	59.09	6.32	88.01	4.07	54.39	11.88	87.76	10.55	5.63	5.37	6.1	6.37	93.76	8.12	5.36	4.59	5.38	5.01	6.77	4.99	6.6	4.94
	1000	81.92	7.07	99.41	4.89	77.63	11.73	99.31	11.04	6.42	6.28	6.72	6.7	99.78	8.62	5.17	4.8	5.26	4.69	5.56	4.63	6.39	4.91
	2000	96.39	7.49	100	5.79	94.66	11.38	100	10.73	6.75	6.82	7.33	7.18	100	8.56	4.81	4.95	5.47	4.94	5.06	5.11	5.36	5.35
$-0.5+\mathbb{I}(t \geq \lfloor 0.75T \rfloor)$	250	1.51	0.02	57.35	1.8	50.35	79.14	56.92	20.45	5.35	5.51	7.42	7.4	70.97	9.5	5.29	4.59	5.68	4.89	7.2	4.62	7.05	4.82
	500	0.63	0	85.96	2.29	65.13	94.25	84	24.77	6.62	6.47	8.79	8.1	91.37	10.41	5.84	4.19	5.46	4.94	8.22	4.98	7.95	4.91
	1000	0.12	0	98.8	1.82	81.65	99.63	98.4	30.64	7.41	6.83	9.03	9.05	99.41	10.66	5.09	4.76	5.42	4.88	6	5.07	6.22	4.82
	2000	0	0	99.99	1.26	95.23	100	99.99	41.26	8.02	6.99	10.09	10.04	100	10.57	5.48	4.91	5.22	4.93	5.2	4.84	5.03	4.43

$\phi(\tilde{\tau})/\theta(\tilde{\tau})$	T	Ta	Ta_{AR}	BH	BH_{AR}	HCL	ALM	WB^{LM}	$\hat{S}^*_{q,QS}$ (s = 8)	$\hat{S}^*_{q,B}$ (s = 8)	$\hat{S}^*_{q,QS}$ (s = 4)	$\hat{S}^*_{q,B}$ (s = 4)
$\phi(\tilde{\tau}) = -0.5+\tilde{\tau}$	250	13.56	12.21	13.77	13.33	3.48	5.44	25.1	4.66	4.76	5.3	4.87
	500	20.38	24.18	17.75	23.3	4.21	5.62	36.99	5.22	5.3	5.02	5.46
	1000	30.31	45.15	25.09	40.8	4.91	6.01	54.96	5.14	5.06	5.21	5.24
	2000	46.52	73.02	38.53	67.96	5.54	6.15	80.3	5.13	5.77	5.36	5.16
$\phi(\tilde{\tau}) = \frac{0.5}{1+\exp(-10+20\tilde{\tau})}$	250	95.33	4.26	93.87	6.43	4.08	3.42	12.61	5.01	5.17	5.28	5.26
	500	99.88	8.95	99.9	10.06	4.76	4.3	17.63	5.05	5.51	5.34	5.57
	1000	100	17.78	100	16.54	5.71	4.54	26.84	5.1	5.37	5.45	5.53
	2000	100	32.57	100	28.34	6.37	5.79	39.69	4.67	5.7	5.57	5.89
$\phi(\tilde{\tau}) = 0.5\sin(\pi\tilde{\tau})$	250	99.2	1.54	98.72	5.08	3.45	3.31	9.38	5.12	5.46	5.09	5.59
	500	100	3.4	100	6.01	3.96	3.81	10.15	4.85	5.15	5.34	5.33
	1000	100	6.2	100	7.25	4.77	3.98	13.5	4.62	5.17	5.33	5.03
	2000	100	11.38	100	9.45	5.47	4.64	16.06	4.82	5.48	5.06	5.49
RC Models												
$\phi^* = 0$	250	51.12	35.23	89.11	39.97	4.65	4.09	43.89	4.98	5.84	25.59	27.31
	500	53.39	51.05	92.62	53.6	11.19	6.45	56.06	12.02	12.68	39.83	40.93
	1000	54.34	62.67	94.65	66.62	10.41	7.19	65.67	4.66	5.35	35.45	37.72
	2000	55.93	70.78	96.29	76.15	12.92	8.79	72.5	5.71	5.12	15.32	16.44
$\phi^* = 0.5$	250	52.4	39.22	90.98	45.33	4.54	4.32	45.9	4.97	6.1	31.11	33.64
	500	54.39	53.98	92.93	58.89	12.52	7.05	57.1	13.68	14.6	46	48.94
	1000	54.57	63.58	94.71	70.5	10.05	7.83	65.39	5.83	5.47	44.05	43.93
	2000	55.57	72.06	96.35	78.24	13.39	9.36	72.5	5.45	5.59	18.62	19.01

Table 3.4: Reported are the rejection frequencies (Size) for testing $d = 0$ in (3.12) for various LM tests under switching in the short-run dynamics (DC (3.13) - (3.14) and RC Models (3.15)). The DGP is generated with conditionally homoskedastic Gaussian innovations, various sample sizes T and $\phi(\tilde{\tau})$ and $\theta(\tilde{\tau})$ present solely in the autoregressive (ar) or the moving average (ma) polynomial, respectively.

conducting the WB^{LM} procedure in this study to enable a fair comparison. Our results are reported in Table 3.4. We observe that all procedures apart from our $\hat{S}_q^*$ test now fail to control their size. This is a key result of our paper since it shows that only our test is capable of conducting correct statistical inference on long memory under the heteroskedastic LS process class (3.1). Additionally, as pointed out in Subsection 3.1.2 the nature of the observed unconditional heteroskedasticity is often not clearly identifiable which effectively only leaves our test among all considered LM procedures to enable correct statistical inference under all considered scenarios whose results comprehensively support (3.11). Furthermore, even under the RC model (3.15) our test is still able to control its size if a bandwidth trimming of $s = 8$ is chosen under sample sizes $T > 500$. Generally spoken, across all these studies, we largely observe a similar performance of our test independently of the choice of kernel. However, in particular for sample sizes $T < 500$ the QS kernel offers a superior size control which is in line with the findings of Andrews (1991). Lastly, we globally find that setting $s = 8$ offers a superior size control since more autocorrelations are trimmed out when calculating $\hat{S}_q^*$ in (3.10).
In light of our findings of the previous section we now turn to analyze the power performance of our $\hat{S}_q^*$ test. Therefore, we repeat all conducted simulation studies reported in Tables 3.1 - 3.4, with respective settings concerning $\{\varepsilon_t\}$ and the pattern of the short-run dynamics, but now set $d \in \{0.2, 0.4\}$ in DGP (3.12). As one can observe all scenarios are now considered under the alternative (see (3.2)). Note that we exclude the RC model specification (3.15) since it does not satisfy our Assumption 3.1. We report all results in Tables 3.5 - 3.6 and 3.9 as well as

		$d = 0.2$								$d = 0.4$							
		$s = 8$				$s = 4$				$s = 8$				$s = 4$			
		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
-0.5	250	29.39	19.54	27.40	18.75	43.10	24.83	41.74	25.11	84.94	80.13	84.47	79.69	94.39	88.09	93.37	87.87
	500	41.99	28.92	41.18	28.74	63.98	39.74	61.84	39.46	95.80	93.21	95.66	93.73	99.58	97.99	99.40	98.05
	1000	55.27	40.75	53.19	39.53	74.87	53.33	72.55	52.99	99.23	98.55	99.21	98.65	99.93	99.84	99.96	99.81
	2000	73.17	57.93	71.86	56.39	88.24	73.35	87.24	73.85	100	99.94	99.95	99.94	100	100	100	99.99
0	250	35.69		36.76		47.88		50.07		81.08		81.88		91.31		91.96	
	500	51.01		51.25		70.60		71.05		93.79		93.66		99.10		99.20	
	1000	66.13		66.26		83.78		84.03		98.77		98.81		99.95		99.96	
	2000	83.09		82.88		96.24		95.92		99.94		99.95		100		100	
0.5	250	32.78	32.59	34.49	34.07	46.67	43.76	49.99	45.81	70.86	74.11	74.60	77.47	86.07	87.14	89.02	89.63
	500	46.73	46.88	48.88	48.13	72.96	67.26	73.94	68.13	88.41	91	90.73	91.81	97.97	98.13	98.81	98.59
	1000	63.60	61.90	63.11	62.65	84.41	82.33	85.79	82.39	97.29	98.01	97.83	97.90	99.78	99.88	99.91	99.89
	2000	81.56	81.76	81.23	81.35	96.56	95.67	96.97	95.68	99.86	99.91	99.89	99.92	100	100	100	100

Table 3.5: Reported are the rejection frequencies (Power) for testing $d = 0$ in (3.12) for the $\hat{S}_q^*$ test under conditional homoskedasticity ($\nu = 1$) for $\alpha = 5\%$. The DGP is generated with fixed parameters ($\tilde{\tau} = 0$), $d \in \{0.2, 0.4\}$, various sample sizes T and short-run dynamics ϕ and θ present solely in the autoregressive (ar) or the moving average (ma) polynomial, respectively.

$\tau = \frac{1}{4}$ and $\nu = \frac{1}{3}$

		$d=0.2$								$d=0.4$							
		$s=8$				$s=4$				$s=8$				$s=4$			
		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	22.99	15.18	22.18	14.94	31.09	18.34	29.63	17.62	72.72	66.67	72.44	65.22	82.59	73.90	82.06	74.11
	500	31.28	20.67	29.58	20.32	43.90	26.46	43.62	25.17	86.19	80.84	85.70	80.09	95.86	90.11	95.24	89.11
	1000	38.17	26.96	37.27	26.10	52.40	33.85	50.74	34.17	95.04	91.53	94.78	92.20	99.12	97.11	98.93	97.25
	2000	51.79	37.98	49.47	37.51	65.87	49.41	64.74	48.41	99.24	98.70	99.17	98.65	99.97	99.80	99.93	99.87
0	250	29.99		30.19		36.80		36.87		71.15		71.61		80.76		81.85	
	500	38.02		37.75		51.03		50.84		83.83		83.72		94.03		94.20	
	1000	47.83		48.05		62.50		62.95		92.84		93.29		98.80		98.74	
	2000	62.64		61.55		79.52		79.34		98.64		98.61		99.99		99.93	
0.5	250	28.73	28.64	31.24	30.08	35.42	34.41	38.18	34.48	65.02	65.65	66.96	69.33	76.80	76.77	79.67	78.39
	500	35.77	35.55	36.57	36.76	50.01	46.95	52.60	47.39	77.26	79.06	79.61	81.76	91.02	91.59	93.25	92.60
	1000	44	43.98	45.63	44.64	60.58	58.97	63.20	60.09	88.72	90.40	90.27	91	96.78	97.28	97.77	97.85
	2000	59.31	59.03	58.63	58.76	80.27	77.98	80.99	77.65	97.39	97.95	97.56	98.23	99.74	99.86	99.77	99.83

$\tau = \frac{3}{4}$ and $\nu = \frac{1}{3}$

		$d=0.2$								$d=0.4$							
		$s=8$				$s=4$				$s=8$				$s=4$			
		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	26.79	17.43	25.80	17.64	40.25	23.48	37.31	22.13	76.95	70.86	76.39	69.45	89.33	81.13	88.83	80.82
	500	37.07	26.36	36.91	25.76	57.13	34.73	55.42	34.25	92.17	87.78	92.05	87.21	98.65	95.57	98.54	95.20
	1000	49.90	35.58	48.68	35.38	67.49	48.13	65.80	47.47	97.92	96.78	98.06	96.43	99.78	99.43	99.81	99.44
	2000	67.24	50.50	65.87	50.53	82.35	67.53	81.91	66.38	99.83	99.71	99.87	99.79	100	100	100	100
0	250	33.65		34.20		44.08		45.38		72.39		73.31		85.19		86.44	
	500	46.91		46.17		64.73		65.01		89.04		88.81		98.11		97.95	
	1000	60.35		60.04		78.21		78		97.24		96.84		99.81		99.82	
	2000	77.17		77.44		93.06		92.78		99.81		99.80		100		100	
0.5	250	30.21	30.69	31.26	31.54	41.63	39.81	44.71	41.79	62.44	65.72	66.25	68.53	79.66	79.08	83.11	82.39
	500	42.19	42.97	43.50	43.85	65.81	59.85	67.92	60.89	81.32	84.24	84.40	85.38	96.12	96.43	97.32	96.92
	1000	56.75	56.81	56.44	57.06	78.79	76	78.53	76.61	93.97	95.48	95.15	95.86	99.46	99.50	99.53	99.70
	2000	75.56	75.14	75.63	73.84	93.81	91.94	93.80	92.07	99.47	99.60	99.49	99.66	100	100	100	100

$\tau = \frac{1}{4}$ and $\nu = 3$

		$d=0.2$								$d=0.4$							
		$s=8$				$s=4$				$s=8$				$s=4$			
		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	24.50	14.61	22.42	14.74	35.14	20.64	34.05	19.05	73.74	65.13	72.25	64.35	87.31	77.79	87.25	77.90
	500	35.27	23.33	33.53	22.73	54.82	32.90	52.06	32.31	90.26	86.06	90.65	86	98.42	94.56	98.35	95.13
	1000	46.34	32.97	45.60	31.98	65.64	45.70	62.94	45.21	98	96.23	97.72	96.30	99.84	99.34	99.86	99.40
	2000	65.26	48.52	63.13	47.35	81.43	65.07	80.59	63.77	99.87	99.78	99.87	99.59	100	100	100	99.99
0	250	28.96		29.74		40.15		40.35		66.63		68.04		82.26		84.25	
	500	42.62		43.49		61.19		61.66		87.21		86.98		97.33		97.44	
	1000	56.72		56.98		76.56		76.38		96.59		96.70		99.76		99.72	
	2000	75.59		74.98		92.24		92.04		99.83		99.71		100		100	
0.5	250	25.65	25.32	26.74	26.47	36.06	34.68	39.57	36.93	52.07	58.23	56.76	60.94	72.75	73.76	78.21	77.93
	500	37.90	38.48	38.23	39.58	61.84	56.44	63.85	58.56	76.56	81.39	80.74	82.44	94.78	94.95	96.09	96.34
	1000	53.14	52.98	54.23	54.45	76.36	73.47	76.92	73.36	93.08	93.99	93.82	95.16	99.15	99.47	99.50	99.66
	2000	72.58	72.87	72.64	72.88	93.42	91.20	92.99	90.64	99.46	99.66	99.55	99.69	99.93	99.99	99.99	100

$\tau = \frac{3}{4}$ and $\nu = 3$

		$d=0.2$								$d=0.4$							
		$s=8$				$s=4$				$s=8$				$s=4$			
		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
ϕ/θ	T	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
−0.5	250	16.14	10.49	15.53	10.24	24.73	14.01	22.80	13.55	68.04	58.58	67.37	59.22	80.58	69.22	79.35	69.43
	500	24.72	15.51	22.64	15.18	38.86	21.74	37.04	20.91	84.87	78.35	84.86	78.98	95.09	89.16	94.59	88.99
	1000	32.92	21.75	30.95	21.47	46.93	30.31	44.74	29.74	94.54	91.11	93.95	90.75	98.99	97.16	98.89	97.09
	2000	47.71	33.09	44.46	31.29	63.65	45.39	60.57	44.84	99.23	98.42	99.25	98.38	99.97	99.83	99.98	99.81
0	250	21.17		21.57		27.93		28.74		62.30		64.36		74.08		75.46	
	500	30.97		31.72		44.18		44.71		79.63		80.89		92.37		92.89	
	1000	41.28		41.02		58.09		57.13		91.74		91.69		98.38		98.38	
	2000	58.01		56.37		77.94		77.21		98.75		98.38		99.97		99.93	
0.5	250	18.43	18.70	19.70	19.89	23.74	23.52	27.03	24.97	50.79	53.93	53.84	57.17	64.47	65.74	69.18	69.98
	500	26.77	27.61	27.62	27.88	42.23	38.76	44.74	39.47	67.84	72.44	71.81	74.75	83.19	85.74	87.41	88.98
	1000	36.47	37.89	37.95	37.57	55.46	53.60	56.30	53.67	82.45	87.50	86.44	87.93	93.27	96.06	95.36	96.91
	2000	54.15	53.52	53.63	53.12	77.43	74.17	77.69	74.12	96.39	97.73	97	97.71	99.02	99.78	99.46	99.82

Table 3.6: Reported are the rejection frequencies (Power) for testing $d=0$ in (3.12) for the $\hat{S}^*_q$ test under directly induced unconditional heterokedasticity. The DGP is generated with fixed parameters ($\tilde{\tau}=0$), $d \in \{0.2, 0.4\}$, various sample sizes T, $\tau \in \{1/4, 3/4\}$, $\nu \in \{1/3, 3\}$ and short-run dynamics ϕ and θ present solely in the autoregressive (ar) or the moving average (ma) polynomial, respectively.

Figure 3.3, where the latter two are moved to the appendix for ease of exposition. Across all of these simulation studies we find that our test displays good power results already in sample sizes $T \leq 500$ which demonstrates that the nonparametric correction does not come at a too high price in terms of power. Moreover, we globally observe that power quickly increases the further one moves into the alternative and with increasing sample size where latter clearly points towards consistency of our test. Under directly induced unconditional heteroskedasticity (see Table 3.6) we find that our procedure in general displays higher power if the length of the regime with the larger variance is longer compared to the one with the smaller variance. Concerning Figure 3.3, we observe that a less complex conditionally heteroskedastic model leads to higher power results compared to the asymmetric type models. Moreover, referring to Table 3.9, higher power is achieved under smooth deterministic switching of the short-run dynamics (3.14) compared to a sudden switch (3.13).

Generally spoken, concerning the choice of kernel, both, the QS and the Bartlett kernel display very similar results for sample sizes $T > 500$. However, for $T \leq 500$ the Bartlett kernel delivers a small power advantage which is in line with the findings of all the previous size simulations. Complementary, if one chooses a bandwidth trimming of $s = 4$ one achieves a notable power boost in particular for sample sizes $T \leq 500$ since less autocorrelations are trimmed out when calculating $\hat{S}_q^*$ in (3.10). Therefore, a practitioner should consider our $\hat{S}^*_{q,QS}$ test with bandwidth trimming $s = 8$ if special emphasis is given on size control and instead our $\hat{S}^*_{q,B}$ test with $s = 4$ if highest power is to be achieved.

3.3 Empirical Analysis

In this section we analyze the memory properties of three different types of ex-post Variance Risk Premium series which can be represented by a fractional cointegration relationship between the Chicago Board of Options Exchange Volatility Index (CBOE VIX) and realized variance of the S&P500 index.

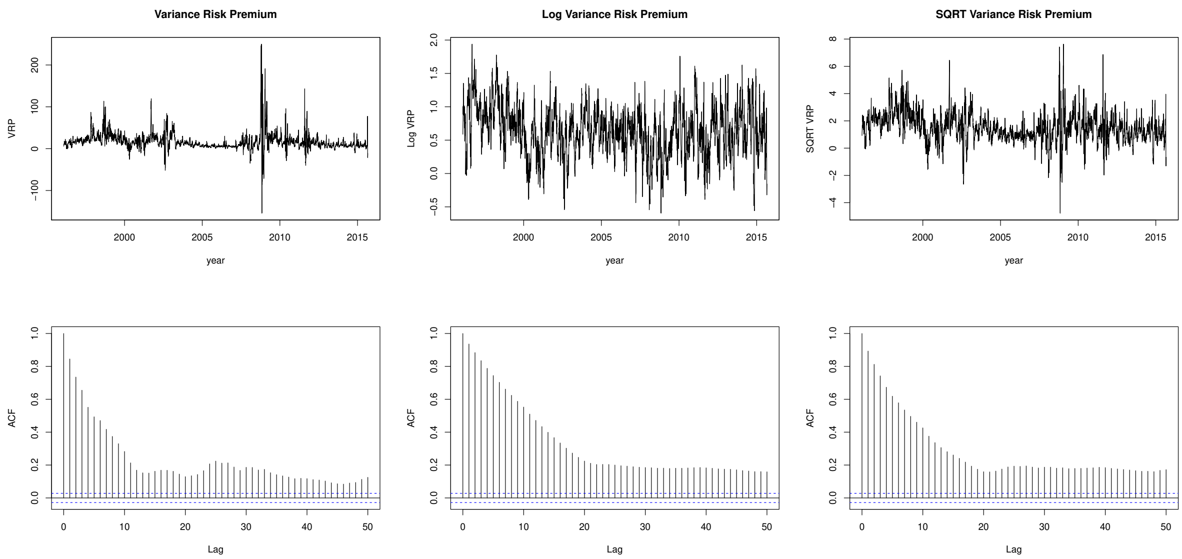


Figure 3.1: Time series plots with corresponding autocorrelation functions of ex-post Variance Risk Premium Series VRP_t, $VRP_t^{\ln}$ and $VRP_t^{\surd}$ shown in panel 1 to 3, respectively.

As is well known the CBOE VIX represents the market expectation of quadratic variation of the S&P500 over the next month, derived under the assumption of risk neutral pricing (e.g. Bekaert and Hoerova (2014)). The existence of a fractional cointegration relationship between implied and realized variance has been frequently studied in the literature i.a. by Christensen and Nielsen (2006), Bandi and Perron (2006), Nielsen (2007) and Bollerslev et al. (2013).
The raw data is obtained from Thomson Reuters Tick History Database and comprises of 5-minute log-returns of the S&P 500 Index from January 2, 1996 to August 31, 2015.[9] Closely following Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2006) we obtain the daily sampled (monthly scaled) realized variance, including close-to-open returns, as follows:

$$RV_t = \sum_{j=1}^{N} r_{t,j}^2 \,, \quad \Rightarrow \quad RV_t^{(22)} = \sum_{j=0}^{21} RV_{t-j} \,,$$

where $r_{t,j}$ defines the j-th, of in total N, intraday returns which ultimately amounts to T $= 4883$ observations in our sample. Then, we incorporate the CBOE VIX to construct three types of

[9]This data set is also studied in Kruse et al. (2016).

ex-post Variance Risk Premium (VRP) series (see Bekaert and Hoerova (2014), Bollerslev et al. (2013) and Bollerslev et al. (2009), respectively):

$$VRP_t = VIX_t^2/12 - RV_{t+22}^{(22)} ,$$
$$VRP_t^{\ln} = \ln(VIX_t^2/12) - \ln RV_{t+22}^{(22)} ,$$
$$VRP_t^{\sqrt{}} = (VIX_t/\sqrt{12}) - \sqrt{RV_{t+22}^{(22)}} .$$

Note that we do not estimate the fractional cointegration vector since the Finance literature often directly considers the above implied $[1,-1]$ setup. However, we would like to point out that the estimation results of Bollerslev et al. (2013) Co-fractional vector autoregression model clearly supports this view. Practically spoken the ex-post Variance Risk Premium corresponds up to a noise term to the ex-post payoff one receives from selling a variance swap (see Bollerslev et al. (2013)). Since this quantity is closely related to economic uncertainty and risk aversion its predictive ability for stock returns has been analyzed i.a. by Bollerslev et al. (2009), Bollerslev et al. (2013) and Bekaert and Hoerova (2014). However, whereas Bekaert and Hoerova (2014) argue that the fractional cointegration relationship between implied and realized variance is strong enough to arrive at a short memory series, Bollerslev et al. (2013) point out that some long memory remains and therefore fractionally difference the series before conducting their predictive regressions. Since returns are well known to follow a martingale difference sequence this prior step is required to ensure a balanced regression and correct subsequent statistical inference. However, if additionally heteroskedasticity is found in these series, robust long memory tests are essential to reliably decide which approach one should follow. Note that in order to conduct out-of-sample return predictions one in fact needs to use the ex-ante Variance Risk Premium series. However, since Bollerslev et al. (2013) argue that both type of series share the same memory degree, we base our following analysis on the ex-post type series.[10]

The plots of all three ex-post Variance Risk Premium series with corresponding acf's are depicted in Figure 3.1. As one can easily observe the acf's all decay very slowly in a fashion typically associated with long memory. This clearly contradicts the view of Bekaert and Hoerova (2014) who claim that the fractional cointegration relationship between implied and realized variance is strong enough to reduce the series to short memory and simultaneously supports the view of Bollerslev et al. (2013). Moreover, in all three series one can observe volatility clustering in particular around the time of the Great Financial Crisis (2007-2008) which gives a first indication for at least conditional heteroskedasticity. In a next step we follow the approach of Cavaliere et al. (2015b) and fit a heteroskedastic $ARFIMA(p,d,q)$ model to all three series and report our findings of the following analysis in Table 3.7. First, the model order was determined using the BIC criterium by considering all permutations of $p,q \in \{4,4\}$. Then we estimate all parameters using the QML estimator (see Cavaliere et al. (2015b, 2017)) and subsequently calculate the

[10]The topic which model forecasts the conditional expectation of realized variance in this context best (see also Bekaert and Hoerova (2014)) and to what extent the ensuing model estimation error biases the following analysis goes beyond the scope of this paper and is left for future research.

	$ARMA$ order (p,q)	$\hat{d}_{QML}$		$ARCH(1)$	$ARCH(5)$	$ARCH(22)$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{K}$	$\mathcal{H}_{AD}$
VRP_t	(3,4)	0.222	$Res.\ VRP_t$	**184.21****	**2093.9****	**2361.4****	**1.597***	**0.543***	**2.243****	**2.547***
		(0.015)		(6.635)	(15.086)	(40.289)	(1.355)	(0.461)	(2.009)	(2.492)
$VRP_t^{\ln}$	(3,3)	0.255	$Res.\ VRP_t^{\ln}$	**137.87****	**196.88****	**280.21****	**3.053****	**2.968****	**3.446****	**15.056****
		(0.002)		(6.635)	(15.086)	(40.289)	(1.63)	(0.743)	(2.009)	(3.85)
$VRP_t^{\surd}$	(1,1)	0.237	$Res.\ VRP_t^{\surd}$	**247.04****	**1036.7****	**1133.7****	**1.859****	**0.615***	**2.575****	**2.854***
		(0.014)		(6.635)	(15.086)	(40.289)	(1.63)	(0.461)	(2.009)	(2.492)

Table 3.7: The first panel reports the model specification results and QML estimates of the memory parameter (standard deviation given in parentheses), whereas the second panel reports the results for cond./uncond. heteroskedasticity tests applied to the residuals (Res.) of the VRP_t, $VRP_t^{\ln}$ and $VRP_t^{\surd}$ series, respectively. Bold-faced values indicate significance at the nominal 10% level; an additional * (**) indicates significance at the nominal 5% (1%) level. Respective critical values are given in parentheses.

corresponding residual series by extracting the estimated parameters.

Our results show that the memory parameter estimates clearly point towards stationary long memory in all three series. Note that these estimates are difficult to compare across other results found in the literature since manifold ways are considered how to construct this type of series. Moreover, contrary to our analysis, data windows are typically selected which focus more on the time at the beginning of the 1990's rather than including the post 2000's (e.g. Christensen and Nielsen (2006)). Nevertheless, our estimates are well in line with the results found in Nielsen (2007).

In order to test for conditional heteroskedasticity we apply the LM test for $ARCH(k)$ dynamics of Engle (1982) to the residuals of the heteroskedastic $ARFIMA(p,d,q)$ model fitted to the VRP_t, $VRP_t^{\ln}$ and $VRP_t^{\surd}$ series, respectively. We set $k \in \{1,5,22\}$ and thereby consider a daily, weekly and monthly frequency when constructing the lags in the LM regression which is carried out on the squared residuals of each series. All results are significant at a $\alpha = 1\%$ level and clearly point towards conditional heteroskedasticity being present in all three types of ex-post Variance Risk Premium series.

To further extend our analysis with respect to unconditional heteroskedasticity, we consider the sample RVP of Cavaliere and Taylor (2007) which plots $\hat{\eta}(\tilde{u}) := \hat{V}_T(\tilde{u})/\hat{V}_T(1)$ against $\tilde{u} \in [0,1]$ with $\hat{V}_T(\tilde{u}) := T^{-1}\sum_{t=1}^{\lfloor T\tilde{u}\rfloor}\hat{\varepsilon}_t^2$, where $\{\hat{\varepsilon}_t\}$ denote the residuals of the heteroskedastic $ARFIMA(p,d,q)$ model fitted to the respective ex-post Variance Risk Premium series. As these authors point out, $\hat{\eta}(\tilde{u}) \approx (\int_0^1 \sigma^2(s)ds)^{-1}\int_0^{\tilde{u}}\sigma^2(s)ds =: \eta(\tilde{u}) = \tilde{u}$ holds under conditional homoskedasticity (approximately under conditional heteroskedasticity) in large samples which results in the sample RVP lying close to the 45 degree line, whereas notable deviations point towards unconditional heteroskedasticity being present in the series (see also Cavaliere et al. (2015b)). We visualize the sample RVP of all three series in Figure 3.2 and clearly find substantial deviations from the 45 degree line in all cases.

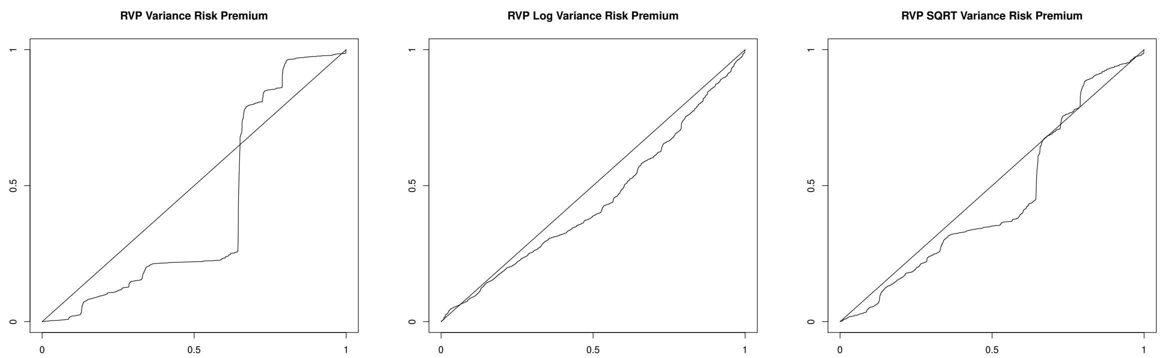


Figure 3.2: Residual Variance Profiles of the ex-post Variance Risk Premium Series VRP_t, $VRP_t^{\ln}$ and $VRP_t^{\sqrt{}}$ shown from left to right, respectively.

In order to further substantiate these results, we conduct four different tests proposed by Cavaliere and Taylor (2007) based on the sample RVP. The null hypothesis is hereby that the series at most displays conditional heterokedasticity versus the alternative of unconditional heteroskedasticity. Define the quantity $\widehat{W}(\tilde{u}) := \hat{\eta}(\tilde{u}) - \tilde{u}$ which will clearly be close to zero $\forall \tilde{u}$ in large samples under the null. To acquire a well defined limiting null distribution for this entity, Cavaliere and Taylor (2007) consider to calculate the following four test statistics and subsequently show under some mild regularity conditions:

$$\mathcal{H}_{KS} := \frac{T^{1/2}\hat{V}_T(1)}{\hat{\lambda}_v}\left(\sup_{\tilde{u}\in[0,1]} |\widehat{W}(\tilde{u})|\right) \xrightarrow{d} \sup_{\tilde{u}\in[0,1]} |\tilde{B}(\tilde{u})|\,, \tag{3.16}$$

$$\mathcal{H}_{CVM} := \frac{T\hat{V}_T(1)^2}{\hat{\lambda}_v^2}\left(\int_0^1 \widehat{W}(\tilde{u})^2\, \mathrm{d}\tilde{u}\right) \xrightarrow{d} \int_0^1 \tilde{B}(\tilde{u})^2\, \mathrm{d}\tilde{u}\,, \tag{3.17}$$

$$\mathcal{H}_{K} := \frac{T^{1/2}\hat{V}_T(1)}{\hat{\lambda}_v}\left(\sup_{\tilde{u}\in[0,1]} \widehat{W}(\tilde{u}) - \inf_{\tilde{u}\in[0,1]} \widehat{W}(\tilde{u})\right) \xrightarrow{d} \sup_{\tilde{u}\in[0,1]} \tilde{B}(\tilde{u}) - \inf_{\tilde{u}\in[0,1]} \tilde{B}(\tilde{u})\,, \tag{3.18}$$

$$\mathcal{H}_{AD} := \frac{T\hat{V}_T(1)^2}{\hat{\lambda}_v^2}\left(\int_0^1 \frac{\widehat{W}(\tilde{u})^2}{\tilde{u}(1-\tilde{u})}\, \mathrm{d}\tilde{u}\right) \xrightarrow{d} \int_0^1 \frac{\tilde{B}(\tilde{u})^2}{\tilde{u}(1-\tilde{u})}\, \mathrm{d}\tilde{u}\,, \tag{3.19}$$

where $\hat{\lambda}_v$ denotes a long-run variance estimator of the series $v_t := \varepsilon_t^2$ and $\tilde{B}(\cdot)$ a standard Brownian Bridge process.[11] The limiting null distributions of these right-sided tests are shown to follow the same as the Kolmogorov-Smirnov (3.16), Cramer-Von-Mises (3.17), Kuiper (3.18) and Anderson-Darling (3.19) tests so that critical values are retrieved from Shorack and Wellner (2009). We apply (3.16) - (3.19) to the residuals of all three ex-post Variance Risk Premium series and observe that all tests throughout reject at least at a $\alpha = 5\%$ nominal significance level (see Table 3.7). Therefore, apart from conditional heteroskedasticity, we additionally find strong evidence for unconditional heteroskedasticity in the data. Lastly, we apply several of the in Subsection 3.1.1 and 3.1.2 reviewed LM tests on long memory on the raw data and the residuals of the VRP_t, $VRP_t^{\ln}$ and $VRP_t^{\sqrt{}}$ series and report our results in Table 3.8. We consider the HCL test of

[11]For $\hat{\lambda}_v$ we implement a standard HAC estimator with Bartlett kernel (c.f Andrews (1991)).

	HCL	ALM	WB^{LM}	$s=8$: $\hat{S}^*_{q,QS}$	$s=8$: $\hat{S}^*_{q,B}$	$s=4$: $\hat{S}^*_{q,QS}$	$s=4$: $\hat{S}^*_{q,B}$
VRP_t	**6.15****	**4.057****	**0.04%****	**10.049****	**9.228****	**8.857****	**8.421****
	(2.326)	(2.326)		(2.326)			
$VRP_t^{\ln}$	**4.085****	**8.475****	10.4%	**12.789****	**12.814****	**12.849****	**12.903****
	(2.326)	(2.326)		(2.326)			
$VRP_t^{\surd}$	**6.158****	**7.518****	68.3%	**11.432****	**10.926****	**11.49****	**10.997****
	(2.326)	(2.326)		(2.326)			
$Res.\ VRP_t$	**1.589**	−1.374	45.1%	−0.292	−0.321	−1.764	−1.812
	(1.282)						
$Res.\ VRP_t^{\ln}$	**2.765****	−2.996	55.4%	−2.377	−2.376	−2.002	−1.999
	(2.326)						
$Res.\ VRP_t^{\surd}$	**2.397****	−1.859	48.3%	−0.60	−0.601	−2.143	−2.16
	(2.326)						

Table 3.8: Reported are the results for testing $d = 0$ for various LM tests for long memory (p-value for WB^{LM}) applied to the raw data and the residuals (Res.) of the VRP_t, $VRP_t^{\ln}$ and $VRP_t^{\surd}$ series, respectively. Bold-faced values indicate significance at the nominal 10% level; an additional * (**) indicates significance at the nominal 5% (1%) level. Respective critical values are given in parentheses.

Harris et al. (2008) (3.8), the ALM test of Demetrescu et al. (2008) (3.9), the WB^{LM} procedure of Cavaliere et al. (2015b) utilizing 9999 Bootstrap replications as well as incorporating the model order given in Table 3.7, and finally our $\hat{S}^*_q$ test (3.10). Setting aside the number of Bootstrap replications, the calibrations for all these tests are the same as in our simulation study in Section 3.2. Note that we only consider the HCL test with its nonparametric correction as a procedure which is non-robust to cond./uncond. heteroskedasticity.

As demonstrated in our simulation study this circumvents additional problems which would arise when considering parametric tests such as the ones of Tanaka (1999) or Breitung and Hassler (2002). When applied to the raw series, the majority of the tests clearly reject the null of short memory which is in line with the first impression given by the acf's in Figure 3.1 and the estimation results of the memory parameter found in Table 3.7. However, when applied to the respective residual series we find that the robust tests do not reject in any case, whereas the HCL test still clearly indicates long memory. In light of our previous findings from above as well as the theoretical results of Cavaliere et al. (2015b) (3.5) one has to interpret these rejections as spurious ones, most likely induced by heteroskedasticity. Furthermore, as previously pointed out, the tests of Cavaliere and Taylor (2007) (3.16) - (3.19) can not be considered to reliably distinguish if the identified unconditional heteroskedasticity is directly or indirectly induced. With respect to our findings in Tab. 3.4 only our test is expected to deliver most secure results in Tab. 3.8. In summary, these results clearly support the approach of Bollerslev et al. (2013) over the one of Bekaert and Hoerova (2014) and demonstrates the practical need for LM tests for long memory that are robust to heteroskedasticity of quite general form.

3.4 Conclusion

In this paper we propose a new nonparametric time-domain based LM test for long memory which is robust against fixed or switching short-run dynamics as well as conditional and unconditional heteroskedasticity of quite general form. Our test is based on the procedure of Harris et al. (2008), utilizing a nonparametric short memory correction, which is subsequently recast in a regression framework in the style of Breitung and Hassler (2002) and standardized using a HAC estimator of Andrews (1991). Contrary to other comparable procedures in the literature like the Wild bootstrap approach of Cavaliere et al. (2015b), our LM test is easy to implement and has a low computational burden by simultaneously avoiding model misspecification problems typically encountered in parametric tests.

We conduct an extensive Monte Carlo simulation study considering several DGP's that display a wide range of symmetric and asymmetric conditional heteroskedasticity, as well as directly and indirectly induced unconditional heteroskedasticity. Out of all considered tests it is revealed that our test is the only one controlling its size well throughout all scenarios and therefore under the heteroskedastic LS process class (3.1) considered here. This justifies our presumption of a standard normal limiting null distribution which has yet to be proven and remains the main focus of our current research. Further simulation results clearly point towards consistency of our test. Additionally, our test shows a good performance in a classic conditionally homoskedastic setup and its nonparametric correction is not too costly in terms of power. We further give some practical recommendations concerning kernel choice as well as bandwidth trimming selection.

Our empirical application concentrates on three different types of ex-post Variance Risk Premium series, represented by a fractional cointegration relationship between CBOE VIX and realized variance of the S&P500 index. We find that long memory remains in these series in addition to conditional and unconditional heteroskedasticity emphasizing the need for robust LM tests for long memory. We close by pointing out a path for future research. Due to its good finite sample performance in a univariate setting we consider to extend our test to a multivariate framework. To the best of our knowledge only the time-domain tests of Nielsen (2005), which primarily extends the test of Tanaka (1999), and Breitung and Hassler (2002) exist in this setup. Since our simulation study reveals that the latter two procedures are outperformed by our test already in a univariate framework this extension seems to be a very promising approach.

3.5 Appendix

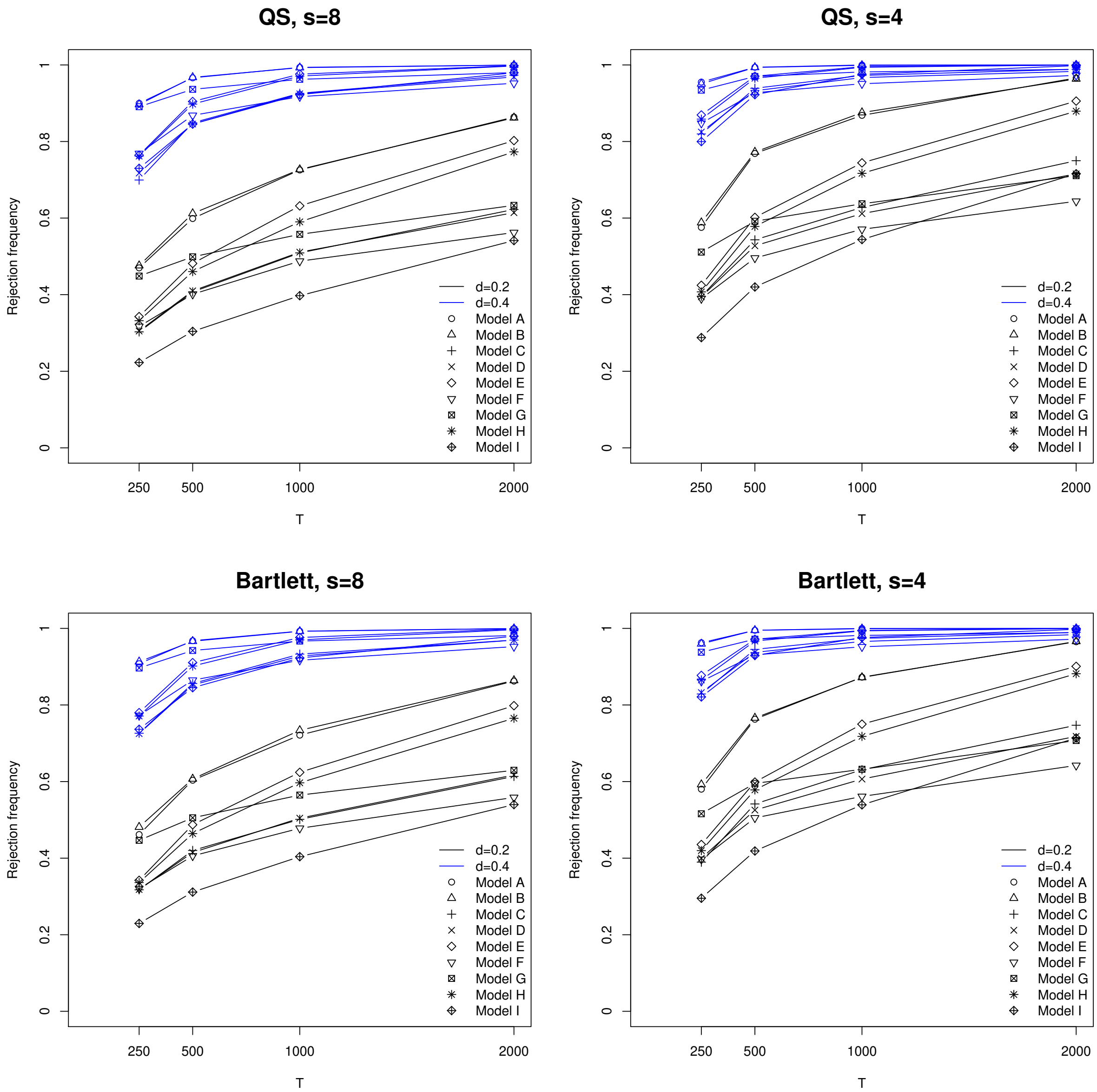


Figure 3.3: Reported are the rejection frequencies (Power) for testing $d = 0$ in (3.12) for the $\hat{S}_q^*$ test employing the Quadratic Spectral (QS) and Bartlett kernel with bandwidth trimming $s \in \{4, 8\}$ under conditional heterokedasticity (Models A-I). The DGP is generated for various sample sizes T, $d \in \{0.2, 0.4\}$ and no short-run dynamics.

$\phi(\tilde{\tau})/\theta(\tilde{\tau})$	T	$d=0.2$								$d=0.4$							
		$s=8$				$s=4$				$s=8$				$s=4$			
		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$		$\hat{S}^*_{q,QS}$		$\hat{S}^*_{q,B}$	
		ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma	ar	ma
$-0.5+1\mathbb{I}(t\geq\lfloor 0.25T\rfloor)$	250	26.8	27.79	28.63	28.4	38.86	38.1	42.14	38.55	64.56	69.14	67.9	71.85	80.86	82.16	85.04	85.25
	500	40.79	41.88	40.78	41.94	63.75	58.69	66.23	60.1	84.39	87.39	86.51	88.76	96.16	97.04	97.48	97.6
	1000	56	55.44	56.61	55.11	77.91	75.03	78.12	75.03	95.58	96.63	96.44	96.74	99.6	99.61	99.74	99.79
	2000	73.91	74.03	73.91	73.92	93.18	91.81	93.75	91.77	99.65	99.72	99.63	99.7	100	100	100	100
$-0.5+1\mathbb{I}(t\geq\lfloor 0.5T\rfloor)$	250	22.7	23.39	25.44	23.58	33.01	30.93	35.13	31.71	57.91	64.45	64.3	68.49	72.58	76.39	78.54	80.35
	500	34.55	35.23	35.6	35.41	55.74	50.94	56.58	50.45	79.32	84.49	83.05	85.28	92.26	93.98	94.71	95.65
	1000	47.51	47.51	48.39	46.78	68.3	66.02	69.36	64.95	92.79	94.54	93.94	95.52	98.42	99.09	98.96	99.28
	2000	66.49	65.45	66.7	65.24	88.47	84.89	88.21	85.04	99.04	99.43	99.22	99.3	99.91	99.99	99.95	99.98
$-0.5+1\mathbb{I}(t\geq\lfloor 0.75T\rfloor)$	250	22.6	18.58	23.82	20.09	30.72	26.32	32.27	26.88	51.06	56.02	56.23	60.47	61.94	66.27	67.75	71.45
	500	31.5	29.92	33.34	29.23	48.96	41.93	50.17	41.12	70.97	78.63	74.85	80.99	81.67	88.09	87.3	90.96
	1000	42.46	39.95	42.44	39.78	61.21	56.25	60.4	55.07	88.09	92.34	89.79	92.96	94.13	97.69	96.27	98.12
	2000	59.07	55.78	58.27	55.23	79.96	75.26	80.06	75.43	97.83	98.72	98.19	98.74	99.25	99.88	99.58	99.89
$\phi(\tilde{\tau})=-0.5+\tilde{\tau}$	250	30		31.21		41.5		42.45		69.9		72.27		81.74		83.93	
	500	44.44		44.56		63.28		62.89		86.78		87.56		95.15		96.44	
	1000	56.2		56.24		76.17		76		95.73		96.15		99.07		99.37	
	2000	74.88		74.27		92.01		91.8		99.72		99.68		99.96		99.91	
$\phi(\tilde{\tau})=\frac{0.5}{1+\exp(-10+20\tilde{\tau})}$	250	30.55		32.5		41.65		42.25		70.08		73.03		82.66		85.89	
	500	42.64		44.32		63.24		64.86		86.12		87.72		96.25		97.3	
	1000	56.64		56.96		77.47		78.61		95.45		96.26		99.35		99.62	
	2000	74.25		74.89		93.38		93.26		99.68		99.76		99.97		99.99	
$\phi(\tilde{\tau})=0.5\sin(\pi\tilde{\tau})$	250	31.53		33.12		43.48		45.28		72.12		74.31		84.22		88.09	
	500	45.74		45.46		66.69		68.32		87.07		89.12		97.8		98.19	
	1000	59.49		59.93		80.35		80.2		96.92		97.1		99.7		99.8	
	2000	78.39		77.75		94.56		94.87		99.77		99.79		99.99		100	

Table 3.9: Reported are the rejection frequencies (Power) for testing $d=0$ in (3.12) for the $\hat{S}^*_q$ test under switching short-run dynamics (DC (3.13) - (3.14) and RC Models (3.15)). The DGP is generated with conditionally homoskedastic Gaussian innovations, $d\in\{0.2,0.4\}$, various sample sizes T and $\phi(\tilde{\tau})$ and $\theta(\tilde{\tau})$ present solely in the autoregressive (ar) or the moving average (ma) polynomial, respectively.

Chapter 4

A Criticism on Spurious Long Memory

A Criticism on Spurious Long Memory

Co-authored with Kai Wenger.

4.1 Introduction

Ever since the emergence of autoregressive fractionally integrated moving average (ARFIMA) processes in the seminal articles of Granger and Joyeux (1980) and Hosking (1981), long memory time series have found numerous applications in economics and finance. Among these range realized volatility series, exchange rates, forward premia, aggregate output and inflation rates (e.g. Andersen et al. (2003), Cheung (1993), Baillie and Bollerslev (1994), Diebold and Rudebusch (1989) as well as Hassler and Wolters (1995) and Baillie et al. (1996), respectively). A survey of the literature is provided by Baillie (1996) and a comprehensive overview of the existing methodology by Beran et al. (2013) and Giraitis et al. (2012).
However, a second branch in the literature quickly developed, questioning the existence of true long memory in the data. The main point of criticism is that a short memory series contaminated by rare level shifts displays features commonly attributed to true long memory like hyperbolically decaying sample autocorrelations and a steep slope of the periodogram for Fourier frequencies near the origin, easily misleading a practitioner in specifying an appropriate model (e.g. Lobato and Savin (1998), Diebold and Inoue (2001), Granger and Hyung (2004) and Perron and Qu (2010)). A distinction of these two classes is however of utmost importance in view of the different laws of asymptotics that apply, subsequent statistical inference, model estimation and ultimately forecasting performance (e.g Perron and Qu (2010) and Varneskov and Perron (2017)). Therefore, estimators of the memory parameter under low-frequency contaminations (e.g. McCloskey and Perron (2013) and Hou and Perron (2014)) and tests on spurious long memory were developed by i.a. Shimotsu (2006), Ohanissian et al. (2008) and Qu (2011), where in particular the latter test benefits from weak assumptions and superior finite sample performance (c.f. Leccadito et al. (2015)). The main idea of the Qu (2011) test is to distinguish these two model classes based on the rate at which the periodogram tapers of at Fourier frequencies near the origin which is much steeper in case of spurious long memory compared to the true long memory case.
In this paper, we argue that a sole distinction based on a spectral rate condition is not sufficient to distinguish spurious from true negative, short or long memory. We loosen the usual assumption of requiring a second-order stationary process and only impose that the second moment exists. A periodogram can therefore be computed in any case and will converge to a pseudo spectral density in the limit. We then study two recently proposed models from the literature, on one hand a subset of locally stationary processes (see Dahlhaus (2000), Palma and Olea (2010)) proposed by Demetrescu and Sibbertsen (2016) and Cavaliere et al. (2015b) and on the other hand a class of random coefficient models, proposed by Giraitis et al. (2014). We take

existing results concerning the former and establish concerning the latter model class that these processes obey the same spectral rate condition typically associated with a true negative, short or long memory process and study the impact on the size of the Qu (2011) test in a large scale Monte Carlo simulation study. We find that a non-negligible bias occurs and pertains even in largest sample sizes being especially pronounced under recommended user-specific bandwidth choices. This shows that not only processes displaying low-frequency contaminations can easily be confused with true long memory, but that under some weaker assumptions, true negative, short or long memory processes can be wrongly attributed to the spurious long memory class. We then extend the prewhitening procedure of Qu (2011) to alleviate the bias induced by the class of locally stationary processes considered here and obtain good size results even in small samples. Additionally our procedure is not too costly in terms of power.
As empirical applications, we on one hand consider log spot prices, corresponding one-period futures contract prices, as well as spreads of gold, silver, platinum and crude oil commodities and on the other hand monthly (seasonally adjusted) inflation rates of G7 countries. We find that in particular the former series are subject to a spurious rejection of the null hypothesis, especially when using recommended user-specific bandwidth choices.
The rest of this paper is structured as follows: Section 4.2 gives a brief survey of the literature on true versus spurious long memory and provides an extended definition of true negative, short and long memory. Locally stationary processes and a class of random coefficient models are discussed in Subsection 4.2.1 and 4.2.2, respectively. Section 4.3 gives the results of our Monte Carlo simulation study and Section 4.4 the empirical analysis of commodities (Subsection 4.4.1) and inflation rates (Subsection 4.4.2), whereas Section 4.5 concludes. A proof is given in the appendix in addition to further simulation results and a practitioners guide on the algorithm of Bai and Perron (1998, 2003a) which is embedded in our extended prewhitening procedure.

4.2 True versus Spurious Long Memory

Following Beran et al. (2013), let the second-order stationary process $\{X_t\}$ with $t = 1,\dots,T$ satisfy the following representation:

$$f(\lambda) \simeq G\lambda^{-2d} \quad \text{as} \quad \lambda \to 0_+ \,, \tag{4.1}$$

where $f_X(\lambda) = \frac{1}{2\pi}(\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k \cos(\lambda k))$ denotes the spectral density of the process at frequency $\lambda \in [-\pi,\pi]$, γ_k the respective autocovariances at lag $k \in \mathbb{Z}$, d the memory parameter with $d \in (-0.5, 0.5)$ and $G \in (0,\infty)$. Under mild additional regularity conditions, one can also consider the following equivalent definition:

$$\begin{aligned} \gamma(k) &\sim L_\gamma(k)|k|^{2d-1} \quad \text{for} \quad k \to \infty \,, \\ L_\gamma(k) &= 2L_f(k^{-1})\Gamma(1-2d)\sin(\pi d) \,, \end{aligned} \tag{4.2}$$

where $d \in (-0.5, 0) \bigcup (0, 0.5)$, $\gamma(k)$ the autocovariance function of $\{X_t\}$ evaluated at lag k, $L_f(\cdot)$ a slowly varying function in Zygmund's sense (most commonly the constant G of (4.1)) and $\Gamma(\cdot)$ the gamma function. Concerning the long-run variance, Beran et al. (2013) conclude:

$$\sum_{k=-\infty}^{\infty} \gamma(k) = \begin{cases} 2\pi f(0) = 0 \quad \text{for} \quad d \in (-0.5, 0)\,, \\ 2\pi f(0) = 0 < 2\pi G < \infty \quad \text{for} \quad d = 0\,, \\ 2\pi \lim\limits_{\lambda \to 0} f(\lambda) = \infty \quad \text{for} \quad d \in (0, 0.5)\,. \end{cases} \tag{4.3}$$

Therefore, the long memory property of a process can be fully characterized by a spectrum being of order $O(\lambda^{-2d})$ for $|\lambda| \to 0$, as well as unboundedness of the spectrum at the origin which in turn implies a non-summable autocorrelation function (acf), since the autocorrelations decay hyperbolically. Complementary, in case of short memory ($d = 0$) the spectrum of the process is $O(1)\ \forall\, \lambda$ and the autocorrelations decay at most at an exponential rate, resulting in a summable acf (c.f. (4.1) and (4.3)).
A prominent example of a process satisfying these conditions and often used in practice to model long memory is the $ARFIMA(p,d,q)$ process of Granger and Joyeux (1980) and Hosking (1981), given by:

$$\phi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \quad \text{for} \quad t = 1, \ldots, T\,, \tag{4.4}$$

where L is the lag operator, $\phi(L) = (1 - \phi_1 L - \ldots - \phi_p L^p)$, $\theta(L) = (1 + \theta_1 L + \ldots + \theta_q L^q)$, $\varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2)$, $(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)L^j}{\Gamma(-d)\Gamma(j+1)}$ and in which case $G = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(1)|^2}{|\phi(1)|^2}$ in (4.1). Now, let the periodogram of $\{X_t\}$ be given by:

$$I_X(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} X_t \exp(i\lambda_j t) \right|^2\,, \tag{4.5}$$

evaluated at the Fourier frequencies $\lambda_j = 2\pi j/T$ with $j = 1, \ldots, [T/2]$. In case the process satisfies (4.1)-(4.3), it directly follows from Beran et al. (2013) and McCloskey and Perron (2013) that $I_X(\lambda_j) = O_p(\lambda_j^{-2d})$ for $\lambda_j = o(1)$ holds and moreover $E(I_X(\lambda_j)/f(\lambda_j)) \to 1$ for all j with $j \to \infty$, $T \to \infty$, and $j/T \to 0$, following Robinson (1995b). In the remainder of this paper, we shall concentrate on the above rate conditions in the frequency domain and refer to true negative, short or long memory processes, as established in the following definition:

Definition 4.1 (True Negative, Short and Long Memory)**.** *A process X_t has true negative, short or long memory, if its (pseudo) spectral density $f_X(\lambda)$ and periodogram $I_X(\lambda_j)$, obey the rate conditions $O(\lambda^{-2d})$ for $|\lambda| \to 0$ and $O_p(\lambda_j^{-2d})$ for $\lambda_j = o(1)$, respectively.*

Contrary to the above discussion, we no longer require the process in Definition 4.1 to be second-order stationary or even linear. We much more focus on a finite second moment and fully rely on the rate condition which is nevertheless satisfied by the periodogram. In any case, this spectral estimator can be computed and will converge at least to a pseudo spectral density in the limit if non-stationarity of the kind studied in this paper is given (e.g. Hou and Perron (2014)). However, recent literature brought forth several types of processes which cause standard statistical inference tools like sample acf and periodogram to spuriously display features commonly attributed to true long memory processes (e.g. Diebold and Inoue (2001), Granger and Hyung (2004), Perron and Qu (2010), Qu (2011) etc.). In particular, the acf of these kind of processes display seemingly hyperbolically decaying autocorrelations and their periodogram is unbounded for Fourier frequencies near the origin. Among these so called spurious long memory processes the following random level shift model is the most well studied (c.f. Perron and Qu (2010)):

$$r_t = z_t + \mu_t \quad \text{with} \quad \mu_t = \mu_{t-1} + \delta_t \eta_t \qquad t = 1, \ldots, T\,. \tag{4.6}$$

Here, the second-order stationary, mean-zero process z_t follows (4.4) with $d = 0$, $\eta_t \overset{iid}{\sim} (0, \sigma_\eta^2)$ and $\delta_t \overset{iid}{\sim} B(1, p_T)$ is a Bernoulli random variable with $p_T = p/T$ and $0 < p < \infty$.[1] This setting lets the expected number of shifts that occur remain bounded which asymptotically results in $r_t \sim I(0)$. By assuming δ_t, η_t and z_t to be mutually independent, Qu (2011) establishes, building on results of Perron and Qu (2010):

$$\begin{aligned} I_r(\lambda_j) &= \frac{1}{2\pi T}\left|\sum_{t=1}^{T} z_t \exp(i\lambda_j t)\right|^2 + \frac{1}{2\pi T}\left|\sum_{t=1}^{T} \mu_t \exp(i\lambda_j t)\right|^2 + \frac{2}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T} z_t \mu_s \cos(\lambda_j (t-s)) \\ &= O_p(1) + O_p(T^{-1}\lambda_j^{-2}) + O_p(T^{-1/2}\lambda_j^{-1})\,, \end{aligned} \tag{4.7}$$

for $\lambda_j = o(1)$. The crucial point in (4.7) is that one can deduce for Fourier frequencies of order $j = O(T^{1/2})$, that the periodogram of (4.6) has a steeper slope, more associated with a unit root process (i.e. $d = 1$ in (4.4)) than with a true long memory process where the slope is $-2d > -1$ for $j = o(T)$. Therefore, although both a true and a spurious long memory process display poles in their periodograms for Fourier frequencies approaching the zero frequency as the sample size increases, the rate at which they taper off clearly differs. Complementary, for large T and λ_j local to zero, McCloskey and Perron (2013) show that for a random level shift process given in (4.6) the following approximation holds:

$$E[I_r(\lambda_j)] \approx \lambda_j^{-2d} f_z(\lambda_j) + T^{-1}\lambda_j^{-2} g(\lambda_j)\,, \tag{4.8}$$

where $f_z(\lambda_j)$ is the spectral density of z_t in (4.6) and $g(\cdot)$ a non-negative even function bounded at zero. Note that (4.6) is not the only model capable of displaying spurious long memory. Qu (2011) establishes the same theoretical results for a short memory process containing a smoothly varying trend and finds similar behavior i.a. in Markov switching models with iid or generalized autoregressive conditional heteroskedastic (GARCH) regimes (see also Baek et al. (2014) and

[1] Note that McCloskey and Perron (2013) extend the results of Perron and Qu (2010) by allowing z_t to have long memory.

Yu (2009)) via a simulation study. Moreover, Hou and Perron (2014) summarize these processes under their property of inducing low-frequency contaminations and propose a modified local Whittle estimator which is consistent in this setup. Complementary to above, we directly follow Qu (2011) and McCloskey and Perron (2013) and define a spurious long memory process solely based on the rate condition of its periodogram as follows:

Definition 4.2 (Spurious Long Memory)**.** *A time series r_t satisfying (4.6)-(4.8) has spurious long memory if its periodogram $I_r(\lambda_j)$ obeys the rate condition $O_p(T^{-1}\lambda_j^{-2})$ for $|\lambda_j| \to 0$.*

Distinguishing between true and spurious long memory goes far beyond a mere technical exercise. Most recently, Varneskov and Perron (2017) use a modified version of (4.6) with $z_t \sim I(d)$ and $d \in [0, 0.5)$ to forecast eight different volatility series and find that their model outperforms, among others, (4.4) in a large scale out-of-sample forecast evaluation exercise. Therefore, correctly discriminating between these two model classes is of considerable interest in applied time series analysis.

A well known procedure to test for spurious long memory which is completely based on the above discussed differing spectral rate conditions is the Lagrange Multiplier test of Qu (2011). The null hypothesis incorporates all second-order stationary true negative, short or long memory processes as discussed in (4.1)-(4.3), with (4.4) being a special case. Under the alternative however, the process has short memory and is contaminated by some level shift, smooth trend, or other low-frequency contamination inducing spurious long memory (see Definition 4.2). The test statistic is based on the local Whittle likelihood function (see Künsch (1986) and Robinson (1995a)) and is given by:

$$W = \sup_{r \in [\epsilon, 1]} \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \left| \sum_{j=1}^{[mr]} \nu_j \left(\frac{I_X(\lambda_j)}{G(\hat{d})\lambda_j^{-2\hat{d}}} - 1 \right) \right|, \tag{4.9}$$

with $\nu_j = \log \lambda_j - (1/m) \sum_{j=1}^{m} \log \lambda_j$, $\hat{d}$ the estimate of the memory parameter obtained via the local Whittle estimator of Robinson (1995a), and ϵ a small trimming parameter. Qu (2011) shows that (4.9) has a well defined limiting null distribution and reports critical values for $\epsilon \in \{0.02, 0.05\}$, where the latter value is recommended for sample sizes $T < 500$. By assuming $m/T^{1/2} \to \infty$ for $T \to \infty$, a recommendation based on extensive simulation study, of using $m = \lfloor T^b \rfloor$ frequency ordinates with bandwidth $b = 0.7$ is given when applying the test. The test is proven to be consistent if the data generating process (DGP) follows a spurious long memory process of the kind discussed in Definition 4.2. Furthermore, it has several desireable properties like not requiring Gaussianity, allowing for conditional heteroskedasticity in the error terms of the DGP, not requiring a precise specification of the low-frequency contamination present in the DGP due to its Score nature and displaying high finite sample power results compared to other similar existing testing procedures like Shimotsu (2006) and Ohanissian et al. (2008) (c.f. Leccadito et al. (2015)). Thus in what follows, we shall concentrate on this test and present in the following two Subsections two classes of processes which can be subsumed under Definition 4.1. These will serve as the basis of our criticism, since any rejection observed in this case is in fact spurious, demonstrating that a distinction between true and spurious long memory solely based on the above discussed spectral rate conditions is not sufficient.

4.2.1 Locally Stationary Processes

Building on results of Dahlhaus (2000), Palma and Olea (2010) define a locally stationary (LS) process $\{X_{t,T}\}$ with $t = 1,\dots,T$, capable of displaying true long memory, by:

$$X_{t,T} = \sigma\left(\frac{t}{T}\right)\sum_{j=0}^{\infty}\psi_j\left(\frac{t}{T}\right)\varepsilon_{t-j}\,, \tag{4.10}$$

with $\varepsilon_t \overset{iid}{\sim} N(0,1)$ and $\{\psi_j\}$ coefficients that obey $\sum_{j=0}^{\infty}\psi_j(\tau)^2 < \infty$ for all $\tau := \lfloor t/T \rfloor \in [0,1]$. It is straightforward to see that (4.10) extends the well known Wold representation of a linear process, by allowing the parameter vector to vary smoothly over time. A particular example of a process satisfying (4.10) is given by Palma and Olea (2010) in the form of the locally stationary autoregressive fractionally integrated moving average (LS-ARFIMA) process $\{X_{t,T}\}$ with $t = 1,\dots,T$:

$$\Phi(\tau,L)X_{t,T} = \sigma(\tau)\Theta(\tau,L)(1-L)^{-d(\tau)}\varepsilon_t\,, \tag{4.11}$$

where $\Phi(\tau,L) = (1-\phi_1(\tau)L-\ldots-\phi_p(\tau)L^p)$ and $\Theta(\tau,L) = (1+\theta_1(\tau)L+\ldots+\theta_q(\tau)L^q)$ are the respective autoregressive and moving average polynomials, $d(\tau)$ a memory parameter, $\sigma(\tau)$ a noise scale factor, all of which are smoothly varying and $\varepsilon_t \overset{iid}{\sim} N(0,1)$. Note that (4.11) naturally extends (4.4) where all parameters are independent of τ. To estimate a LS-ARFIMA model, Palma and Olea (2010) derive a Whittle likelihood technique and rigorously establish consistency, normality and efficiency (see also Beran (2009)).

At this point, we would like to emphasize that model (4.11) is non-stationary in a global sense, since already a single switch in one of the parameters almost surely induces unconditional heteroskedasticity and/or smooth or sudden changes in the off-diagonal elements of the autocovariance matrix of the process. However, since $\sum_{j=0}^{\infty}\psi_j(\tau)^2 < \infty$ for all τ in (4.10) holds, the second moment is clearly bounded and the process itself is locally second-order stationary for any regime between the breakpoints. Importantly, Palma and Olea (2010) establish the following semiparametric representation by letting the time-varying spectral density of the process satisfy:

$$f_\varrho(\tau,\lambda) \simeq C_f(\varrho,\tau)|\lambda|^{-2d_\varrho(\tau)} \quad \text{as} \quad |\lambda| \to 0\,. \tag{4.12}$$

Here, ϱ is a vector containing all parameters of (4.11), $C_f(\varrho,\tau) > 0$, $0 < \inf_{\varrho,\tau} d_\varrho(\tau)$, $\sup_{\varrho,\tau} d_\varrho(\tau) < 1/2$ and $d_\varrho(\tau)$ having a bounded first derivative with respect to τ. This establishes that (4.12) can be viewed as a natural extension of (4.1), which in turn shows that the order of the pseudo or time-varying spectral density of a LS-ARFIMA process is clearly $O(\lambda^{-2d_\varrho(\tau)})$ as $|\lambda| \to 0$ for all

τ, establishing that it falls under Definition 4.1 of a true short or long memory process with a globally bounded second moment.

Since all parameters of a LS-ARFIMA model are dependent on τ, this opens a near endless flexibility to model persistence, variance scaling and non-stationarity. Therefore, in what follows we concentrate on a subset of (4.11) given in form of a model studied in Demetrescu and Sibbertsen (2016). Let $\{x_t\}$ with $t = 1, \ldots, T$ be defined by:

$$x_t = \sigma(\tau) v_t \,. \tag{4.13}$$

In this case $\{v_t\}$ follows an $ARFIMA(p,d,q)$ process given in (4.4) with $d \in (-0.5, 0.5)$ and the scaling factor $\sigma(\tau)$ is capable of inducing breaks in the unconditional variance leading to unconditional heteroskedasticity, also termed non-stationary volatility. It can easily be seen that $\{x_t\}$ restricts (4.11) in the sense that only the variance scaling factor $\sigma(\tau)$ is allowed to vary smoothly, whereas all remaining parameters are independent of τ. We sharpen the further analysis in this way, since breaks in the unconditional variance have received increasing interest in the recent literature concerning applications on financial and macroeconomic data where the so-called "Great Moderation", present i.a. in inflation rates, during the begin of the 1990's serves as the most prominent example. Notable contributions include, among others, McConnell and Perez-Quiros (2000), Sensier and Van Dijk (2004), Cavaliere et al. (2015b) and see also Harris and Kew (2017) for an extensive overview. More precisely, we directly adopt from Demetrescu and Sibbertsen (2016):

Assumption 4.1 (Unconditional Heteroskedasticity and Long Memory)**.** *Let $\sigma(\tau)$, with non-negative $\sigma(\cdot)$ and $\tau \in (0,1)$, satisfy a uniform Lipschitz condition at all but a finite number of discontinuity points. The process $\{x_t\}$ is generated as in (4.13), where $\{v_t\}$ is Gaussian and follows (4.4) with $d \in (-0.5, 0.5)$ and obeying (4.1) - (4.3), respectively.*

This implies that $\sigma(\tau)$ is bounded resulting in a finite number of breaks in the unconditional variance of the process at deterministic relative sample points τ. Complementary, Cavaliere et al. (2015b) study a heteroskedastic $ARFIMA(p,d,q)$ model by replacing the error terms in (4.4) with $\varepsilon_t = \sigma(\tau)\tilde{z}_t$ and $\tilde{z}_t \overset{iid}{\sim} (0, \sigma^2_{\tilde{z}})$ for $t = 1, \ldots, T$ and thus consider scaling the noise term rather than the entire process as done in (4.13). We would like to refer to Demetrescu and Sibbertsen (2014) who establish under weak regularity conditions and by rewriting both model types into their respective Wold representation form that both ways to model unconditional heteroskedasticity can be treated asymptotically equivalent. Therefore, $\{x_t\}$ is nested in the heteroskedastic $ARFIMA(p,d,q)$ model proposed by Cavaliere et al. (2015b) in the sense of Assumption 4.1.[2] Lastly, it directly follows from Demetrescu and Sibbertsen (2014) that the

[2]More concretely, $\{x_t\}$ fufills Assumption $\mathcal{R}$ and $\mathcal{V}$ *a*) on p. 559 of Cavaliere et al. (2015b), but not Assumption $\mathcal{V}$ *b*) dealing with conditional heteroskedasticity in the error terms which is not subject of this paper.

time-varying autocovariance function and the periodogram of $\{x_t\}$ obey:

$$\gamma_k^{t,T}(x_t) = \gamma_k \sigma_t \sigma_{t-k} \quad \text{as} \quad T \to \infty\,, \tag{4.14}$$

$$I_x(\lambda_j) = O_p(\lambda_j^{-2d}) \quad \text{as} \quad \lambda_j \to 0_+\,. \tag{4.15}$$

Thus, we conclude this subsection by emphasizing that under Assumption 4.1, $\{x_t\}$ is globally non-stationary, but will nevertheless have a finite second moment as can easily be seen in (4.14). Its periodogram in (4.15) fulfills the same rate condition as a true negative, short or long memory process, enabling us to subsume it under Definition 4.1.

4.2.2 Random Coefficient Models

Up to this point, changes in the parameter vector were modelled deterministically, ultimately resulting in structural breaks that remain unforecastable. Contrary to that, the random coefficient (RC) models literature relies on short-run dynamics driven by persistent stochastic processes, capable of inducing breaks in the off-diagonal elements of the autocovariance matrix in addition to indirectly causing changes in the unconditional variance of the observed process. In general such a modeling approach is justified in the literature by an evolution of beliefs of policy makers, changing cultural norms and overall time-varying policy in an economy. A notable contribution is Cogley and Sargent (2005), who employ a VAR model with parameters following a bounded random walk process to analyze i.a. inflation rates during the time of the so-called "Great Moderation". Similar work has been brought forth by Benati and Surico (2008, 2009) and Cogley et al. (2010), among others. Giraitis et al. (2014) provide an extensive survey of the literature in addition to proposing a new class of RC models on which we shall focus on in the following. Let $\{y_t\}$ for $t = 1,\ldots,T$, be given by:

$$\begin{aligned} y_t &= \alpha_t + \rho_{t-1} y_{t-1} + u_t\,, \\ \rho_t &= \rho \frac{a_t}{\max\limits_{0\le k\le t}|a_k|}, \quad \text{for} \quad t \ge 0\,, \end{aligned} \tag{4.16}$$

where $\rho_t \in [-\rho, \rho] \subset (-1,1)$ is a drifting random coefficient, α_t a random intercept, the process $\{u_t\}$ a stationary ergodic martingale difference sequence and $\{a_t\}$ a stochastic process driving the random drift. Concerning this model, we impose, directly following Giraitis et al. (2014):

Assumption 4.2 (RC Model). *The process $\{y_t\}$ for $t = 1,\ldots,T$ is generated according to (4.16) and the parameters ρ_t, α_t, initialization y_0 and processes $\{u_t\}$ and $\{a_t\}$ satisfy, respectively:*

1.) ρ_t and a_t for $t = 0,\ldots,T$ are measurable with respect to some filtration $\mathfrak{F}_t$, $E(a_0^4) < \infty$, $E(y_0^4) < \infty$ and $E(u_1^4) < \infty$. The process $v_t := \{a_t - E(a_t)\} - \{a_{t-1} - E(a_{t-1})\}$ for $t = 1,\ldots,T$ is stationary with zero mean and finite variance.

2.) There exists $\gamma \in (0,1)$ such that $T^{-\gamma}(a_{[\tau T]} - E(a_{[\tau T]})) \overset{D[0,1]}{\Rightarrow} W_\tau$ for $0 \le \tau \le 1$, where $\overset{D[0,1]}{\Rightarrow}$ denotes weak convergence in Skorokhod space and W_τ a zero mean random process with finite *variance like a standard or fractional Brownian motion. Moreover,* $|E(a_t) - E(a_{t+k})| \leq Ck^{\gamma}$ *for* $1 \leq k < t$*, holds for some generic constant* C*.*

3.) α_t *for* $t = 1, \ldots, T$ *is* $\mathfrak{F}_t$ *measurable,* $\max_j E(\alpha_j^4) < \infty$ *and for some* $\tilde{\beta} \in (0,1]$*,* $E(\alpha_t - \alpha_{t+k})^2 \leq C(k/t)^{2\tilde{\beta}}$ *for* $1 \leq k < t/2$ *holds.*

More concrete, Assumption 4.2 ensures that $\{y_t\}$ and the parameter processes $\{\rho_t\}$ and $\{\alpha_t\}$ of the AR(1) type RC model (4.16), are bounded and display persistence which validate the use of the nonparametric kernel estimation methods of Giraitis et al. (2014) discussed below. One can easily observe that $\{\rho_t\}$ is driven by a restricted random walk process, confining it into the stationary domain where the differences $v_t := a_t - a_{t-1}$ for $t = 0, \ldots, T$ can be modelled for example by (4.4) with $\gamma = 1/2$ ($\gamma = 1/2 + d$) for $d = 0$ ($|d| < 1/2$). Moreover, concerning the random intercept α_t, Giraitis et al. (2014) show that the following relations hold under Assumption 4.2 with $\tilde{\beta} = \gamma$ for $T \to \infty$:

$$\begin{aligned} \mu_t &= \frac{\alpha_t}{(1-\rho_t)} + o_p(1)\,, \\ \alpha_t &= \mu_t - \rho_{t-1}\mu_{t-1}\,. \end{aligned} \tag{4.17}$$

The persistent random process $\{\mu_t\}$ is the so-called "attractor", which can be interpreted in a classic AR(1) model setup as the mean of $\{y_t\}$, but is time-varying here. We would like to refer to Giraitis et al. (2014) concerning more technical details on (4.16), but emphasize that $E(y_t) \to 0$ for $\alpha_t \equiv 0$ holds. Moreover, in any case $\max_j E(y_j^2) < \infty$ for $j = 0, \ldots, T$ is ensured, since $\max_j E(\mu_j^2) < \infty$ is assumed. Using previous terminology, the process $\{y_t\}$ is therefore globally non-stationary, with a switch in the elements of the autocovariance matrix occuring almost surely at each lag, by simultaneously displaying a global finite second moment.[3]

It is well known from the literature that statistical inference conducted on RC models in general requires casting them in state space form where parameter estimation is then carried out by using variants of the Kalman filter. This brings along several disadvantages like a huge computational effort and an often not sufficiently developed asymptotic theory. Contrary to that, Giraitis et al. (2014) propose convenient to compute, nonparametric kernel-based estimators for their RC models and establish consistency and studentized asymptotic normality under very mild regularity conditions. To estimate μ_t, ρ_t and α_t for each point in time in (4.16)-(4.17), they consider:

$$\bar{y}_t = \frac{\sum_{j=1}^T b_{tj} y_j}{\sum_{j=1}^T b_{tj}}\,, \qquad \hat{\rho}_{T,t} := \frac{\sum_{k=1}^T b_{tk}(y_k - \bar{y}_t)(y_{k-1} - \bar{y}_t)}{\sum_{k=1}^T b_{tk}(y_{k-1} - \bar{y}_t)^2}\,, \qquad \hat{\alpha}_{T,t} = \bar{y}_t - \hat{\rho}_{T,t}\bar{y}_t\,, \tag{4.18}$$

where $b_{tk} := K\left(\frac{t-k}{H}\right)$ with $K(\mathrm{x}) \geq 0$ and $\mathrm{x} \in \mathbb{R}$, is a continuous bounded kernel obeying $\int_{-\infty}^{\infty} K(\mathrm{x})\mathrm{dx} = 1$.

[3]Note that Giraitis et al. (2014) refer to LS processes of Subsection 4.2.1 as deterministic coefficient models.

The following kernel functions are most common in the literature:

$$K(\mathrm{x}) := \begin{cases} (1/2)\mathbb{I}(|\mathrm{x}| \leq 1) \quad \text{flat kernel}\,, \\ (3/4)(1-\mathrm{x}^2)\mathbb{I}(|\mathrm{x}| \leq 1) \quad \text{Epanechnikov kernel}\,, \\ (1/\sqrt{2\pi})\exp(-\mathrm{x}^2/2) \quad \text{Gaussian kernel}\,. \end{cases} \tag{4.19}$$

Theoretically and supported by extensive simulation evidence, Giraitis et al. (2014) recommend using the Gaussian kernel from (4.19) in combination with a MSE-optimal bandwidth choice of $H = \sqrt{T}$ in practice. For what follows, define:

$$B_{1t} := \sum_{k=1}^{T} b_{tk}\,, \qquad B_{2t}^2 := \sum_{k=1}^{T} b_{tk}^2\,, \qquad \hat{\sigma}^2_{\hat{Y}\hat{u},t} := \sum_{k=1}^{T} b_{tk}^2 \hat{y}_{k-1}^2 \hat{u}_k^2\,,$$
$$\hat{\sigma}^2_{\hat{Y},t} := \sum_{k=1}^{T} b_{tk}\hat{y}_{k-1}^2\,, \qquad \widehat{B}_{3t}^2 := \sum_{j=1}^{T} b_{Tj}^2 \left(1 - \bar{y}_t \frac{B_{1t}}{\hat{\sigma}^2_{\hat{Y},t}}(y_{j-1} - \bar{y}_t)\right)^2 \hat{u}_j^2\,,$$

where $\hat{y}_k = y_k - \bar{y}_t$, and $\hat{u}_k = y_k - \hat{\rho}_t y_{k-1} - \hat{\alpha}_{T,t}$ for $k = 1,\ldots,T$ are the residuals from (4.16). To complete statistical inference, Giraitis et al. (2014) establish for $T \to \infty$:

$$\begin{aligned} \left|\frac{1-\hat{\rho}_{T,t}}{1+\hat{\rho}_{T,t}}\right|^{1/2} \frac{B_{1t}^{3/2}}{B_{2t}\hat{\sigma}_{\hat{Y},t}}(\bar{y}_t - \mu_t) &\xrightarrow{d} N(0,1)\,, \\ \frac{\hat{\sigma}^2_{\hat{Y},t}}{\hat{\sigma}_{\hat{Y}\hat{u},t}}(\hat{\rho}_{T,t} - \rho_t) &\xrightarrow{d} N(0,1)\,, \\ \frac{B_{1t}}{\widehat{B}_{3t}}(\hat{\alpha}_{T,t} - \alpha_t) &\xrightarrow{d} N(0,1)\,, \end{aligned} \tag{4.20}$$

which allows the construction of point-wise confidence bands for all parameters of interest.
In light of our criticism on distinguishing true from spurious long memory solely based on a spectral rate condition, we establish:

Corollary 1 (Spectral Behavior of RC Models)**.** *Under Assumption 4.2 the pseudo spectral density of the process* $\{y_t^* := (y_t - \mu_t)\}$ *for* $t = 1,\ldots,T$*, generated by (4.16) with* $\alpha_t \equiv 0$*, suffices for* $T \to \infty$*:*

$$f_{y_t^*}(\lambda) = O_p(1) \quad \textit{for} \quad \lambda \in [-\pi,\pi]\,.$$

Proof: *See the Appendix.*

Our Corollary 1 gives the important result that the process $\{y_t^*\}$ can be subsumed under Definition 4.1 since its pseudo spectral density behaves like a true short memory process across all frequencies.

Any rejection observed when applying a test on spurious long memory, like the one proposed by Qu (2011), must therefore be interpreted as being spurious if one distinguishes true from spurious long memory solely based on a spectral rate paradigm.

4.3 Monte Carlo Simulations

After confirming that LS processes and RC models of the form given in our Corollary 1 can be subsumed under Definition 4.1, we now turn to analyze the finite sample effects these models have on the size of the Qu (2011) test on spurious long memory through an extensive Monte Carlo simulation study. First, we consider LS processes from Subsection 4.2.1 and define DGP1, $\{x_t\}$ for $t = 1,\ldots,T$, by:

$$\begin{aligned}(1-\phi L)(1-L)^d x_t &= \sigma_t \tilde{z}_t =: \varpi_t\,, \\ \sigma_t^2 &= \sigma_1^2 + (\sigma_2^2 - \sigma_1^2)\mathbb{I}(t \geq \lfloor \tau T \rfloor)\,, \\ \nu &:= \sigma_2/\sigma_1 = 1/3\,,\end{aligned} \tag{4.21}$$

where $\tilde{z}_t \overset{iid}{\sim} N(0,1)$ and $\tau \in \{0.25, 0.5, 0.75\}$ denotes the relative break fraction. This heteroskedastic $ARFIMA(p,d,q)$ model which allows for a single deterministic break in the unconditional variance, as well as the corresponding parametrization, are taken from Cavaliere et al. (2015b) (see also (4.4) and c.f. (4.13)). Furthermore, we choose $\phi \in \{-0.5, 0, 0.5\}$, $d \in \{0, 0.2, 0.4\}$, sample sizes $T \in \{500, 1000, 3000, \ldots, 9000\}$, a nominal significance level $\alpha = 5\%$ and report the rejection frequencies obtained through $M = 10{,}000$ replications. The test statistic is calculated as in (4.9), using various frequency ordinates $m = \lfloor T^b \rfloor$ with bandwidths $b \in \{0.5, 0.55, \ldots, 0.8\}$ and with trimming parameter $\epsilon = 0.02$ which is in line with the recommendations of Qu (2011) for the sample sizes considered here.[4] The results are reported in Figures 4.1 - 4.2 and in Figures 4.6 - 4.7 which are moved to the appendix for ease of exposition.

If the test is applied on the pure series we find that a bias is present already when using small bandwidths, leading to a liberal behavior of the test which prevails even in largest samples and completely invalidates correct statistical inference. The effect is slightly more pronounced for higher values of d and decreasing with the relative breakpoint τ lying closer at the end of the sample. Naturally, the bias is stronger if short-run dynamics are present ($\phi \neq 0$) which can be attributed to the resulting high-frequency leakage when calculating the periodogram. Therefore, we also analyzed the extent of the effect if the prewhitening setup of Qu (2011) is used in a prior step. In this procedure a low-order $ARFIMA(p,d,q)$ model (4.4) is estimated and the resulting short-run parameters are subsequently extracted from the DGP, which robustifies the test against the influence of high-frequency leakage when calculating the periodogram.[5] We find ce-

[4] All simulations carried out with DGP1 were repeated with $\epsilon = 0.05$ which leads to qualitatively identical results. Moreover, further simulations with MA, ARMA components and $\nu = 3$ in (4.21), were additionally conducted and led to similar results which are available upon request.

[5] We refer to Qu (2011) for an exact description of the individual steps of his prewhitening procedure.

teris paribus that the prewhitening of Qu (2011) indeed decreases the bias, but nevertheless the test remains liberal showing no proper size control even in largest samples and especially when τ lies close to the beginning of the sample. Most important this effect is especially pronounced under the bandwith recommendation $b = 0.7$ given by Qu (2011) for practical applications.

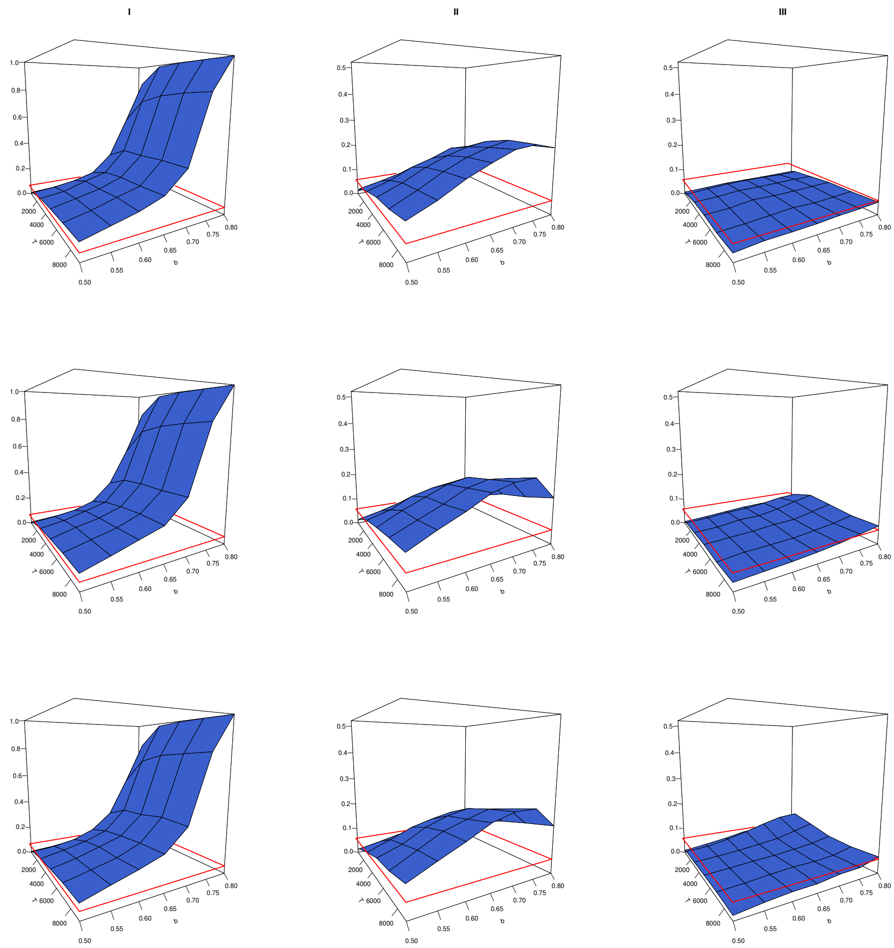
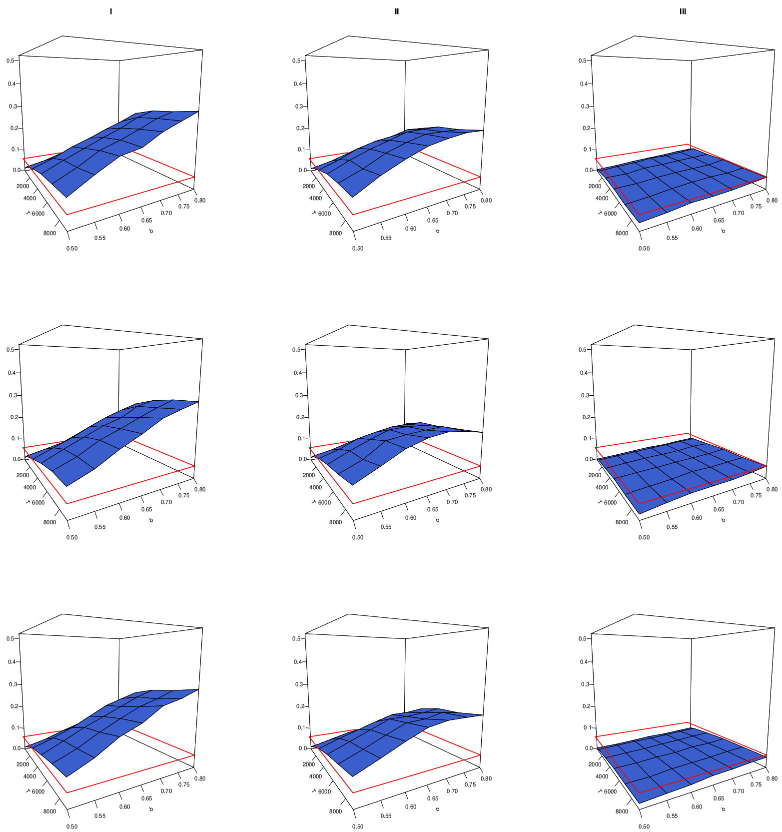


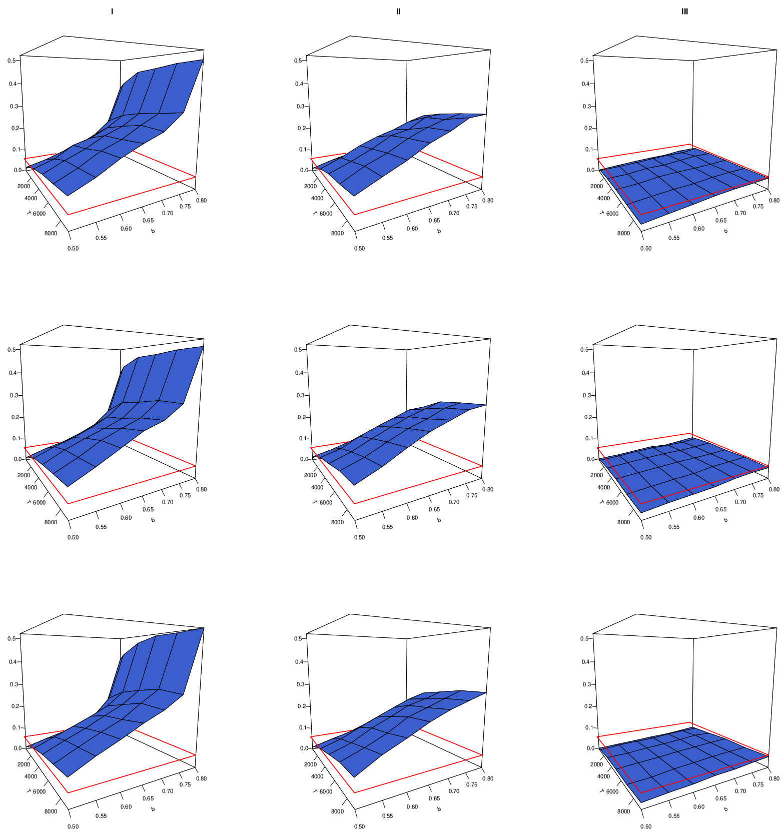
Figure 4.1: The 3D plots report the rejection frequency for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory, using trimming parameter $\epsilon = 0.02$, dependent on the sample size T and frequency bandwidth b under DGP1 (4.21) with parameters $\phi = 0.5$, $\tau = 0.25$, $\nu = 1/3$ and $d \in \{0, 0.2, 0.4\}$ by row 1-3, respectively. The columns report results if no prewhitening procedure is applied (I), if the prewhitening procedure of Qu (2011) is considered (II) and in case our extended prewhitening procedure is used (III).

Therefore, these simulations confirm that a LS process, represented by DGP1, is indeed capable of invalidating statistical inference on spurious long memory, easily leading to a spurious conclusion, when in fact a true short or long memory process in the sense of Definition 4.1 is observed.[6] To alleviate the bias induced by the unconditional heteroskedasticity of a LS process represented by (4.13), we propose an extension of the prewhitening procedure of Qu (2011).

[6]Note that these simulations can also be interpreted as analyzing the finite sample performance of the Qu (2011) test in case of loosening his Assumption 2 which allows the error terms to display at most conditional heteroskedasticty.



(a) $\phi = 0$ and $\tau = 0.25$



(b) $\phi = -0.5$ and $\tau = 0.25$

Figure 4.2: The 3D plots report the rejection frequency for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory, using trimming parameter $\epsilon = 0.02$, dependent on the sample size T and frequency bandwidth b under DGP1 (4.21) with parameters ϕ and τ as given in (a) and (b), $\nu = 1/3$ and $d \in \{0, 0.2, 0.4\}$ by row 1-3, respectively. The columns report results if no prewhitening procedure is applied (I), if the prewhitening procedure of Qu (2011) is considered (II) and in case our extended prewhitening procedure is used (III).

Denote by $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d)'$ a vector containing the short-run dynamics and the memory parameter driving $\{v_t\}$ in (4.13) and equivalently in (4.21). Cavaliere et al. (2017) prove under some regularity conditions that the Conditional Sum of Squares (CSS) (Quasi Maximum Likelihood) estimator $\hat{\beta}$ is consistent for β even under conditional and unconditional heteroskedasticity of i.a. the form studied in this paper. We use this key result to propose an extension of the prewhitening procedure of Qu (2011) which is carried out as follows when dealing with processes of the type given in (4.13) and equivalently in (4.21):

1. Define a model set with a realistic conjecture concerning the upper bound for the model order p and q. Following the arguments of Sin and White (1996) the usual information criteria remain valid in this setup (see also Cavaliere et al. (2015b)) and can be used to consistently specify the model if its true order is a subset of the studied model set.

2. For a given model order p^* and q^*, obtain $\hat{\beta}$ through CSS estimation and construct the residual series $\{\hat{\varpi}_t\}$ (see (4.21)), where $\hat{\varpi}_t \xrightarrow{p} \varpi_t$ for $T \to \infty$ is ensured by Cavaliere et al. (2017). By constructing the squared residual series $\{\hat{\varpi}_t^2\}$ it is straightforward to see that any shift in the scaling factor σ_t becomes a shift in the unconditional mean of the process.[7]

3. Use the procedure of Bai and Perron (1998, 2003a) to consistently estimate the unknown breakpoints and breakdates in $\{\hat{\varpi}_t^2\}$ utilizing their $\sup F$-type l vs. $(l+1)$ test.[8] This approach only requires a realistic conjecture for the upper bound of possible breakpoints present in the unconditional mean and is known to display good finite sample results in particular if breakpoints occur at the beginning or the end of the sample.[9] Obtain the estimated breakpoints and corresponding breakdates and subsequently demean the different regimes resulting in the prewhitened mean-zero residual series $\{\hat{\varpi}_t^{2*}\}$.

4. Apply the test of Qu (2011) on the transformed mean-zero residual series $\{\hat{\varpi}_t^*\}$ to test for spurious long memory.

In a next step we analyzed the extent of the bias correction on the size of the Qu (2011) test if our extended prewhitening procedure is applied. To sharpen the analysis we imposed knowledge of the model order and in a first setup specified a single existing breakpoint as an upper bound for the algorithm of Bai and Perron (1998, 2003a). We find that independently of ϕ and τ the liberal bias induced by the unconditional heteroskedasticity completely vanishes leading to a slightly conservative behavior of the test which however is a well known phenomenon already discussed in Qu (2011). Moreover, our correction shows good results already in small sample sizes and especially for a bandwidth choice of $b = 0.7$ which remains the recommendation for

[7] Note that the residual series $\{\hat{\varpi}_t\}$ is in any case asymptotically an iid process with different white noise regimes. Therefore, following Dittmann and Granger (2002) the squared series $\{\hat{\varpi}_t^2\}$ remains I(0).

[8] An alternative approach would be to time-transform $\{\hat{\varpi}_t\}$ using the residual variance profile proposed by Cavaliere and Taylor (2007). Although this procedure is capable of dealing with a broader class of unconditional heteroskedasticity than studied in this paper, it lacks the possibility to consistently estimate the break dates.

[9] We follow a specific-to-general modeling strategy and use their $\sup F$-type l vs. $(l+1)$ test to determine an appropriate number of possible breakpoints. A detailed description of the algorithm is given in the appendix.

practical use. Furthermore, we analyzed the effect if ceteris paribus an overspecification of two possible breakpoints as an upper bound in the third step of our algorithm is imposed. The results, reported in Table 4.5 in the appendix, clearly show that this only has a negligible effect apart from a greater computational effort. As shown in Bai and Perron (1998, 2003a), their procedure remains consistent since $\{\hat{\varpi}_t^{2*}\}$ exhibits only a single breakpoint in the unconditional mean. The Qu (2011) test, although still conservative, nevertheless approaches its nominal significance level with an increasing sample size.
Complementary, we also considered applying the Qu (2011) test with our extended prewhitening procedure in case DGP1 displays no break in its unconditional variance ($\nu = 1$ in (4.21)) creating a scenario where a correction for unconditional heteroskedasticity is in fact not needed. The results are reported in Figure 4.8 in the appendix. Compared to the classic prewhitening approach which is in fact required in this setup, we find that overall our procedure only leads to a slightly more conservative behavior of the Qu (2011) test. This demonstrates that one does not pay a too high price for the additional flexibility in a classic conditional homoskedastic setting.
In conclusion, applying our extended prewhitening procedure recuperates correct statistical inference on spurious long memory, since it robustifies the test against a LS process of the form studied in (4.13) and generated by DGP1 which is in line regarding Definition 4.1 of true short and long memory.
Next we analyze the potential power loss of the Qu (2011) test in case our extended prewhitening procedure is applied when in fact the DGP follows a spurious long memory process. We consider all processes with corresponding parametrization also studied in the simulations of Qu (2011) and adopt his notation for reasons of convenience. Thus, let $\{y_t\}$ for $t = 1, \ldots, T$, be respectively given by:

1. Non-stationary random level shift: $y_t = \mu_t + \varepsilon_t$, $\mu_t = \mu_{t-1} + \pi_t\eta_t$, $\pi_t \overset{iid}{\sim} B(1, 6.1/T)$, $\varepsilon_t \overset{iid}{\sim} N(0,5)$, $\eta_t \overset{iid}{\sim} N(0,1)$.

2. Stationary random level shift: $y_t = \mu_t + \varepsilon_t$, $\mu_t = (1-\pi_t)\mu_{t-1} + \pi_t\eta_t$, $\pi_t \overset{iid}{\sim} B(1, 0.003)$, ε_t and $\eta_t \overset{iid}{\sim} N(0,1)$.

3. Markov switching with iid regimes: $y_t \overset{iid}{\sim} N(1,1)$ if $s_t = 0$ and $y_t \overset{iid}{\sim} N(-1,1)$ if $s_t = 1$, with state transition probabilities $p_{10} = p_{01} = 0.001$.

4. Markov switching with GARCH regimes: $r_t = \sqrt{h_t}\varepsilon_t$ and $h_t = 1 + 2s_t + 0.4r_{t-1}^2 + 0.3h_{t-1}$, where $\varepsilon_t \overset{iid}{\sim} N(0,1)$, $s_t = 0, 1$ for $p_{10} = p_{01} = 0.001$ and $y_t = \log r_t^2$.

5. White noise with a monotonic deterministic trend: $y_t = 3t^{-0.1} + \varepsilon_t$, $\varepsilon_t \overset{iid}{\sim} N(0,1)$.

6. White noise with a non-monotonic deterministic trend: $y_t = \sin(4\pi t/T) + \varepsilon_t$, $\varepsilon_t \overset{iid}{\sim} N(0,3)$.

The results of this simulation, reported in Table 4.1, show that our extended prewhitening procedure does not inflict a notable power loss when applied to a spurious long memory process

$\#\tau$	1		2		1		2	
T	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.02$	$\epsilon = 0.05$
	Non-stationary random level shift				Stationary random level shift			
500	55.50	53.25	54.20	52.41	41.25	40.75	41.77	41.27
1000	80	78.50	79.22	77.72	77.75	77.50	78.10	77.68
3000	96	94.50	95.57	94.05	99.50	99.50	99.65	99.70
5000	97.50	96.75	97.17	96.59	100	100	99.95	99.97
7000	98	98	98.14	97.60	100	100	99.99	100
9000	98.75	98.25	98.67	98.38	100	100	100	100
	Markov Switching with iid regimes				Markov Switching with GARCH regimes			
500	31.25	32.25	31.96	32.93	2.75	4	2.58	3.98
1000	57.25	58.25	57.52	58.47	8	7.75	7.13	7.53
3000	93.50	93.75	93.90	94.03	52.75	30.75	50.74	28.19
5000	99	99	99.10	99.12	82.50	67	83.02	66.18
7000	100	100	99.89	99.90	95	86.25	94.69	85.83
9000	100	100	99.96	99.96	98.50	95.75	98.38	95.59
	White noise with a monotonic det. trend				White noise with a non-monotonic det. trend			
500	10.50	10.50	10.49	10.21	99	93	98.85	92.83
1000	30.50	28.25	28.16	26.55	100	100	100	99.93
3000	88.25	78.25	85.98	74.53	100	100	100	100
5000	98.75	96	98.33	94.96	100	100	100	100
7000	100	99	99.83	98.91	100	100	100	100
9000	100	100	100	99.76	100	100	100	100

Table 4.1: Reported are the rejection frequencies for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory using trimming parameter $\epsilon = \{0.02, 0.05\}$ and a frequency bandwidth choice of $b = 0.7$, in case our extended prewhitening procedure is used with $\#\tau \in \{1,2\}$ denoting the maximum value of possible breakpoints in the algorithm of Bai and Perron (1998, 2003a). The DGPs are generated by various spurious long memory processes with sample sizes T and parameters taken from Qu (2011).

obeying Definition 4.2. The Qu (2011) test clearly remains consistent and also an overspecification concerning the upper bound of possible breakpoints in the algorithm of Bai and Perron (1998, 2003a) induces only negligible power losses that are more pronounced in smaller sample sizes.

In a next step we turn to analyze the finite sample performance of the Qu (2011) test in case the DGP follows a RC model as discussed in Subsection 4.2.2. Following Giraitis et al. (2014), we let DGP2, which fulfills Assumption 4.2, be generated by $\{y_t\}$ for $t = 1, \ldots, T$, with:

$$
\begin{aligned}
y_t &= \rho_{t-1} y_{t-1} + e_t\,, \\
\rho_t &= \rho \frac{a_t}{\max\limits_{0\le j\le t} |a_j|}\,, \\
a_t - a_{t-1} := v_t &= (1-\phi L)^{-1}(1-L)^{-d}\varepsilon_t\,,
\end{aligned}
\tag{4.22}
$$

where the parametrization e_t and $\varepsilon_t \overset{iid}{\sim} N(0,1)$, $\rho = 0.9$, $\phi \in \{0, 0.8\}$, $d \in \{0, 0.2, 0.4\}$ and the sample sizes $T \in \{50, 100, 200, 400, 800\}$ are directly taken from these authors. Note that (4.22) fulfills our Corollary 1 since its random intercept is $\alpha_t \equiv 0$ and thus falls under Definition 4.1 of a true short memory process. The results of these simulations are reported in Figure 4.3 and 4.9 where the latter is moved to the appendix. If the raw series are applied to the Qu (2011) test, we again observe that even in largest samples a non-negligible liberal bias is induced which is increasing with a higher bandwidth choice. The results are more pronounced the higher the persistence in $\{v_t\}$. Importantly, the bandwidth recommendation $b = 0.7$ given by Qu (2011) clearly leads to a complete invalidation of statistical inference. Although a smaller bandwidth choice of for example $b = 0.55$ shows good size control, its use is clearly not recommendable in practice since further unreported simulations reveal that it leads to notable power losses (c.f. Qu (2011)).

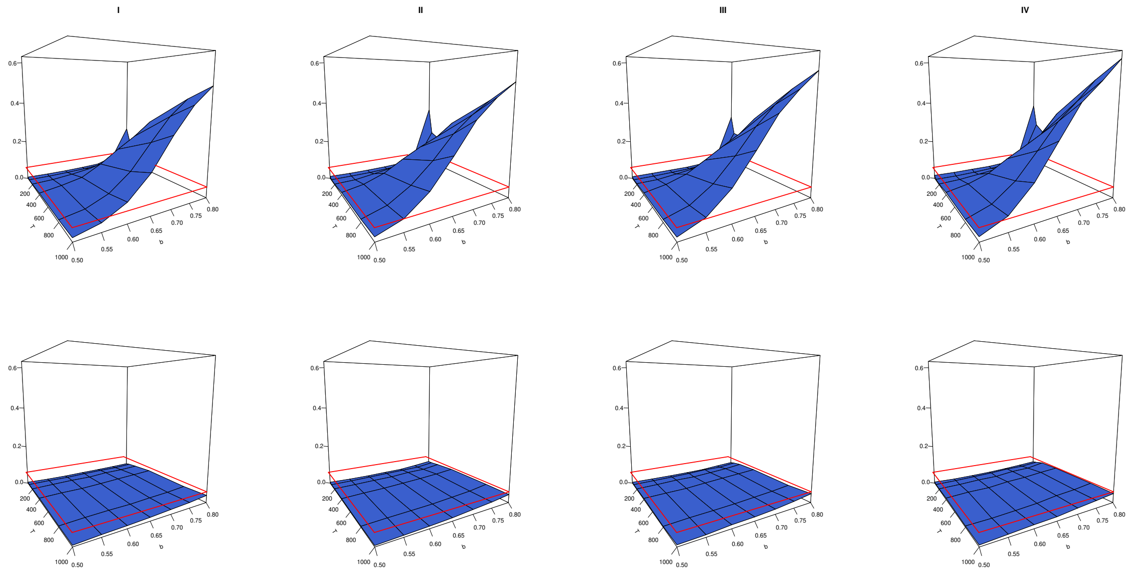


Figure 4.3: The 3D plots report the rejection frequency for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory, using trimming parameter $\epsilon = 0.05$ ($\epsilon = 0.02$) for $T < 500$ ($T > 500$) dependent on the sample size T and frequency bandwidth b under DGP2 (4.22). The columns report results if $\{v_t\}$ displays the following short-run dynamics: $\phi = 0$ and $d = 0$ (I), $\phi = 0.8$ and $d = 0$ (II), $\phi = 0$ and $d = 0.2$ (III), $\phi = 0$ and $d = 0.4$ (IV). The first row gives results if the pure series and the second if the residual series $\{\hat{e}_t\}$ is applied, constructed with parameter estimates using the Gaussian kernel from (4.19) with MSE-optimal bandwidth choice $H = \sqrt{T}$.

Since neither the prewhitening procedure of Qu (2011) nor our extended prewhitening procedure are valid under DGP2, we employ the nonparametric kernel-based estimators and kernels (4.18)-(4.19) with corresponding MSE-optimal bandwidth choice $H = \sqrt{T}$ of Giraitis et al. (2014) to compute the residual series $\{\hat{e}_t\}$ from (4.22) on which we subsequently applied the Qu (2011) test. Although we once more observe a conservative behavior, a bandwidth choice of $b = 0.7$ now achieves good size control already in smaller sample sizes and therefore, in combination with the results in Table 4.1, remains the recommendation for practical use. Moreover, utilizing the Gaussian kernel from (4.19) gives best finite sample results compared to the flat or Epanechnikov kernel which is in line with the findings of Giraitis et al. (2014) (see Figure 4.9 in the appendix). We conclude this section by emphasizing that both DGPs, which respectively represent a subset of LS processes and RC models, notably deteriorate statistical inference on spurious long memory in finite samples if applied to the Qu (2011) test in an unfiltered form or if an insufficient prewhitening procedure is used. Although both process classes violate the assumptions of Qu (2011) they nevertheless are subsumed under Definition 4.1 of true short or long memory. Therefore, we criticize the existing paradigm of distinguishing true and spurious long memory solely based on a spectral rate condition. Our simulations reveal that true short or long memory processes defined in a wider sense can easily be attributed to the spurious long memory class due to a spurious rejection of the null.

4.4 Empirical Analysis

In light of our simulation results we now revisit two recently studied empirical applications from the literature where LS processes and RC models are applied and analyze the potential of a spurious rejection when conducting statistical inference using the Qu (2011) test.

4.4.1 Commodity Application (Locally Stationary Processes)

The first data set is studied in Cavaliere et al. (2015b) and originates from Westerlund and Narayan (2013) to whom we refer concerning the exact details of its construction. It consists of daily data of logged spot prices (s_t) and corresponding one-period futures contract prices (f_t) of gold, silver, platinum and crude oil commodities and spans from July 5, 2005 - November 22, 2011, yielding a total of $T = 1665$ observations.[10] In particular the former authors test the well known weaker form of the Efficient Market Hypothesis (EMH) (e.g. Luo (1998) and Westerlund and Narayan (2013)) which states that s_t and $f_t^{(k)} \sim I(1)$ form a strong fractional cointegration relationship with cointegrating vector $(1,-1)'$ such that the spread $\{u_t^{(k)}\}$ suffices for some small generic constant c:

$$E(u_t^{(k)} = s_t - f_{t-k}^{(k)} | \mathcal{I}_t) = c \, ,$$

where $\mathcal{I}_t$ is the given information set, the sigma-algebra, containing all $(s_j, f_j)'$ for $j = 0, \ldots, t$. Thus, one can define $\tilde{b}$ such that $\{u_t^{(k)}\} \sim I(0 \leq 1 - \tilde{b} < 0.5)$ holds. Since $\{u_t^{(k)}\}$ is allowed to display unconditional heteroskedasticity, Cavaliere et al. (2015b) use four different corresponding tests of Cavaliere and Taylor (2007) and find significant breaks in the unconditional variance in all but the silver commodity series. By utilizing their Wild bootstrap score-based test on long memory, which is robust against i.a. this kind of non-stationary volatility, they find evidence that the weak form of EMH holds for silver and platinum with $\{u_t^{(1)}\} \sim I(0)$, as well as gold with $\{u_t^{(1)}\} \sim I(0 \leq d < 0.5)$ and advocate a heteroskedastic $ARFIMA(p,d,q)$ modelling approach for Δs_t, $\Delta f_t^{(1)}$ and $u_t^{(1)} := s_t - f_{t-1}^{(1)}$. As discussed in Subsection 4.2.1 this model is nested in the class of LS processes and thus we extend their analysis in two ways. First, we analyze the magnitude of distortion the unconditional heteroskedasticity has when conducting inference on spurious

[10] We thank Morten Ø. Nielsen for sharing the data with us. Cavaliere et al. (2015b) originally obtained the raw untransformed data from Bloomberg.

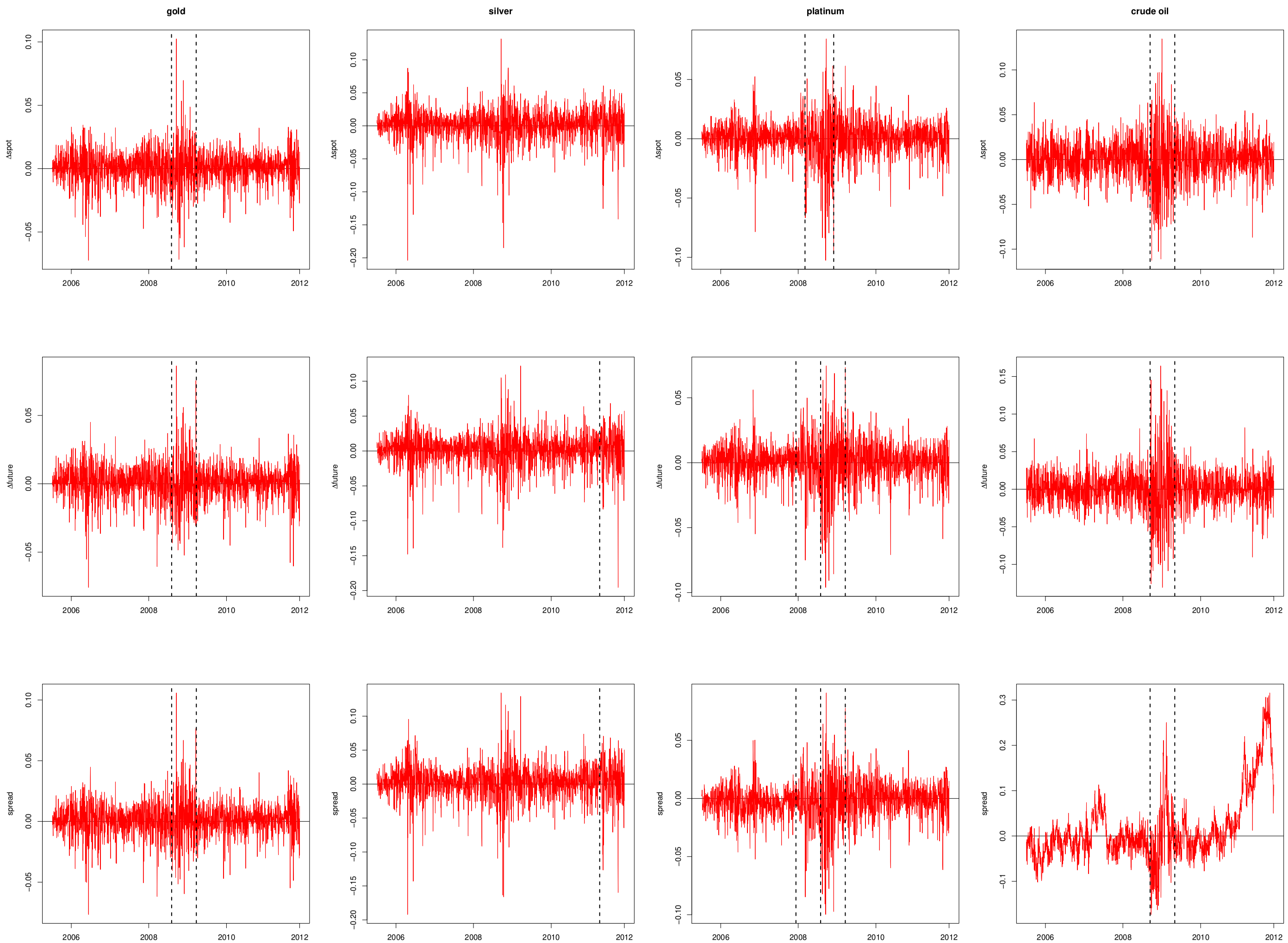


Figure 4.4: Time series plots of Δs_t, $\Delta f_t^{(1)}$ and $u_t^{(1)} = s_t - f_{t-1}^{(1)}$ of gold, silver, platinum and crude oil, shown in panel 1 to 4, respectively. The breakpoints in the unconditional variance are obtained using the algorithm of Bai and Perron (1998, 2003a) ($\alpha = 5\%$), embedded in our extended prewhitening procedure with prespecified $\#\tau = 5$, are marked by dashed vertical lines.

b	corrected				uncorrected				corrected				uncorrected			
	gold	silver	platinum	crude oil	gold	silver	platinum	crude oil	gold	silver	platinum	crude oil	gold	silver	platinum	crude oil
	Δs_t								$\Delta f_t^{(1)}$							
0.70	1.034	0.326	0.540	0.457	0.981	0.326	**1.244**	**1.260***	0.957	0.430	1.026	0.577	0.942	0.461	**1.324***	**1.697****
							(1.118)	(1.252)							(1.252)	(1.517)
0.72	0.979	0.342	0.736	0.600	1.019	0.342	**1.423***	**1.271***	0.926	0.448	**1.169**	0.551	1.029	0.424	**1.470***	**1.535****
							(1.252)				(1.118)				(1.252)	(1.517)
0.74	1.014	0.402	0.676	0.534	1.022	0.402	**1.288***	1.096	0.958	0.368	1.040	0.485	1.027	0.332	**1.331***	**1.368***
							(1.252)								(1.252)	
0.76	1.106	0.612	0.620	0.693	**1.181**	0.612	**1.130**	**1.229**	1.072	0.559	0.949	0.632	**1.207**	0.447	**1.154**	**1.410***
					(1.118)		(1.118)						(1.118)		(1.118)	(1.252)
0.78	**1.224**	0.744	0.758	0.642	**1.328***	0.744	**1.502***	1.075	**1.256***	0.956	**1.124**	0.609	**1.377***	0.874	**1.370***	**1.286***
	(1.118)				(1.252)		(1.252)		(1.252)		(1.118)		(1.252)		(1.252)	
0.80	1.110	0.636	0.694	0.630	1.103	0.636	**1.391***	**1.262***	1.091	0.654	1.045	0.669	1.090	0.662	**1.339***	**1.433***
							(1.252)								(1.252)	

Table 4.2: Reported are results of the Qu (2011) test statistic (4.9), applied to the uncorrected and the corrected residuals of the differenced logged spot (Δs_t) and corresponding differenced one-period futures ($\Delta f_t^{(1)}$) commodity data for different bandwidths b and $\epsilon = 0.02$. Bold-faced values indicate significance at the nominal 10% level; an additional * (**) indicates significance at the nominal 5% (1%) level. Respective critical values are given in parentheses.

long memory using the Qu (2011) test and complementary to which extent our extended prewhitening procedure alleviates the effect. Secondly, using the algorithm of Bai and Perron (1998, 2003a) embedded in our procedure, we date the breakpoints in the unconditional variance which allows some implications concerning the economic background.

We use the model order specified in Table 6 on p. 570 of Cavaliere et al. (2015b) to estimate the uncorrected residual series of Δs_t, $\Delta f_t^{(1)}$ and $u_t^{(1)}$ via the CSS estimator which basically corresponds to the prewhitening procedure of Qu (2011) with pre-determined model order. Since his approach does not correct for unconditional heteroskedasticity, we complementary also applied our extended prewhitening procedure to the same commodity series and obtained the corrected residual series. Both types of residual series are then applied to the Qu (2011) test using all practically relevant bandwidths $b \in \{0.7, 0.72, \ldots, 0.8\}$. We summarize our results in Tables 4.2 and 4.3, whereas Figure 4.4 visualizes Δs_t, $\Delta f_t^{(1)}$ and $u_t^{(1)}$ in several plots which also indicate all breakpoints found in the unconditional variance obtained through the algorithm of Bai and Perron (1998, 2003a) ($\alpha = 5\%$), embedded in our extended prewhitening procedure with prespecified $\#\tau = 5$ maximum allowed number of breakpoints by dashed vertical lines. Note that Cavaliere et al. (2015b) find non-stationary long memory in the spread of the crude oil series ($\hat{d} = 0.78$) which imposes an additional violation of the Assumptions of Qu (2011) that require $|d| < 0.5$. Therefore, we omit any further interpretation of these results in Table 4.3 and report them only for the sake of completeness.[11]

We find that the uncorrected residual series of platinum and crude oil in Table 4.2 clearly lead to a rejection of the Qu (2011) test indicating spurious long memory, whereas the corrected residual series of the same commodities show no rejection of the null apart from two negligible exceptions. The corresponding gold series shows qualitatively the same results which are however somewhat

[11] All other series display an estimate of d close to zero (c.f. Table 6 on p. 570 of Cavaliere et al. (2015b))

b	corrected					uncorrected			
	gold	silver	platinum	crude oil	|	gold	silver	platinum	crude oil
					$u_t^{(1)}$				
0.70	0.628	0.515	0.922	**1.151**	|	0.614	0.420	**1.233**	**1.457*****
				(1.118)	|			(1.118)	(1.252)
0.72	0.592	0.655	**1.220**	**1.490***	|	0.634	0.511	**1.517****	**1.544****
			(1.118)	(1.252)	|			(1.517)	
0.74	0.678	0.586	**1.137**	**1.444***	|	0.680	0.373	**1.409***	**1.537****
			(1.118)	(1.252)	|			(1.252)	(1.517)
0.76	0.862	0.684	1.033	**1.951****	|	0.938	0.635	**1.187**	**2.024****
				(1.517)	|			(1.118)	(1.517)
0.78	**1.129**	0.928	**1.427***	**2.095****	|	**1.121**	0.903	**1.556****	**2.108****
	(1.118)		(1.252)	(1.517)	|	(1.118)		(1.517)	
0.80	0.909	0.655	**1.308***	**2.738****	|	0.848	0.742	**1.327***	**2.852****
			(1.252)	(1.517)	|			(1.252)	(1.517)

Table 4.3: Reported are results of the Qu (2011) test statistic (4.9), applied to the uncorrected and the corrected residuals of the commodity spread data ($u_t^{(1)} = s_t - f_{t-1}^{(1)}$) for different bandwidths b and $\epsilon = 0.02$. Bold-faced values indicate significance at the nominal 10% level; an additional * (**) indicates significance at the nominal 5% (1%) level. Respective critical values are given in parentheses.

weaker pronounced. These findings are completely in line with our simulation results from Section 4.3. The unconditional heteroskedasticity still present in the uncorrected residual series leads to a spurious rejection of the Qu (2011) test, whereas the corrected residual series controls for this effect due to our extended prewhitening procedure. Interestingly, the residual series of silver shows no rejection at all independently of which prewhitening procedure is used (see also $u_t^{(1)}$ of silver in Table 4.3). This is nevertheless in line with our findings since Cavaliere et al. (2015b) find no unconditional heteroskedasticity in these series in the first place. Therefore, no spurious rejection should occur and our extended prewhitening procedure is equivalent up to a nuisance step, concerning the algorithm of Bai and Perron (1998, 2003a), to the prewhitening setup of Qu (2011) which remains valid in this situation. Nevertheless, Figure 4.4 shows a single breakpoint at the end of the sample of $\Delta f_t^{(1)}$ and $u_t^{(1)}$ of silver which might seem contradictive at first sight. A possible explanation might be that the localization of the breakpoint and/or its magnitude is to weak to induce a spurious rejection.

Concerning the results of $u_t^{(1)}$ of platinum in Table 4.3, we again observe that the uncorrected residuals induce a far more significant rejection over several considered bandwidths than the corrected residual series which again is completely in line with our simulation studies. However, the correction effect of our extended prewhitening procedure is less pronounced and interestingly Figure 4.4 shows three breakpoints. A possible interpretation might be that $\Delta f_t^{(1)}$ and $u_t^{(1)}$ of platinum displays no sudden shifts in its unconditional variance like modelled in (4.13) but rather a smooth trend.[12]

Summing up, our results on one hand show that spurious rejections induced by unconditional heteroskedasticity indeed occur in these series and distort inference on spurious long memory as predicted by our simulation studies and on the other hand strengthen the analysis of Cavaliere et al. (2015b) by supporting their arguments to model Δs_t, $\Delta f_t^{(1)}$ and $u_t^{(1)}$ with heteroskedastic

[12]See Assumption $\mathcal{V}$ on p. 559 of Cavaliere et al. (2015b)

$ARFIMA(p,d,q)$ processes. Lastly, please note that the localization of the breakpoints in general imply that the unconditional variance significantly increased during the time of the Great Recession (2007-2009) as a result of the Great Financial Crisis (2007-2008). A detailed analysis of this issue is however left for future research.

4.4.2 Inflation Rates Application (Random Coefficient Models)

In our second analysis we focus on monthly (seasonally adjusted) inflation rates of the G7 countries Canada, France, Germany, Italy, Japan, UK and US. The modelling approach for inflation rates has been quite controversially discussed in the recent literature. On one hand a true short or long memory process in line with Definition 4.1 and more concrete in form of a $ARFIMA(p,d,q)$ process (4.4) (e.g. Hassler and Wolters (1995)) is advocated which may even display one or multiple switches in the memory parameter (c.f. Hassler and Meller (2014)). However, on the other hand the literature identifies level shifts (e.g. Bos et al. (1999) and Kumar and Okimoto (2007)) or non-linear changes in the mean of the process (e.g. Baillie and Kapetanios (2007)) as a source in the data that ultimately causes spurious long memory, subsuming at least this part of the process under Definition 4.2. Most recently, Rinke et al. (2017) find spurious long memory mainly induced by smooth trends in all G7 countries excluding the US and propose a semiparametric fractional autoregressive ($SEMIFAR$) modelling approach. However, since Giraitis et al. (2014) apply their RC models on i.a. quarterly inflation rates of Australia, Canada, Japan, Switzerland, UK as well as the US and in light of our findings in Corollary 1, we use their nonparametric kernel-based estimators to fit corresponding models to the data set of Rinke et al. (2017) and subsequently analyse the potential of a spurious rejection when conducting inference with the Qu (2011) test. The data spans from January 1970 - February 2015, yielding a total of $T = 541$ observations and we refer to Rinke et al. (2017) concerning the exact details of its construction.[13]

Closely following Giraitis et al. (2014), we specify an RC model containing a random intercept term (4.16) and use the estimators (4.18) with Gaussian kernel (4.19), by simultaneously considering a MSE-optimal bandwidth choice of $H = \sqrt{T}$ to estimate $\{\rho_t\}$, $\{\alpha_t\}$ and $\{\mu_t\}$, respectively. We then utilize the asymptotic results of Giraitis et al. (2014) given in (4.20) to construct point-wise 90% confidence bands for the parameter estimates and display all results in Figure 4.5. Firstly, we observe that all series except the US display a pronounced estimated attractor process $\{\hat{\mu}_t\}$ which is complementary to the findings of Rinke et al. (2017). Secondly, we observe that the estimated AR coefficient process $\{\hat{\rho}_t\}$ indicates highest persistence during the time of the "Great Inflation" from 1970 up to 1980, followed by a decline up to the 2000s due to the "Great Moderation". These findings are especially pronounced for France, Germany, UK and US and in line with those of Giraitis et al. (2014) and for example Cogley et al. (2010). Similar to the results of the former authors we observe that in no case the confidence bands of the $\{\hat{\rho}_t\}$ process

[13]We thank Marie Busch for sharing the data with us. Rinke et al. (2017) originally obtained the raw Consumer Price Index data from Thomson Reuters Datastream.

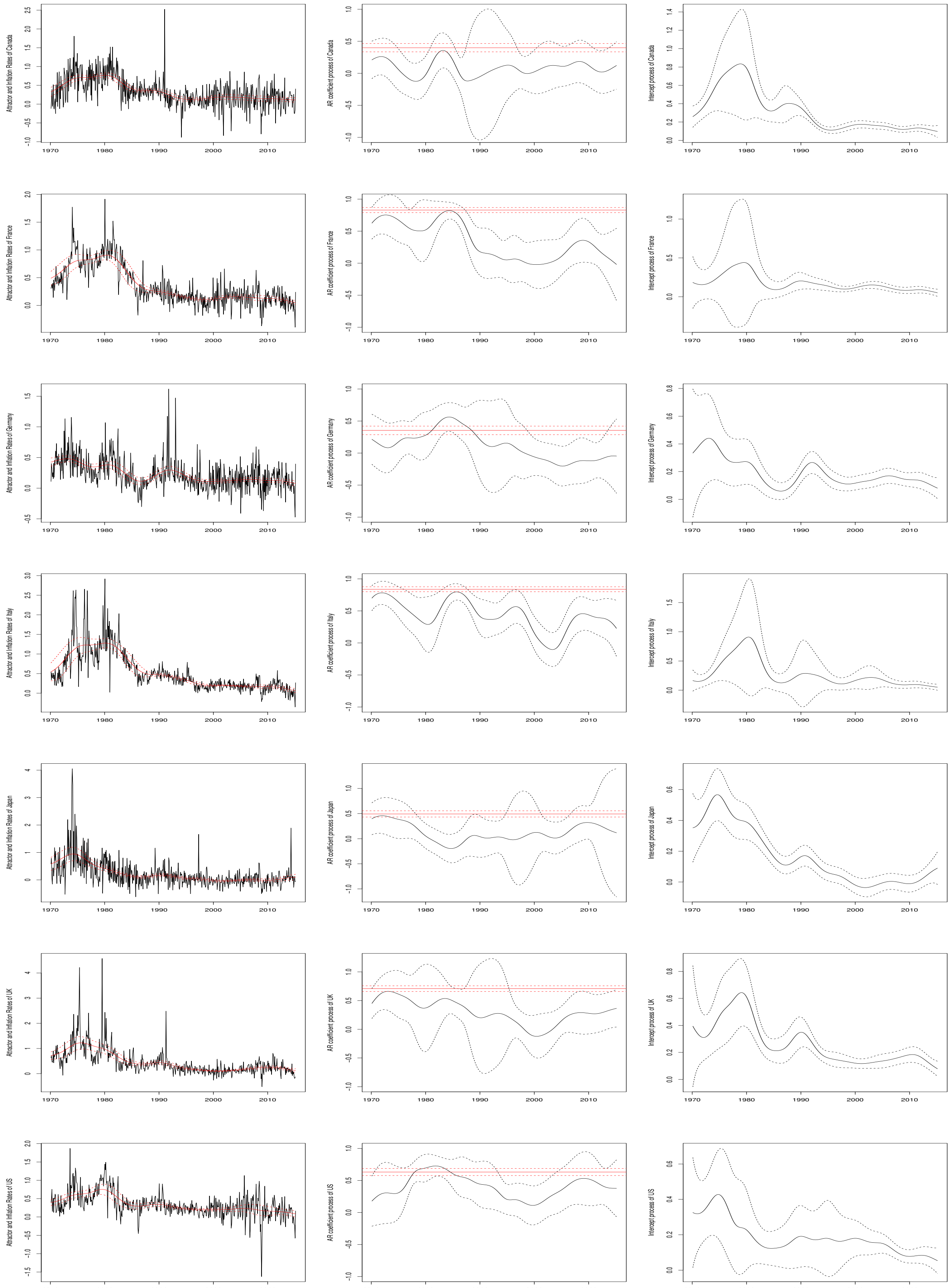


Figure 4.5: Reported by rows are Inflation rates data with attractor (in red), time-varying AR coefficient and intercept, all with their respective 90% confidence bands, obtained using the Gaussian kernel for G7 countries: Canada, France, Germany, Italy, Japan, UK and US. The AR coefficient panels also report the estimate in a fixed coefficient AR(1) model, together with its 90% confidence bands (in red).

b	residual series: $\hat{u}_t$							extracted attractor series: $y_t^* := (y_t - \hat{\mu}_t)$							pure series: y_t						
	Canada	France	Germany	Italy	Japan	UK	USA	Canada	France	Germany	Italy	Japan	UK	USA	Canada	France	Germany	Italy	Japan	UK	USA
0.70	1.060	0.960	0.885	**1.208**	**1.182**	1.030	**1.184**	0.988	0.954	0.853	1.025	**1.215**	**1.153**	1.079	**1.607****	**1.972****	**1.797****	**1.686****	**1.353***	**1.286***	**1.143**
				(1.118)			(1.118)					(1.118)	(1.118)			(1.517)			(1.252)		(1.118)
0.72	1.025	0.804	0.889	0.960	**1.153**	0.978	**1.192**	0.954	0.972	0.844	1.001	**1.191**	**1.121**	1.117	**1.823****	**1.793****	**1.677****	**1.715****	**1.410***	**1.395***	0.883
					(1.118)		(1.118)					(1.118)	(1.118)			(1.517)			(1.252)		
0.74	0.969	0.737	0.818	0.879	1.083	0.930	1.103	0.887	0.923	0.801	0.981	**1.139**	1.095	1.051	**2.265****	**2.121****	**2.053****	**1.794****	**1.688****	**1.491***	**1.277****
												(1.118)					(1.517)			(1.252)	(1.517)
0.76	0.932	0.720	0.791	0.881	1.059	0.947	1.065	0.856	0.909	0.778	0.984	1.085	1.112	1.047	**2.412****	**2.185****	**2.137****	**1.639****	**1.967****	**1.293***	**1.223**
																	(1.517)			(1.252)	(1.118)
0.78	0.900	0.699	0.776	0.854	1.060	0.961	1.073	0.826	0.892	0.758	0.973	1.068	**1.127**	1.081	**2.574****	**2.211****	**2.149****	**1.590****	**1.931****	1.076	0.884
													(1.118)				(1.517)				
0.80	0.854	0.837	0.806	0.834	**1.153**	0.890	1.013	0.783	0.852	0.723	0.962	1.045	1.088	1.048	**2.838****	**2.498****	**2.418****	**1.506***	**1.991****	**1.230**	1.026
					(1.118)											(1.517)		(1.252)	(1.517)	(1.118)	

Table 4.4: Reported are results of the Qu (2011) test statistic (4.9), applied to the pure series (y_t), the extracted attractor series ($y_t^* := y_t - \hat{\mu}_t$) and the residuals ($\hat{u}_t$) of the G7 inflation rates for different bandwidths (b) and $\epsilon = 0.02$. Bold-faced values indicate significance at the nominal 10% level; an additional * (**) indicates significance at the nominal 5% (1%) level. Respective critical values are given in parentheses.

are fully covered by those of a fixed coefficient AR(1) model. This demonstrates the benefits of the RC modelling approach implying that inflation persistence is closely linked to dynamic monetary policy rather than rigid price-setting as stated in Giraitis et al. (2014). Lastly, we emphasize that the AR coefficient process of Canada, Germany and Japan display throughout a very low persistence level which is in line with the results of Rinke et al. (2017) who estimate a memory parameter close to zero using the robust Local Whittle estimator of Hou and Perron (2014). In a next step we apply the pure series (y_t), the extracted attractor series ($y_t^* := y_t - \hat{\mu}_t$) that obey our Corollary 1 and the residual series ($\hat{u}_t$) to the Qu (2011) test and report our results in Table 4.4. Ignoring minor deviations, the results for y_t are qualitatively identical to those in Rinke et al. (2017) and are only re-reported here for the sake of completeness. One clearly observes a rejection in favor of spurious long memory for all G7 countries, except for the US where the results are somewhat weaker pronounced. Interestingly, Giraitis et al. (2014) show that the case of a time-varying mean ($\mu_t = g(t/T)$) is nested in their Assumptions concerning the attractor term. Following the arguments of Qu (2011) an RC model including a random intercept term like in (4.16) can therefore be subsumed under Definition 4.2 in this case. Moreover, the results for the extracted attractor series y_t^* show basically no rejection at all, implying that the attractor term $\{\mu_t\}$ is indeed causing the spurious long memory by inducing a time-varying mean in the series. This is an interesting result in its own sense since it extends the class of processes capable of causing spurious long memory. However, these findings are practically equal to those obtained when applying the residual series $\hat{u}_t$ which is at first sight contradictive to the results in our simulation studies in Section 4.3. We would rather expect to observe more spurious rejections induced by the random switching of the AR coefficient process $\{\rho_t\}$. Nevertheless, a possible explanation might be that aside from the rather small sample size, the persistence present in the $\{v_t\}$ process (see (4.22)) is not high enough to lead to the spurious rejections which we ultimately observe in our Monte Carlo simulations.

4.5 Conclusion

This paper criticizes the existing paradigm of distinguishing between true and spurious long memory solely based on a spectral rate condition. We argue that by loosening the usual assumption of second-order stationarity in the sense that only the second moment remains bounded, one is nevertheless able to compute a periodogram that converges to a (pseudo) spectral density in the limit. This entity retains the usual rate condition for frequencies near the origin which is commonly associated with a true negative, short or long memory process. On one hand we consider a subset of LS processes (c.f. Dahlhaus (2000) and Palma and Olea (2010)) in the form of models studied by Demetrescu and Sibbertsen (2016) and Cavaliere et al. (2015b) and on the other hand RC models by Giraitis et al. (2014), take existing results concerning the former and establish concerning the latter that these processes can be subsumed under Definition 4.1 of true short or long memory.

We apply these DGPs to the Qu (2011) test in a large scale Monte Carlo simulation study and find that a non-negligible liberal bias occurs that prevails even in largest samples invalidating statistical inference on spurious long memory. Therefore, not only processes displaying low-frequency contaminations can easily be confused with true long memory, but under some weaker assumptions, true short or long memory processes can be wrongly attributed to the spurious long memory class. Since the prewhitening procedure of Qu (2011) is demonstrated to be ineffective against this type of non-stationarity, we propose an extended prewhitening procedure for the subset of LS processes considered in this paper. Our extended prewhitening builds on results of Qu (2011), Cavaliere et al. (2017) as well as the algorithm of Bai and Perron (1998, 2003a) and reestablishes good size control even in small samples. Although we provide additional robustness to the Qu (2011) test, we are still able to retain good power results under a wide set of processes displaying spurious long memory. Moreover, a potential overspecification of admissible breakpoints is shown to induce only mild losses in terms of size and power.

We then revisit two recently studied empirical applications from the literature where LS processes and RC models are applied and analyze the potential of a spurious rejection when conducting statistical inference on spurious long memory using the Qu (2011) test. On one hand we consider daily data of logged spot prices, corresponding one-period futures contract prices as well as spreads of gold, silver, platinum and crude oil commodities studied by Cavaliere et al. (2015b) and on the other hand monthly (seasonally adjusted) inflation rates of the G7 countries Canada, France, Germany, Italy, Japan, UK and US studied by Rinke et al. (2017). In particular concerning the former data set we clearly observe that a spurious rejection of the true short or long memory null hypothesis occurs, especially when using bandwidths recommended for the use in practice. Nevertheless, our extended prewhitening procedure proves to be effective in these cases.

We conclude this paper by pointing out some avenues for future research. Although both LS processes and RC models almost surely induce unconditional heteroskedasticity and additional switching in the off-diagonal elements of the covariance matrix of the process, to the best of our knowledge no possibility exists to distinguish between these two process classes. Even utilizing the four tests of Cavaliere and Taylor (2007) seems futile since they are based on the residual variance profile which should be affected equally in both cases. Moreover, instead of using our extended prewhitening procedure a direct modification of the Qu (2011) test seems promising. By exchanging the standard periodogram with its localized counterpart (c.f. Dahlhaus (2000) and Palma and Olea (2010)), which will clearly influence the limiting null distribution of the test, at least equal robustness features should be achieved. This however may come at the price of a high power loss. Lastly, an interesting path could be to drop Definition 4.2 of spurious long memory by Qu (2011) and much more consider a redefinition based on the concept of summability as a generalization of the fractional integration paradigm as recently proposed by Berenguer-Rico and Gonzalo (2014).

4.6 Appendix

Proof (Corollary 1). *Note that all referenced equations and Theorems are taken from Giraitis et al. (2014). In this setting the attractor is extracted at each point in time which nests the interesting case of $\alpha_{T,t} \equiv 0$ in (4.16) and results in $E(y_t) \to 0$ for $T \to \infty$. Furthermore, using equation 2.35 on p. 52, a purely deterministic constant attractor $\mu_t = t^{-1}(\mu + \ldots + \mu) = \mu \in \mathbb{R}$ directly results in $E(y_t) \to \mu$ for $T \to \infty$. Thus, this setup can also be interpreted as observing a demeaned model in (4.16). First note that using equation 2.28 on p. 51, we have for $t = 1, \ldots, T$:*

$$\underbrace{(y_t - \mu_t)}_{y_t^*} = \rho_{t-1} \underbrace{(y_{t-1} - \mu_{t-1})}_{y_{t-1}^*} + u_t \,,$$

which allows us to use the results of Theorem 2.2. on p. 50 and corresponding proofs on p. 57, since $\{y_t^*\}$ follows (4.16) with $\alpha_{T,t} \equiv 0$. Assuming mutual independence of ρ_t, u_t and y_0^* $\forall t$, we adopt the result:

$$\begin{aligned} E(y_t^{2*}) &\le (\sigma_u^2 + E(y_0^{2*}))(1-\rho^2)^{-1} < \infty \,, \\ Cov(y_t^*, y_{t+k}^*) &\le \rho^k E(y_t^{2*}) \quad for \quad k \ge 1 \,. \end{aligned}$$

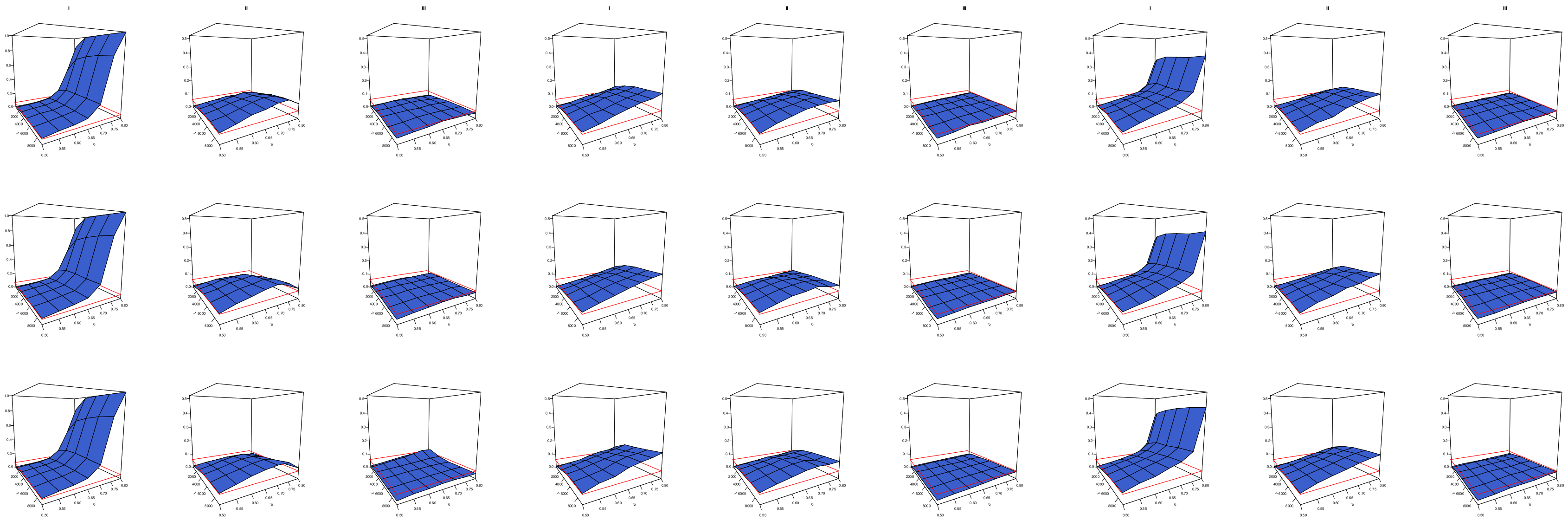
Since $\rho \in (0,1)$ and under Assumption 4.2, we have:

$$\begin{aligned} \sum_{k=0}^{\infty} Cov(y_t^*, y_{t+k}^*) &\le \frac{(1-\rho^2)^{-1}(\sigma_u^2 + E(y_0^{2*}))}{1-\rho} < \infty \,, \\ Var(y_t^*) + 2\sum_{k=1}^{\infty} Cov(y_t^*, y_{t+k}^*) &\le \frac{(1+\rho)(\sigma_u^2 + E(y_0^{2*}))}{(1-\rho^2)(1-\rho)} < \infty \,. \end{aligned} \tag{4.23}$$

It can easily be seen that the left-hand side of (4.23) is the long-run variance of $\{y_t^*\}$ which is bounded. Therefore, a Fourier transformation of the acf of $\{y_t^*\}$ will converge to a pseudo spectral density $f_{y_t^*}(\cdot)$ due to Parseval's theorem for $T \to \infty$:

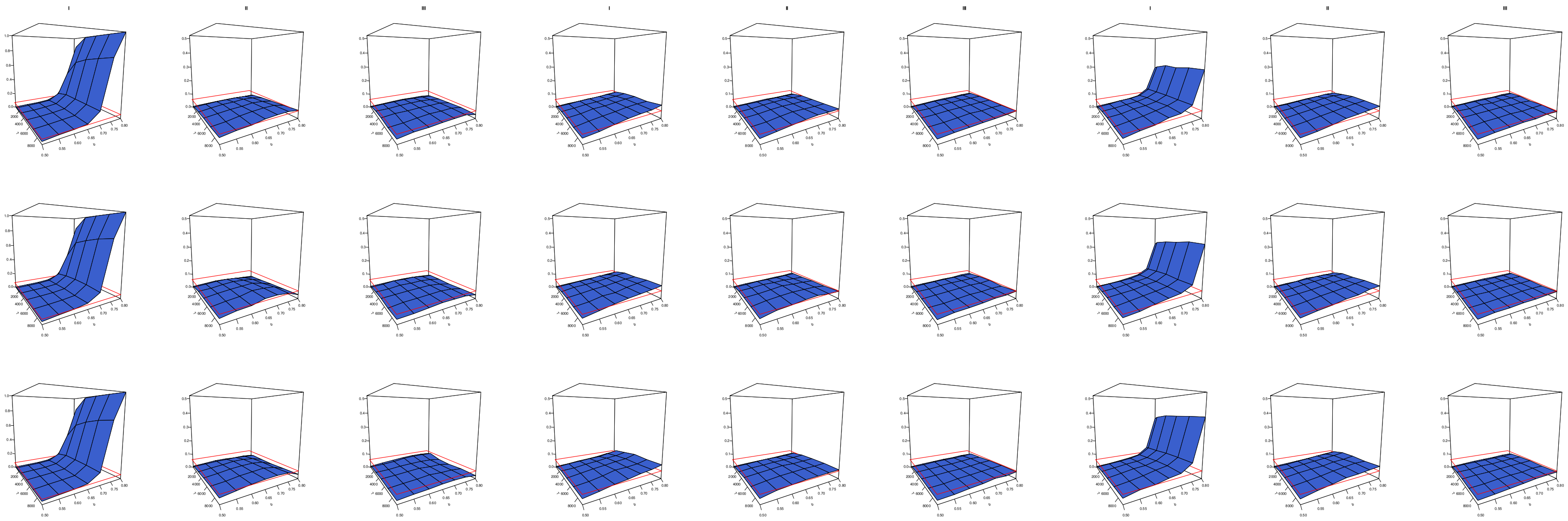
$$f_{y_t^*}(\lambda) = \underbrace{\frac{1}{2\pi}[Var(y_t^*) + 2\sum_{k=1}^{\infty} Cov(y_t^*, y_{t+k}^*)\cos(\lambda k)]}_{O_p(1)} < \frac{(1+\rho)(\sigma_u^2 + E(y_0^{2*}))}{(1-\rho^2)(1-\rho)} < \infty \,,$$

for $\lambda \in [-\pi, \pi]$ which establishes the result. □



(a) $\phi = 0.5$ and $\tau = 0.5$ **(b)** $\phi = 0$ and $\tau = 0.5$ **(c)** $\phi = -0.5$ and $\tau = 0.5$

Figure 4.6: The 3D plots report the rejection frequency for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory, using trimming parameter $\epsilon = 0.02$, dependent on the sample size T and frequency bandwidth b under DGP1 (4.21) with parameters ϕ and τ as given in (a)-(c), $\nu = 1/3$ and $d \in \{0, 0.2, 0.4\}$ by row 1-3, respectively. The columns report results if no prewhitening procedure is applied (I), if the prewhitening procedure of Qu (2011) is considered (II) and in case our extended prewhitening procedure is used (III).

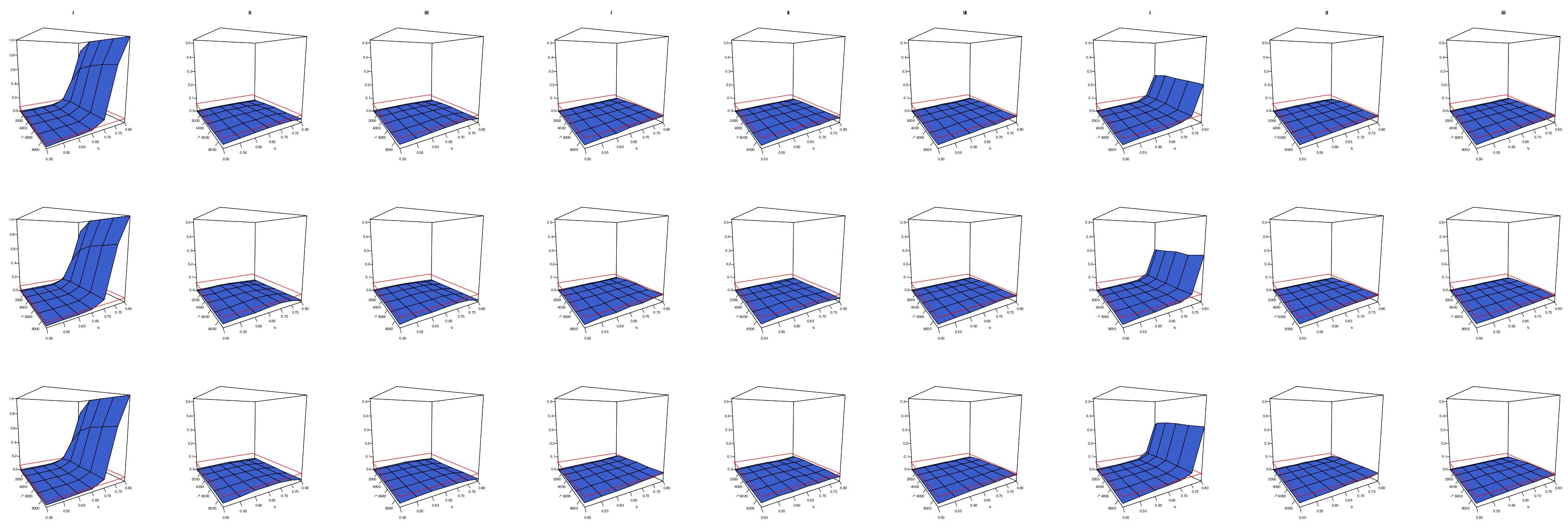


(a) $\phi = 0.5$ and $\tau = 0.75$

(b) $\phi = 0$ and $\tau = 0.75$

(c) $\phi = -0.5$ and $\tau = 0.75$

Figure 4.7: The 3D plots report the rejection frequency for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory, using trimming parameter $\epsilon = 0.02$, dependent on the sample size T and frequency bandwidth b under DGP1 (4.21) with parameters ϕ and τ as given in (a)-(c), $\nu = 1/3$ and $d \in \{0, 0.2, 0.4\}$ by row 1-3, respectively. The columns report results if no prewhitening procedure is applied (I), if the prewhitening procedure of Qu (2011) is considered (II) and in case our extended prewhitening procedure is used (III).



(a) $\phi = 0.5$ **(b)** $\phi = 0$ **(c)** $\phi = -0.5$

Figure 4.8: The 3D plots report the rejection frequency for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory, using trimming parameter $\epsilon = 0.02$, dependent on the sample size T and frequency bandwidth b under DGP1 (4.21) with parameters ϕ as given in (a)-(c), $\nu = 1$ and $d \in \{0, 0.2, 0.4\}$ by row 1-3, respectively. The columns report results if no prewhitening procedure is applied (I), if the prewhitening procedure of Qu (2011) is considered (II) and in case our extended prewhitening procedure is used (III).

T/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
τ	0.25		0.5		0.75		0.25		0.5		0.75		0.25		0.5		0.75	
ϕ	0						0.5						-0.5					
d=0																		
500	2.37	2.84	2.23	2.78	2.29	3.01	1.43	2.04	1.65	2.22	1.37	1.76	2.02	2.75	2.60	3.08	2.40	2.89
1000	2.72	3.43	3.40	3.71	2.94	3.49	2.07	2.80	2.54	3.17	2.22	2.64	2.90	3.59	3.28	3.64	2.78	3.26
3000	3.31	3.76	4.48	4.59	4.05	3.94	3.19	3.82	3.58	3.70	3.35	3.53	3.56	3.97	4.18	4.35	3.64	3.75
5000	3.80	4.31	4.41	4.57	3.60	4.00	3.51	4.10	4.23	4.34	4.00	4.37	3.52	3.90	4.25	4.70	4.05	4.19
7000	3.67	4.18	4.61	4.59	4.21	4.24	3.80	4.17	4.19	4.39	3.93	3.89	4.23	4.28	4.43	4.29	4.30	4.35
9000	4.17	4.44	4.42	4.60	4.74	4.66	3.94	4.26	3.78	4.03	4.11	4.42	4.29	4.64	4.48	4.67	3.88	4.49
d=0.2																		
500	1.90	2.28	2.30	2.67	2.07	2.66	2.77	2.80	2.53	2.54	1.81	1.78	1.83	2.57	2.46	2.91	1.91	2.50
1000	2.49	3.34	3.04	3.56	2.70	3.34	4.03	4.33	3.65	4.02	2.65	2.78	2.51	3.40	2.69	3.40	2.94	3.68
3000	3.39	3.98	3.95	4.07	3.43	3.55	4.41	4.50	4.27	4.07	3.58	3.64	3.36	3.89	3.93	4.19	3.40	3.91
5000	4.25	4.50	4.02	4.34	3.94	4.48	3.98	4.79	4.12	4.26	3.68	4.05	3.32	3.71	4.13	4.47	3.55	4.11
7000	3.84	4.04	4.12	4.21	4.08	4.52	4.12	4.41	3.85	4.12	4.18	4.25	3.80	4.26	4.08	4.26	3.94	4.00
9000	3.98	4.41	3.98	4.47	4.16	4.55	3.92	4.39	4.13	4.19	4.08	4.27	3.93	4.24	4.07	4.12	4.03	4.50
d=0.4																		
500	1.64	2.49	2.25	2.76	1.92	2.48	5.84	5.75	4.94	4.84	3.68	3.66	1.79	2.53	2.12	2.78	2.01	2.55
1000	2.19	2.86	2.17	3.11	2.29	3.20	7.86	8.48	5.84	6.23	3.99	4.43	2.26	3.31	2.47	3.36	2.12	3.15
3000	3.00	3.69	3.28	3.88	3.16	3.86	5.16	5.21	3.60	3.84	3.40	3.34	3.09	3.82	3.03	3.59	3.27	3.71
5000	3.21	3.91	3.56	4.02	3.42	4.13	3.61	4.19	3.38	4.14	3.18	3.60	3.15	3.88	3.51	4.10	3.56	4.25
7000	3.37	3.94	3.41	3.78	3.87	4.19	3.60	4.31	3.46	3.92	3.53	3.79	3.46	3.93	3.79	4.44	3.36	3.76
9000	3.70	4.32	4.03	4.66	3.61	3.89	3.75	4.36	3.34	3.80	3.44	3.86	3.81	4.18	3.80	4.13	3.72	4.30

Table 4.5: Reported are the rejection frequencies for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory using trimming parameter $\epsilon = \{0.02, 0.05\}$ and a frequency bandwidth choice of $b = 0.7$, in case our extended prewhitening procedure is used with two possible breakpoints as the maximum value in the algorithm of Bai and Perron (1998, 2003a). DGP1 (4.21) is generated for various sample sizes T and with parameters $\phi \in \{-0.5, 0, 0.5\}$, $\tau \in \{0.25, 0.5, 0.75\}$, $\nu = 1/3$ and $d \in \{0, 0.2, 0.4\}$.

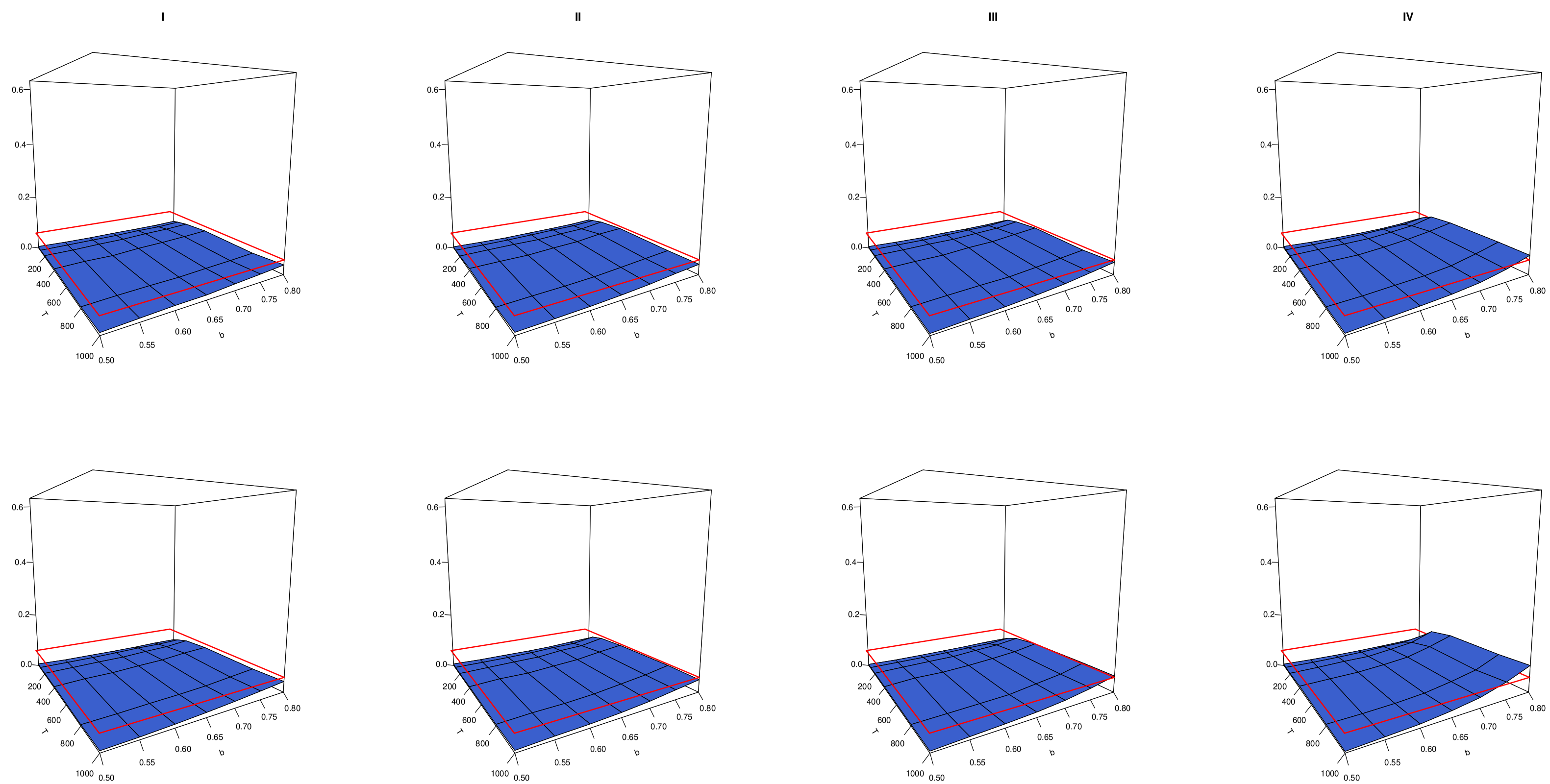


Figure 4.9: The 3D plots report the rejection frequency for $\alpha = 5\%$ of the Qu (2011) test on spurious long memory, using trimming parameter $\epsilon = 0.05$ ($\epsilon = 0.02$) for $T < 500$ ($T > 500$) dependent on the sample size T and frequency bandwidth b under DGP2 (4.22). The columns report results if $\{v_t\}$ displays the following short-run dynamics: $\phi = 0$ and $d = 0$ (I), $\phi = 0.8$ and $d = 0$ (II), $\phi = 0$ and $d = 0.2$ (III), $\phi = 0$ and $d = 0.4$ (IV). The first (second) row gives results if the residual series $\{\hat{e}_t\}$ is applied, constructed with parameter estimates using the flat (Epanechnikov) kernel from (4.19) with MSE-optimal bandwidth choice $H = \sqrt{T}$.

The Algorithm of Bai and Perron (1998, 2003a)

Bai and Perron (1998, 2003a) propose a testing procedure for multiple structural changes in a time series that occur at unknown points in time and prove its consistency if the true number of breakpoints is smaller or equal to a pre-specified upper bound of $\#\tau$ breakpoints. It is based on sequentially testing the null hypothesis of l versus $l+1$ structural breaks in the coefficients of a linear regression model and proceeds as follows:

1. A model with l break points ($l = 0$ in the initialization) is estimated which is done by first separating the sample into l partitions $(T_1, ..., T_l)$ where for each regime the regression parameters are estimated by minimizing the sum of squared residuals. Since we are interested in mean shifts in the residual series $\{\hat{\varpi}_t^2\}$, we get the corresponding estimator $\widehat{\mu}(T_j)$ for each segment $j = 1, ..., l$. The obtained $\widehat{\mu}(T_j)$ are substituted into the objective function such that the total resulting sum of squared residuals is denoted as $S_T(T_1, ..., T_l)$. The l estimated break points are then acquired via:

$$(\hat{T}_1, ..., \hat{T}_l) = \arg\min_{T_1,...,T_l} S_T(T_1, ..., T_l)\,,$$

which is a global minimization of the sum of squared residuals.

2. Denote with $(\hat{T}_1, ..., \hat{T}_l)$ all breakpoints obtained in the previous step. One then sequentially tests the hypothesis H_0: No additional break against the alternative of a single shift for each of the $(l+1)$ regimes. Among all possible break points obtained, the additional break point taken into account in the $(l+1)$ model is the one that minimizes the sum of squared residuals over all segments. In a next step one tests whether the sum of squared residuals of the $(l+1)$ model is significantly smaller than the sum of squared residuals of the l model by calculating the statistic:

$$F_T(l+1|l) = \left\{S_T(\widehat{T}_1, ..., \widehat{T}_l) - \min_{1\leq i\leq l+1} \inf_{\tilde{\tau}\in\Lambda_{i,\eta}} S_T(\widehat{T}_1, ..., \widehat{T}_{i-1}, \tilde{\tau}, \widehat{T}_i, ..., \widehat{T}_l)\right\}/\widehat{\sigma}^2\,,$$

with $\widehat{\sigma}^2$ a consistent estimator for σ^2 under the null hypothesis, $\tilde{\tau}$ as the additional break date and where for some small number η:

$$\Lambda_{i,\eta} = \left\{\tilde{\tau}; \widehat{T}_{i-1} + (\widehat{T}_i - \widehat{T}_{i-1})\eta \leq \tilde{\tau} \leq \widehat{T}_i - (\widehat{T}_i - \widehat{T}_{i-1})\eta\right\},$$

holds which ensures that the break dates are bounded away from the boundaries. The test statistic converges to a scaled Brownian bridge and critical values for various η and l are provided by Bai and Perron (1998, 2003a,b).

3. If one rejects the null using the $F_T(l+1|l)$ test, the l break points in the first step increases by one and the algorithm is subsequently repeated. The sequential procedure stops when $F_T(l+1|l)$ in the second step is smaller than its corresponding critical value i.e. the sum of squared residuals does not decrease significantly taking one additional break into account or if the pre-specified upper bound $\#\tau$ is reached which ultimately results in a number of $0 \leq l^* \leq \#\tau$ breaks being selected.

Bibliography

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.

Agiakloglou, C. and Newbold, P. (1994). Lagrange multiplier tests for fractional difference. *Journal of Time Series Analysis*, 15(3):253–262.

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.

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

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

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. and Sun, Y. (2004). Adaptive local polynomial whittle estimation of long-range dependence. *Econometrica*, 72(2):569–614.

Baek, C., Fortuna, N., and Pipiras, V. (2014). Can markov switching model generate long memory? *Economics Letters*, 124(1):117–121.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Bai, J. and Perron, P. (2003a). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Bai, J. and Perron, P. (2003b). Critical values for multiple structural change tests. *The Econometrics Journal*, 6(1):72–78.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1):5–59.

Baillie, R. T. and Bollerslev, T. (1994). The long memory of the forward premium. *Journal of International Money and Finance*, 13(5):565–571.

Baillie, R. T., Chung, C.-F., and Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated arfima-garch model. *Journal of Applied Econometrics*, 11(1):23–40.

Baillie, R. T. and Kapetanios, G. (2007). Testing for neglected nonlinearity in long-memory models. *Journal of Business & Economic Statistics*, 25(4):447–461.

Bandi, F. M. and Perron, B. (2006). Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics*, 4(4):636–670.

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.

Benati, L. and Surico, P. (2008). Evolving U.S. monetary policy and the decline of inflation predictability. *Journal of the European Economic Association*, 6(2-3):634–646.

Benati, L. and Surico, P. (2009). VAR analysis and the Great Moderation. *The American Economic Review*, 99(4):1636–1652.

Beran, J. (2009). On parameter estimation for locally stationary long-memory processes. *Journal of Statistical Planning and Inference*, 139(3):900–915.

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

Berenguer-Rico, V. and Gonzalo, J. (2014). Summability of stochastic processes - A generalization of integration for non-linear processes. *Journal of Econometrics*, 178(Part 2):331–341.

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.

Bos, C. S., Franses, P. H., and Ooms, M. (1999). Long memory and level shifts: Re-analyzing inflation rates. *Empirical Economics*, 24(3):427–449.

Box, G. E. and Jenkins, G. M. (1976). *Time series analysis: forecasting and control, revised ed.* Holden-Day.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Breitung, J. and Hassler, U. (2002). Inference on the cointegration rank in fractionally integrated processes. *Journal of Econometrics*, 110(02):167–185.

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.

Cavaliere, G., Angelis, L. D., Rahbek, A., and Robert Taylor, A. (2015a). A comparison of sequential and information-based methods for determining the co-integration rank in heteroskedastic VAR models. *Oxford Bulletin of Economics and Statistics*, 77(1):106–128.

Cavaliere, G., Nielsen, M. Ø., and Taylor, A. M. R. (2015b). Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price dynamics in commodity spot and futures markets. *Journal of Econometrics*, 187(02):557–579.

Cavaliere, G., Nielsen, M. Ø., and Taylor, A. R. (2017). Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form. *Journal of Econometrics*, 198(1):165–188.

Cavaliere, G. and Taylor, A. R. (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics*, 140(2):919–947.

Cavaliere, G. and Taylor, A. R. (2008). Testing for a change in persistence in the presence of non-stationary volatility. *Journal of Econometrics*, 147(1):84–98.

Cavaliere, G. and Taylor, A. R. (2009). Heteroskedastic time series with a unit root. *Econometric Theory*, 25(5):1228–1276.

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.

Cheung, Y.-W. (1993). Long memory in foreign-exchange rates. *Journal of Business & Economic Statistics*, 11(1):93–101.

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.

Cogley, T., Primiceri, G. E., and Sargent, T. J. (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, 2(1):43–69.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

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.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics*, 28(6):1762–1794.

Demetrescu, M. and Hanck, C. (2012a). A simple nonstationary-volatility robust panel unit root test. *Economics Letters*, 117(1):10–13.

Demetrescu, M. and Hanck, C. (2012b). Unit root testing in heteroskedastic panels using the Cauchy estimator. *Journal of Business & Economic Statistics*, 30(2):256–264.

Demetrescu, M., Kuzin, V., and Hassler, U. (2008). Long memory testing in the time domain. *Econometric Theory*, 24(01):176–215.

Demetrescu, M. and Sibbertsen, P. (2014). Inference on the long-memory properties of time series with non-stationary volatility. Technical report, Leibniz Universität Hannover, Wirtschaftswissenschaftliche Fakultät.

Demetrescu, M. and Sibbertsen, P. (2016). Inference on the long-memory properties of time series with non-stationary volatility. *Economics Letters*, 144:80–84.

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.

Diebold, F. X. and Rudebusch, G. D. (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics*, 24(2):189–209.

Dittmann, I. and Granger, C. W. (2002). Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics*, 110(2):113–133.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. (1990). Stock Volatility and the Crash of '87: Discussion. *Review of Financial Studies*, 3(1):103–106.

Fitzsimmons, P. and McElroy, T. (2010). On joint fourier–laplace transforms. *Communications in Statistics – Theory and Methods*, 39(10):1883–1885.

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.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Giraitis, L., Kapetanios, G., and Yates, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, 179(1):46–65.

Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large sample inference for long memory processes*. London: Imperial College Press.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 123(1):89–120.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11(3):399–421.

Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.

Harris, D. and Kew, H. (2017). Adaptive Long Memory Testing under Heteroskedasticity. *Econometric Theory*, 33(3):755–778.

Harris, D., McCabe, B., and Leybourne, S. (2008). Testing for long memory. *Econometric Theory*, 24(01):143–175.

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. and Meller, B. (2014). Detecting multiple breaks in long memory the case of US inflation. *Empirical Economics*, 46(2):653–680.

Hassler, U., Rodrigues, P. M., and Rubia, A. (2009). Testing for general fractional integration in the time domain. *Econometric Theory*, 25(06):1793–1828.

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, 13(1):37–45.

Hosking, J. R. (1981). Fractional differencing. *Biometrika*, 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.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–808.

Kapetanios, G. and Yates, T. (2014). Evolving UK and US macroeconomic dynamics through the lens of a model of deterministic structural change. *Empirical Economics*, 47(1):305.

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.

Kew, H. and Harris, D. (2009). Heteroskedasticity-robust testing for a fractional unit root. *Econometric Theory*, 25(6):1734–1753.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(6):1130–1164.

Kim, T.-H., Leybourne, S., and Newbold, P. (2002). Unit root tests with a break in innovation variance. *Journal of Econometrics*, 109(2):365–387.

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, Discussion Paper, Wirtschaftswissenschaftliche Fakultät, Leibniz University of Hannover.

Kumar, M. S. and Okimoto, T. (2007). Dynamics of persistence in international inflation rates. *Journal of Money, Credit and Banking*, 39(6):1457–1479.

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

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.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21(01):21–59.

Leschinski, C. (2017). On the memory of products of long range dependent time series. *Economics Letters*, 153:72–76.

Li, J. and Patton, A. J. (2015). Asymptotic inference about predictive accuracy using high frequency data. *unpublished*.

Lobato, I. N. and Robinson, P. M. (1998). A nonparametric test for I(0). *The Review of Economic Studies*, 65(3):475–495.

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.

Loretan, M. and Phillips, P. C. (1994). Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of Empirical Finance*, 1(2):211–248.

Luo, G. Y. (1998). Market efficiency and natural selection in a commodity futures market. *Review of Financial Studies*, 11(3):647–674.

Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437.

Mariano, R. S. and Preve, D. (2012). Statistical tests for multiple forecast comparison. *Journal of Econometrics*, 169(1):123–130.

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(6):1196–1237.

McConnell, M. M. and Perez-Quiros, G. (2000). Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review*, 90(5):1464–1476.

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.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

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. (2005). Multivariate lagrange multiplier tests for fractional integration. *Journal of Financial Econometrics*, 3(3):372–398.

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

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.

Palma, W. and Olea, R. (2010). An efficient estimator for locally stationary gaussian long-memory processes. *The Annals of Statistics*, 38(5):2958–2997.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

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. 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.*

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

Rao, C. R. (1973). *Linear statistical inference and its applications*, volume 2. John Wiley & Sons.

Rinke, S., Busch, M., and Leschinski, C. (2017). Long memory, breaks, and trends: On the sources of persistence in inflation rates. Technical report, Leibniz Universität Hannover, Wirtschaftswissenschaftliche Fakultät.

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47(1):67–84.

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(3):1048–1072.

Robinson, P. M. (2005). Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory*, 21(1):171–180.

Rossi, B. (2005). Testing long-horizon predictive ability with high persistence, and the Meese–Rogoff puzzle. *International Economic Review*, 46(1):61–92.

Schwert, G. W. (2002). Tests for unit roots: A Monte Carlo investigation. *Journal of Business & Economic Statistics*, 20(1):5–17.

Sensier, M. and Van Dijk, D. (2004). Testing for volatility changes in US macroeconomic time series. *Review of Economics and Statistics*, 86(3):833–839.

Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Technical Report 1101, Queen's Economics Dept. Working Paper.

Shorack, G. R. and Wellner, J. A. (2009). *Empirical Processes with Applications to Statistics*, volume 59. SIAM.

Sin, C.-Y. and White, H. (1996). Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics*, 71(1):207–225.

Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *Review of Economics and Statistics*, 87(3):503–522.

Sun, Y., Phillips, P. C., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, 76(1):175–194.

Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory*, 15(04):549–582.

Varneskov, R. T. and Perron, P. (2017). Combining long memory and level shifts in modelling and forecasting the volatility of asset returns. *Quantitative Finance*, pages 1–23.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084.

Westerlund, J. and Narayan, P. (2013). Testing the efficient market hypothesis in conditionally heteroskedastic futures markets. *Journal of Futures Markets*, 33(11):1024–1045.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.

Yu, W.-C. (2009). Markov switching and long memory: A monte carlo analysis. *Applied Economics Letters*, 16(12):1205–1210.