

Kreye, Tom Jannik; Sibbertsen, Philipp

**Working Paper**

## Testing for a forecast accuracy breakdown under long memory

Hannover Economic Papers (HEP), No. 729

**Provided in Cooperation with:**

School of Economics and Management, University of Hannover

*Suggested Citation:* Kreye, Tom Jannik; Sibbertsen, Philipp (2024) : Testing for a forecast accuracy breakdown under long memory, Hannover Economic Papers (HEP), No. 729, Leibniz Universität Hannover, Wirtschaftswissenschaftliche Fakultät, Hannover

This Version is available at:

<https://hdl.handle.net/10419/307743>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Testing for a Forecast Accuracy Breakdown under Long Memory

Tom Jannik Kreye\*

Institute of Statistics, Leibniz University Hannover  
kreye@statistik.uni-hannover.de

and

Philipp Sibbertsen

Institute of Statistics, Leibniz University Hannover  
sibbertsen@statistik.uni-hannover.de

November 7, 2024

## Abstract

We propose a test to detect a forecast accuracy breakdown in a long memory time series and provide theoretical and simulation evidence on the memory transfer from the time series to the forecast residuals. The proposed method uses a double sup-Wald test against the alternative of a structural break in the mean of an out-of-sample loss series. To address the problem of estimating the long-run variance under long memory, a robust estimator is applied. The corresponding breakpoint results from a long memory robust CUSUM test. The finite sample size and power properties of the test are derived in a Monte Carlo simulation. A monotonic power function is obtained for the fixed forecasting scheme. In our practical application, we find that the global energy crisis that began in 2021 led to a forecast break in European electricity prices, while the results for the U.S. are mixed.

*Keywords:* Electricity prices; Forecast failure; Local power; Out-of-sample forecast; Structural change

---

\*The authors gratefully acknowledge support from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummern 258395632 and INST 187/742-1 FUGG. We would like to thank the participants of the Statistical Week 2023 in Dortmund and the participants of the Annual Conference of the International Association for Applied Econometrics 2024 in Thessaloniki.

# 1 Introduction

The presence of low frequency contaminations, such as structural breaks, poses a challenge in distinguishing true long memory from spurious long memory. This is due to the fact that both concepts share similar characteristics, such as a singularity in the periodogram near the zero Fourier frequency and significant autocorrelations at large lags ([Diebold and Inoue \(2001\)](#), [Granger and Hyung \(2004\)](#), [Mikosch and Stărică \(2004\)](#)). Tests for structural breaks are closely related to tests for forecast breakdowns, which aim to detect changes in forecasts based on a given loss function. [Perron \(2006\)](#) reviews the extensive literature on structural breaks. [Perron and Yamamoto \(2021\)](#) extend classical tests for structural breaks to the context of forecast failure. [Giacomini and Rossi \(2009\)](#) propose a test to retrospectively assess whether a given forecast model provides stable forecasts by comparing in-sample and out-of-sample averages of a forecast error loss series. A forecast breakdown is defined as a decline in the out-of-sample performance of a forecasting model relative to its in-sample performance. To measure the forecast performance, the popular squared error loss function is used in this paper. Unlike structural break tests applied to the forecasting model, forecast accuracy breakdown tests allow for misspecified models and they can detect variance changes when a quadratic loss function is used. Furthermore, they assist the applied econometrician in determining the accuracy of a given forecast model and whether it requires modification. It is important to note, however, that the detection of a forecast breakdown does not necessarily indicate that the forecast model needs to be altered. For instance, [Christoffersen and Diebold \(1997\)](#) demonstrate that the optimality of a point forecast does not depend on the variance of the errors when a symmetric loss function is used. Nevertheless, a change in the variance can cause a forecast breakdown, since tests for forecast failures cannot distinguish between symmetric and asymmetric loss functions.

In practical applications, it is natural to distinguish between in-sample and out-of-sample periods. In retrospective testing, however, an artificial separation between these periods is required, since the test period serves as a pseudo-out-of-sample period. The forecasting model is estimated in the in-sample period, and to assess whether the forecast accuracy changes, it should provide stable forecasts. To obtain forecasts, [Perron and Yamamoto \(2021\)](#) recommend using a fixed forecast scheme, since it is the only forecast scheme that leads to a monotonic power function, mainly because it ensures the maximum difference between the in-sample and out-of-sample means of the loss series. Rolling and recursive schemes can induce power losses due to contamination problems. Under a fixed scheme, the test proposed by [Giacomini and Rossi \(2009\)](#) is a Wald test for a single mean change in the total loss series at a predetermined date. The total loss series combines in-sample and out-of-sample losses. The test suffers from power losses if the break date is not set correctly. To address this issue, [Rossi and Inoue \(2012\)](#) propose a sup-Wald test that maximizes the [Giacomini and Rossi \(2009\)](#) test over a predefined range of potential break dates. When multiple changes occur, the test still suffers from non-monotonic power functions. For this reason, [Perron and Yamamoto \(2021\)](#) propose a double sup-Wald (DSW) test for a structural break in the mean of the out-of-sample loss series, estimating the breakpoint with the [Bai and Perron \(1998\)](#) estimator. The proposed method involves performing SW tests for each choice of the in-sample period. Ultimately, the DSW test statistic value is the largest value among all in-sample period choices. In this paper, we extend the DSW test to the long memory framework by using a memory and autocorrelation consistent (MAC) estimator of the long-run variance proposed by [Robinson \(1995\)](#) and by incorporating a long memory robust estimator of the breakpoint. A related topic is discussed by [Paza Mboya and Sibbertsen \(2023\)](#). They obtain optimal forecasts in a long memory time series setting under discrete structural breaks. To the best of our knowledge, our study is the first to connect long memory time series with forecast failures.

Kruse et al. (2018) derive theoretical evidence for a long memory transfer from a time series to forecasts and to forecast differentials, especially in the presence of biased forecasts. We show that this transfer also applies to the squared forecast error, which is a component of a forecast differential. We consider biased and unbiased forecasts, as well as the presence and absence of a fractional cointegration relationship between a long memory time series and its forecast.

The remainder of the paper is organized as follows. Section 2 contains the theoretical derivation of the long memory transfer from the time series to the forecasts and to the squared forecast error. This justifies the need for a long memory robust forecast accuracy breakdown test, which is presented in Section 3. To derive the finite sample size and power properties of the test, we perform a Monte Carlo study in Section 4. Section 5 applies the proposed test to European and U.S. electricity prices, while Section 6 contains concluding remarks. All proofs are consolidated in the Appendix.

## 2 Long Memory in Forecast Residuals

Before presenting the test to detect a change in forecast accuracy, we analyze the transfer of long memory from a time series and its forecast to the loss function under specific settings. We distinguish between biased and unbiased forecasts. Our analysis follows Kruse et al. (2018), who derive long memory transfer properties for forecast error loss differentials. The proofs of all propositions are provided in the Appendix. To create a framework, we restrict the time series and its forecast to be stationary long memory processes.

**Assumption 1.** *The time series  $y_t$  and the forecast  $\hat{y}_t$ , with  $t = 1, \dots, T$  and expectations  $\mathbb{E}(y_t) = \mu_y$  and  $\mathbb{E}(\hat{y}_t) = \mu_{\hat{y}}$ , are causal Gaussian long memory processes having the spectral density  $f_a(\lambda) \sim L_f(\lambda)|\lambda|^{-2d_a}$ , for  $a_t \in \{y_t, \hat{y}_t\}$ , as  $\lambda \rightarrow 0$ , where  $L_f(\lambda) \geq 0$  is a symmetric function that varies slowly at the origin. We write  $a_t \sim LM(d_a)$ , where  $d_a \in [0; 1/2)$ .*

Assumption 1 is convenient since Dittmann and Granger (2002) derive long memory properties for squares and cross products of Gaussian long memory processes. We now distinguish between the presence of fractional cointegration between the time series and its forecast and the absence of this property. A fractional cointegration relationship refers to the existence of a linear combination of long memory time series that is integrated of reduced order. The memory transfer results are based on the asymptotic behavior of autocovariance functions of products and squares of long memory time series. In addition, Proposition 3 of Chambers (1998) can be used, as it states that the memory order of a linear combination of fractionally integrated series is equal to the maximum order of the individual components. Leschinski (2017) generalizes this result for the broader class of long memory processes and discusses the memory of products of long memory time series. We derive the long memory transfer for the squared error loss function, which we also use for our test in Section 3,

$$L_t = (y_t - \hat{y}_t)^2. \quad (1)$$

First, consider the case of the absence of fractional cointegration.

**Assumption 2.** *If  $y_t, \hat{y}_t \sim LM(d)$ , then  $y_t - \phi_0 - \phi_1 \hat{y}_t \sim LM(d) \forall \phi_0, \phi_1 \in \mathbb{R}$ .*

Under Assumption 2, the long memory transfer is summarized in the following proposition.

**Proposition 1.** *Under Assumptions 1 and 2, the squared forecast error in Eq. (1) is  $L_t \sim LM(d_L)$ , where*

$$d_L = \begin{cases} \max\{d_y, d_{\hat{y}}\}, & \text{if } \mu_y \neq \mu_{\hat{y}} \\ \max\{2\max\{d_y, d_{\hat{y}}\} - 1/2, 0\}, & \text{if } \mu_y = \mu_{\hat{y}}. \end{cases}$$

**Proof.** *See the Appendix.*

The result is obtained by replacing  $y_t$  and  $\hat{y}_t$  by their centered series  $a_t^* = a_t - \mu_a$  for

$a_t \in \{y_t, \hat{y}_t\}$ . As shown in the Appendix, the squared forecast error in Eq. (1) can then be rearranged as

$$L_t = 2[y_t^*(\mu_y - \mu_{\hat{y}}) - \hat{y}_t^*(\mu_y - \mu_{\hat{y}})] - 2y_t^* \hat{y}_t^* + y_t^{*2} + \hat{y}_t^{*2} + \text{const.} \quad (2)$$

Proposition 1 shows that if the forecast is biased, the memory order of the original terms dominates. If the forecast is unbiased, the first part of Eq. (2) drops and the product and squared series remain. As shown in the Appendix, the memory orders of the squared series dominate the memory of the product series and the result follows.

We now relax Assumption 2 and allow for fractional cointegration between the series and the forecast.

**Assumption 3.** *Let  $x_t$  be a stationary long memory process according to Assumption 1 with memory parameter  $d_x$ . Then, if  $y_t$  and  $\hat{y}_t$  are  $LM(d_x)$  and fractionally cointegrated, they can be represented as  $y_t = \beta_y + \kappa_y x_t + \epsilon_{y,t}$  and  $\hat{y}_t = \beta_{\hat{y}} + \kappa_{\hat{y}} x_t + \epsilon_{\hat{y},t}$  with  $\kappa_y, \kappa_{\hat{y}} \neq 0$  and  $\epsilon_{y,t}, \epsilon_{\hat{y},t}$  are centered causal Gaussian long memory processes with  $1/2 > d_x > d_{\epsilon_{y,t}}, d_{\epsilon_{\hat{y},t}}$ . We write  $y_t, \hat{y}_t \sim FCI(d_x, d_x - \delta)$ .*

$d_x - \delta$  refers to the memory order of the linear combination of  $y_t$  and  $\hat{y}_t$ , and they are fractionally cointegrated for  $\delta > 0$ . As can be seen from Assumption 3,  $x_t$  is the common factor that drives the long memory properties of  $y_t$  and  $\hat{y}_t$ . We restrict the fractional cointegration to a form where the two series can be represented as linear functions of their common factor. In fact, our previous Assumption 2 of the absence of fractional cointegration may be inappropriate in a forecasting setup. We nevertheless retain the case of the absence of fractional cointegration in Proposition 1 for completeness. It follows directly from Assumption 3 that

$$y_t - \hat{y}_t = \beta_y - \beta_{\hat{y}} + x_t(\kappa_y - \kappa_{\hat{y}}) + \epsilon_{y,t} - \epsilon_{\hat{y},t}.$$

There are three cases to consider under fractional cointegration.

- (i) Biased forecast and  $\kappa_y \neq \kappa_{\hat{y}}$ ,
- (ii) biased forecast and  $\kappa_y = \kappa_{\hat{y}}$ ,
- (iii) unbiased forecast and  $\kappa_y = \kappa_{\hat{y}}$ .

The memory transfer results for the three cases are obtained by substituting the relations from Assumption 3 into the squared forecast error in Eq. (1). By rearranging the resulting equation, we obtain an expression similar to that in Eq. (2). We obtain the following result.

**Proposition 2.** *Under Assumptions 1 and 3, the squared forecast error in Eq. (1) is  $L_t \sim LM(d_L)$ , where*

$$d_L = \begin{cases} d_x, & \text{if } \mu_y \neq \mu_{\hat{y}} \text{ and } \kappa_y \neq \kappa_{\hat{y}} \\ \tilde{d}, & \text{if } \mu_y \neq \mu_{\hat{y}} \text{ and } \kappa_y = \kappa_{\hat{y}} \\ \max\{2d_x - 1/2, 0\}, & \text{if } \mu_y = \mu_{\hat{y}} \text{ and } \kappa_y = \kappa_{\hat{y}} \end{cases}$$

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

**Proof.** *See the Appendix.*

In case of a biased forecast with  $\kappa_y \neq \kappa_{\hat{y}}$ , the memory order of  $x_t$  dominates. In the other two cases, the memory is reduced. Under a biased forecast and  $\kappa_y = \kappa_{\hat{y}}$ , the resulting memory parameter of the squared forecast error is  $\tilde{d}$ . The exact order of  $\tilde{d}$  cannot be determined because squares and products of the errors  $\epsilon_{y,t}$  and  $\epsilon_{\hat{y},t}$  are involved, for which the memory orders are unknown. In the remaining case of an unbiased forecast with  $\kappa_y = \kappa_{\hat{y}}$ , the forecast is an accurate prediction of the underlying series. Here, the memory of  $x_t^{*2}$  dominates, but this memory is also smaller than that of  $x_t^*$ , as shown in the Appendix. In summary, our main results in Propositions 1 and 2 show that long memory can be transferred from the underlying time series to the forecast and finally, to the squared forecast error. The memory transfer depends crucially on the (un)biasedness of the forecast and the presence of fractional cointegration.



### 3 Double Sup-Wald Test

After deriving the long memory transfer from the time series to the forecast and finally to the squared forecast error, we now turn to the detection of a forecast failure. The accuracy of the point forecasts is evaluated in the out-of-sample period. We use the popular out-of-sample squared error loss function

$$L_{t+\tau}^o(m) = (y_{t+\tau} - \hat{y}_{t+\tau})^2, \quad (3)$$

where  $L^o(m) = L_{m+\tau}^o, \dots, L_T^o$  is the out-of-sample loss sequence of size  $n = T - m - \tau + 1$ .

Under the null hypothesis,

$$H_0 : E[L_t^o] = \mu_0, \quad \forall t = \tau + 1, \dots, T - \tau + 1,$$

there is no structural break in the mean of the out-of-sample loss series. The alternative hypothesis is given by:

$$H_1 : E[L_t^o] \neq E[L_{t+1}^o]$$

for at least one  $t = \tau + 1, \dots, T - \tau$ . Thus, our test is capable of detecting a single break in the forecast performance, but it does not rule out the existence of multiple breaks. We do not take into account the presence of multiple breaks because there is no long memory robust breakpoint estimator yet.

The underlying test for a break in the predictive accuracy is a Wald-type test proposed by [Perron and Yamamoto \(2021\)](#). To compute the test statistic, the out-of-sample loss series is calculated for each in-sample period of size  $m$  in an interval  $[m_0, m_1]$ , where the range between  $m_0$  and  $m_1$  is usually a fraction of the size of the largest out-of-sample period. The SW test is then applied to each out-of-sample loss series, and we take the maximum over all SW test statistics for all choices of the in-sample period, resulting in a DSW test. The test statistic is given by

$$DSW_{L^o(m)} = \max_{m \in [m_0, m_1]} SW_{L^o(m)},$$

where

$$SW_{L^o(m)} = \max_{T_b(m) \in [m+\tau+\varepsilon n, m+\tau+(1-\varepsilon)n]} \frac{SSR_{L^o(m)} - SSR(T_b(m))_{L^o(m)}}{\hat{V}_{L^o(m)}}. \quad (4)$$

$\varepsilon$  is a trimming parameter that is set to 0.1 throughout the simulations and application. For a given choice of  $m$ , the test maximizes the difference between the demeaned restricted sum of squared residuals of the out-of-sample loss series ( $SSR_{L^o(m)}$ ) and the demeaned unrestricted SSR of the out-of-sample loss series ( $SSR(T_b(m))_{L^o(m)}$ ).  $T_b(m)$  denotes the breakpoint, which results from a long memory robust CUSUM test proposed by [Horváth and Kokoszka \(1997\)](#) for Gaussian processes and extended by [Wang \(2008\)](#) to general linear processes. Both papers derive the limit distributions that converge to the supremum of a fractional Brownian bridge. The difference between the SSR is standardized by the MAC long-run variance estimate of the sample mean proposed by [Robinson \(2005\)](#) and [Abadir et al. \(2009\)](#), which is applied to the demeaned out-of-sample loss series with a break at time  $T_b(m)$ . The MAC estimator is defined as

$$\hat{V}_{L^o(m)}(d_L) = \hat{b}_w(d_L)p(d_L),$$

where

$$\hat{b}_w(d_L) = w^{-1} \sum_{j=1}^w \lambda_j^{2d_L} I_n(\lambda_j)$$

and

$$I_n(\lambda_j) = (2\pi n)^{-1} \left| \sum_{t=1}^n e^{it\lambda_j} L_t^o(m) \right|^2$$

is the periodogram,  $\lambda_j = 2\pi j/n$  are the Fourier frequencies and the bandwidth  $w$  converges to infinity and satisfies  $w = o(n/(\log n)^2)$ . Finally,

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

The MAC estimator requires an estimate of the long memory parameter of the out-of-sample loss series under the alternative. For this purpose, we use the local Whittle estimator

of [Künsch \(1987\)](#) and [Robinson \(1995\)](#),

$$\hat{d}_L = \underset{d_L \in [-1/2, 1/2]}{\operatorname{argmin}} \log \left( \frac{1}{w_{LW}} \sum_{j=1}^{w_{LW}} j^{2d_L} I_n(\lambda_j) \right) - \frac{2d_L}{w_{LW}} \sum_{j=1}^{w_{LW}} \log j.$$

For the bandwidth choices, we follow the simulation results of [Abadir et al. \(2009\)](#) and choose  $w = \lfloor n^{0.8} \rfloor$  for the bandwidth of the MAC estimator and  $w_{LW} = \lfloor n^{0.65} \rfloor$  for the bandwidth of the local Whittle estimator.

Since we are trying to detect a change in forecast accuracy, for each  $m \in [m_0, m_1]$  the selected in-sample period should be stable in order to provide stable forecasts. To obtain stable forecasts, a fixed forecasting scheme should be chosen where the in-sample period consists of observations  $1, \dots, m$ , because a rolling and a recursive scheme ensure that the mean of the in-sample losses approaches the mean of the out-of-sample losses rather quickly. This prevents forecast accuracy breakdown tests from generating power under a rolling and a recursive scheme. A fixed scheme may not optimize a given loss function, but that is not the goal of forecast breakdown tests.

We require the following assumption and additionally that  $T, m$  and  $n$  converge to infinity at the same rate.

**Assumption 4.** *Under the null hypothesis of no change in forecast accuracy, it holds for the total loss series  $L_{t+\tau} \equiv \{L_{t+\tau}\}_{t=1}^{T-2\tau+1}$ :  $E[L_{t+\tau}] = \mu \forall t$ ,  $T^{-1+d_L} E \left[ \sum_{t=1}^{\lfloor r(T-2\tau+1) \rfloor} (L_{t+\tau} - \mu) \right]^2 \xrightarrow{p} r\Omega$ , as  $T \rightarrow \infty$ , for  $r \in [0, 1]$  with  $\tau$  fixed, where  $\xrightarrow{p}$  denotes convergence in probability,  $m \in [m_0, m_1]$ ,  $\Omega$  is a full rank nonrandom matrix and  $T^{-1/2+d_L} \sum_{t=1}^{\lfloor r(T-2\tau+1) \rfloor} (L_{t+\tau} - \mu) \Rightarrow \Omega^{1/2} W_{d_L}(r)$ , where  $\Rightarrow$  denotes weak convergence in distribution and  $W_{d_L}(r)$  is a fractional Brownian motion defined on  $r$  with memory parameter  $d_L$ .*

The limit distribution of the test is derived in the Appendix and summarized in the following proposition.

**Proposition 3.** *Under Assumption 4, we have for the DSW test*

$$T^{d_L} DSW \Rightarrow \sup_{\mu \in [0, \bar{\mu}]} \sup_{\lambda \in [\mu + \varepsilon(1-\mu), 1 - \varepsilon(1-\mu)]} \frac{[(\lambda - \mu)W_{d_L}(1) + (1 - \lambda)W_{d_L}(\mu) - (1 - \mu)W_{d_L}(\lambda)]^2}{(1 - \lambda)(1 - \mu)(\lambda - \mu)},$$

as  $T, m, n \rightarrow \infty$  at the same rate.  $W_{d_L}(r)$  is a fractional Brownian motion defined on  $r \in [0, 1]$  and  $\bar{\mu} = \lim_{T \rightarrow \infty} \frac{m_1 - m_0}{n_0}$  with  $n_0 = T - m_0 - \tau + 1$ .

**Proof.** *See the Appendix.*

## 4 Monte Carlo Study

To examine the finite sample size and power properties of our proposed test, we conduct a Monte Carlo simulation study. The critical values in Table 1 are obtained under the

$\alpha/d$	0.1	0.2	0.3	0.4
10%	7.572	7.686	7.475	7.410
5%	9.022	9.335	9.095	9.167
1%	12.693	12.837	13.046	12.824

Table 1: Critical values for FWN processes with 5,000 repetitions,  $T = 1,000$ ,  $m_0 = \lfloor 0.2T \rfloor$ ,  $\bar{\mu} = 0.3$ ,  $\tau = 1$  and  $\varepsilon = 0.1$ .

null hypothesis with 5,000 repetitions of  $T = 1,000$  observations of fractional white noise (FWN) processes with a fractional parameter  $d \in \{0.1, 0.2, 0.3, 0.4\}$ , Gaussian innovations and a significance level of 10%, 5%, and 1%. In addition, we account for short-run dynamics by adding an autoregressive parameter  $\phi \in \{0.1, 0.2, 0.3, 0.4\}$  to the FWN processes. The corresponding critical values can be found in Table 4 in the Appendix. The critical values, the size and the power results are obtained with  $m_0 = \lfloor 0.2T \rfloor$ , so that 20% of the sample data is used for the smallest choice of the in-sample period,  $\bar{\mu} = 0.3$ , which implies that

30% of the largest out-of-sample period is used for the window defining  $m_1$ , the size of the largest in-sample period. We consider a forecast horizon of  $\tau = 1$  and a trimming value of  $\varepsilon = 0.1$ . Changing these values slightly does not significantly change the results. We only report results for a fixed forecast scheme because, similar to Perron and Yamamoto (2021), we find that this is the only forecast scheme that ensures a monotonic power function. We elaborate on the loss of power under a recursive and a rolling forecast scheme in the supplementary material.

We conduct a local power analysis comparing the Perron and Yamamoto (2021) test with the heteroscedasticity and autocorrelation consistent estimator of Andrews (1991) for the long-run variance correction factor in Eq. (4) and the Bai and Perron (1998) breakpoint estimator (hereafter abbreviated as HAC test) with our proposed test (hereafter abbreviated as MAC test). For the size and the power we use 2,000 Monte Carlo repetitions and  $T = 500$  observations. Moreover, we let  $d \in \{0.1, 0.2, 0.3, 0.4\}$  and  $\phi \in \{0, 0.1, 0.2, 0.3, 0.4\}$  and we consider mean shifts in the simulated autoregressive fractionally integrated moving average (ARFIMA) processes starting from the three hundredth observation of increasing size  $kT^{d-1/2}$  for  $k = 0, 1, \dots, 25$ .

Table 2 shows the empirical size for nominal significance levels of 10%, 5%, and 1%. On average, the empirical size of our MAC test is closer to the nominal size than the empirical size of the HAC test. At the same time, our test is slightly undersized on average, while the HAC test is oversized for all parameter combinations. The liberalness of the HAC test increases on average with increasing fractional parameter value. For increasing short-run dynamics, the liberalness of the HAC test seems to decrease, at least for  $d = 0.4$ .

Figure 1 shows the local power functions for FWN processes with a nominal significance level of 5%. Overall, the MAC test is superior in terms of power for all values of the fractional parameter. The power differences increase with increasing fractional parameter value and the HAC test with  $d = 0.4$  even loses power.

$\phi = 0$								
$\alpha/d$	MAC				HAC			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
10%	0.072	0.092	0.092	0.088	0.111	0.106	0.112	0.128
5%	0.043	0.048	0.043	0.052	0.062	0.062	0.065	0.071
1%	0.011	0.011	0.013	0.011	0.013	0.025	0.020	0.015
$\phi = 0.1$								
10%	0.082	0.086	0.090	0.130	0.112	0.114	0.130	0.174
5%	0.046	0.045	0.049	0.065	0.066	0.068	0.072	0.101
1%	0.012	0.012	0.012	0.022	0.021	0.013	0.022	0.050
$\phi = 0.2$								
10%	0.090	0.084	0.095	0.119	0.102	0.112	0.140	0.168
5%	0.053	0.053	0.045	0.071	0.070	0.070	0.086	0.110
1%	0.012	0.015	0.009	0.018	0.014	0.020	0.028	0.039
$\phi = 0.3$								
10%	0.079	0.087	0.096	0.104	0.112	0.107	0.136	0.154
5%	0.056	0.052	0.044	0.044	0.063	0.063	0.073	0.096
1%	0.012	0.012	0.012	0.012	0.020	0.015	0.022	0.040
$\phi = 0.4$								
10%	0.078	0.084	0.096	0.096	0.114	0.113	0.128	0.135
5%	0.042	0.041	0.045	0.045	0.060	0.059	0.063	0.081
1%	0.005	0.009	0.016	0.012	0.013	0.015	0.025	0.035

Table 2: Empirical size of the MAC and HAC test for ARFIMA( $\phi, d, 0$ ) processes with 2,000 repetitions,  $T = 500$ ,  $m_0 = \lfloor 0.2T \rfloor$ ,  $\bar{\mu} = 0.3$ ,  $\tau = 1$  and  $\varepsilon = 0.1$ .

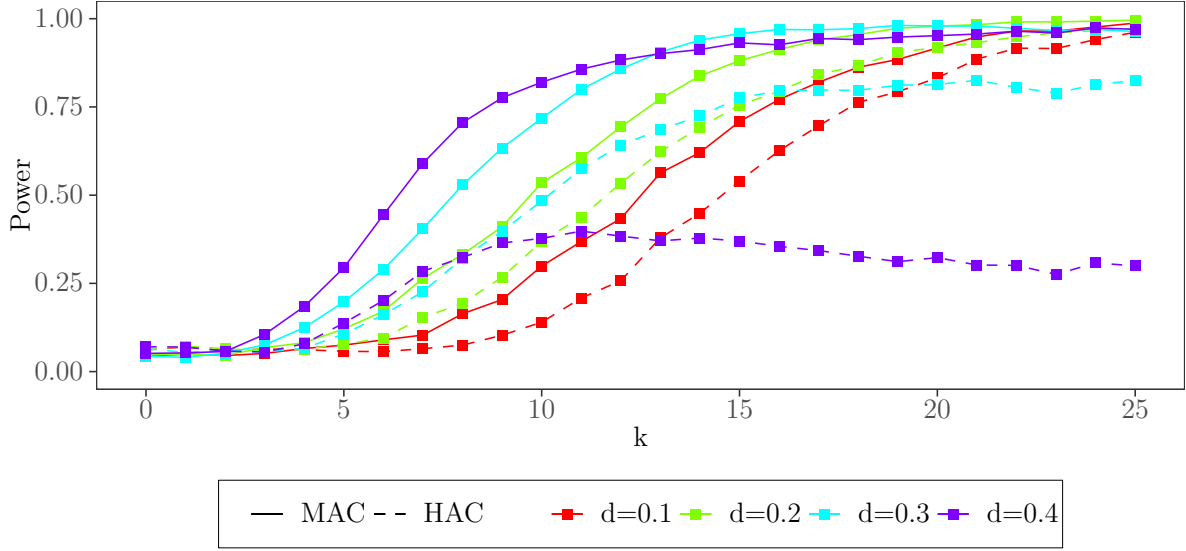


Figure 1: Local power for a mean break in FWN processes with long memory parameter  $d$ , 2,000 repetitions,  $T = 500$ ,  $m_0 = \lfloor 0.2T \rfloor$ ,  $\bar{\mu} = 0.3$ ,  $\tau = 1$  and  $\varepsilon = 0.1$ .

Figure 2 introduces short-run dynamics of increasing size to the simulated FWN processes. Our test is still superior in terms of power, but the power of the HAC test increases overall. On average, both tests lose power when the short-run dynamics increase.

We also considered variance changes in the simulated ARFIMA processes, as [Giacomini and Rossi \(2009\)](#) state that forecast accuracy breakdown tests with a squared error loss function can detect variance changes as well. However, as we have shown in Section 2, a memory reduction to zero is possible when the forecast is unbiased. Both tests are still consistent under a variance break, but they perform similarly in terms of power, so we do not report the results.

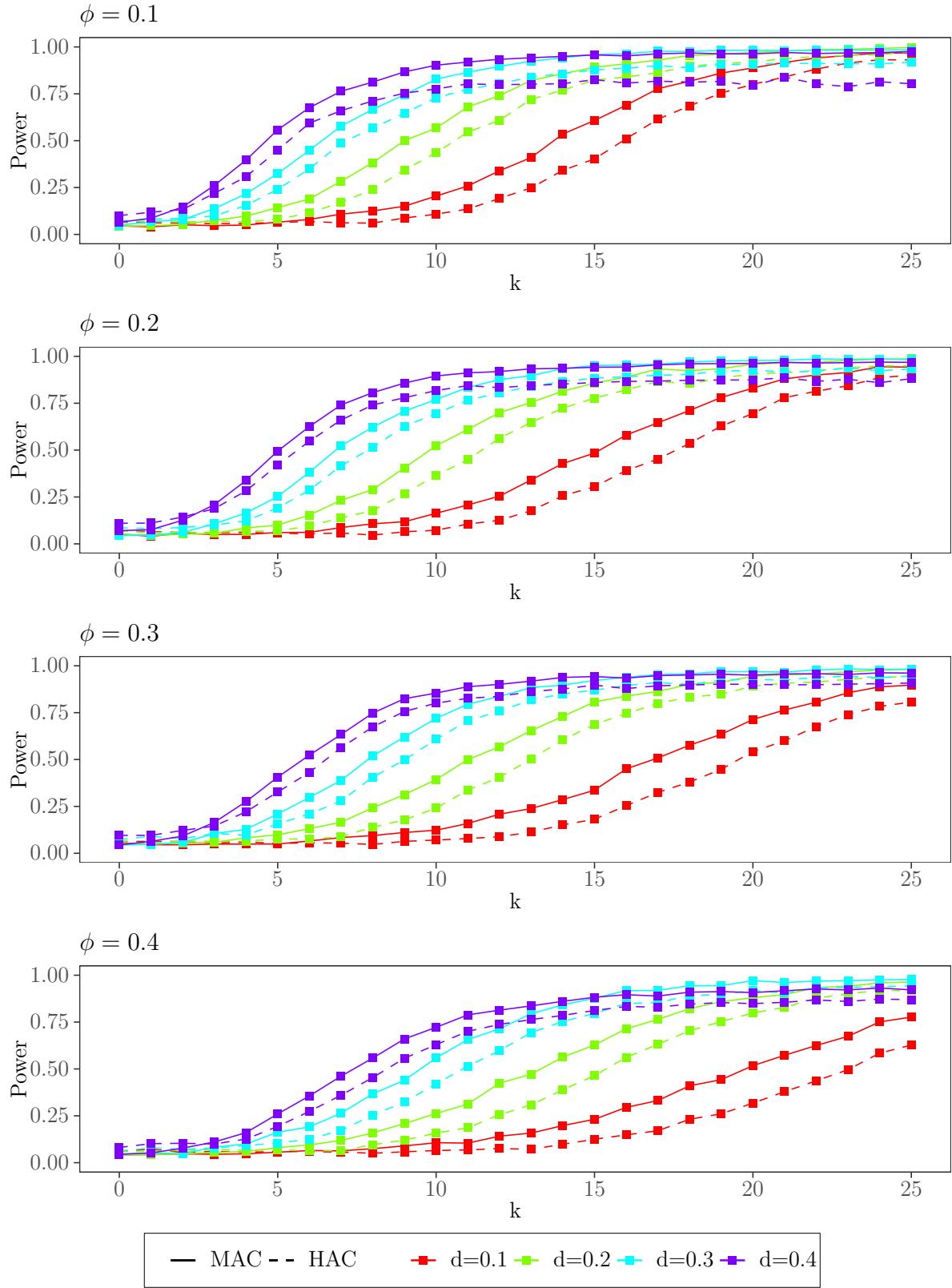


Figure 2: Local power for a mean break in ARFIMA( $\phi, d, 0$ ) processes with 2,000 repetitions,  $T = 500$ ,  $m_0 = \lfloor 0.2T \rfloor$ ,  $\bar{\mu} = 0.3$ ,  $\tau = 1$  and  $\varepsilon = 0.1$ .



## 5 Breaks in Forecast Accuracy during the Global Energy Crisis

The global energy crisis that commenced in 2021 resulted in a significant surge in electricity prices. In this analysis, we apply our proposed test to European and U.S. daily wholesale electricity prices with the objective of identifying differences in the severity and timing of the changes in forecasting performance in response to the crisis. The European electricity price series include data from Denmark, France, Germany, the Netherlands, and Norway. They are obtained from the German Federal Network Agency. The U.S. electricity price hubs considered are Mass Hub, Mid-C, Palo Verde, PJM West, and SP-15. The data set was obtained from the U.S. Energy Information System. The regions the price hubs cover can be found on the organization’s website. It should be noted that additional U.S. price hubs are not considered in this analysis due to incomplete data. In order to facilitate comparison with the U.S. data, weekends have been removed from the European data set. In view of the sharp rise in inflation following the global pandemic, all price series were adjusted for inflation using an electricity-specific price index. The initial year of our analysis, 2015, has been designated as the base year. The electricity price indices for the European countries are obtained from the corresponding national statistical authorities<sup>1</sup>, while the corresponding index for the U.S. has been obtained from the U.S. Bureau of Labor Statistics. Figure 3 shows the inflation-adjusted electricity price series for Germany and the price hub PJM West, spanning from 01/01/2015 to 05/31/2024. It also illustrates the autocorrelation function (ACF) and the periodogram of the corresponding series. The remaining price series, ACFs and periodograms are shown in Figures 4, 5, 6 and 7 in the Appendix. The series are highly persistent with a pole at frequency zero, indicating that the

---

<sup>1</sup>Statistics Denmark, The National Institute of Statistics and Economic Studies (France), Statistisches Bundesamt (Germany), Centraal Bureau voor de Statistiek (The Netherlands), Statistics Norway.

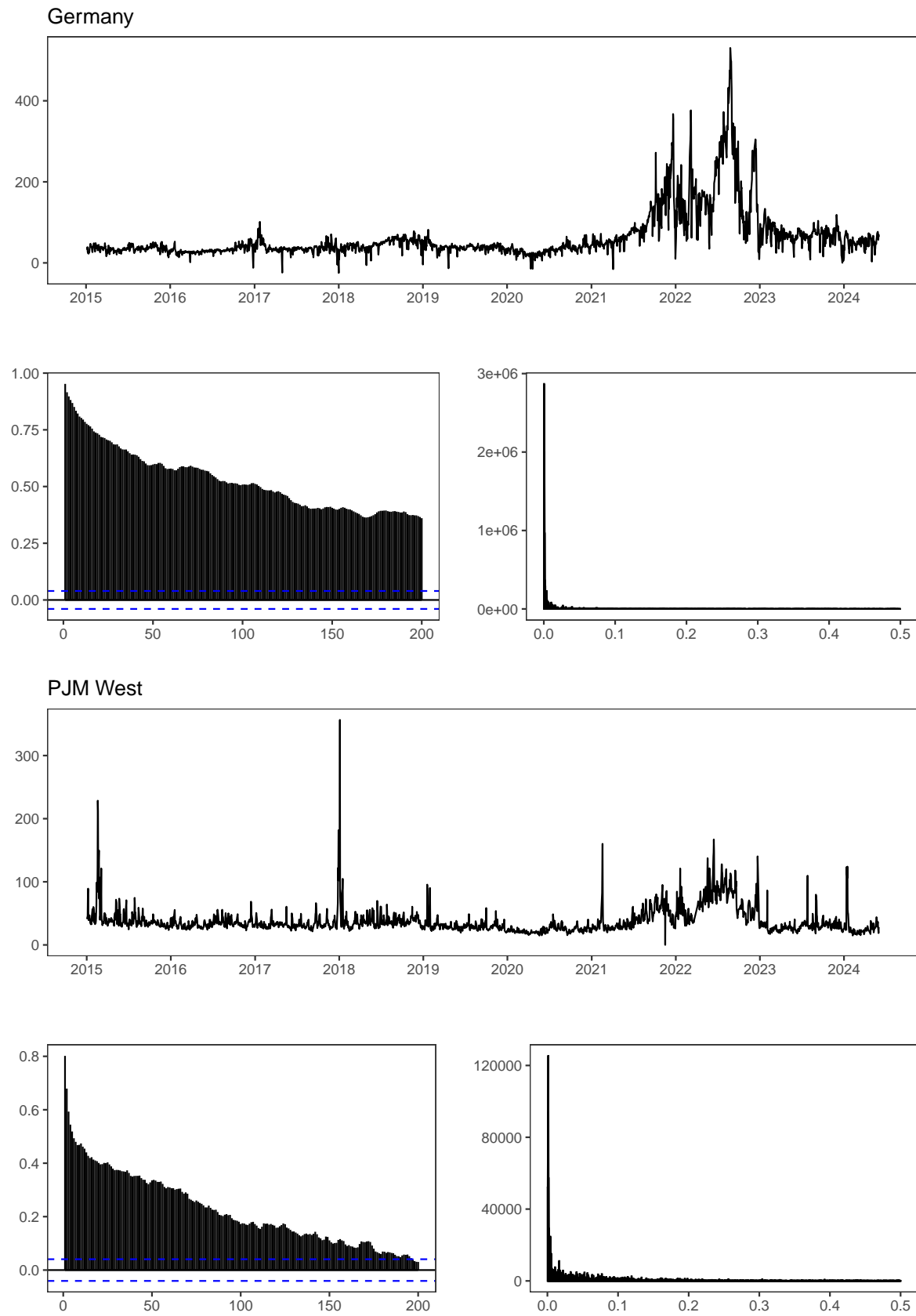


Figure 3: Price series, ACF and periodogram for Germany and PJM West.

Electricity Price Series	90% Crit.	95% Crit.	99% Crit.	Test Statistic	Break Date
Mass Hub	11.004	13.747	21.940	20.852	12/06/2021
Mid-C	11.325	14.348	24.593	8.914	-
Palo Verde	10.527	12.841	18.209	3.901	-
PJM West	10.159	11.946	16.208	22.567	02/10/2021
SP-15	10.324	12.327	18.910	10.657	12/08/2020
Denmark	10.778	13.273	21.481	262.739	09/22/2021
France	11.964	14.980	25.609	91.794	10/05/2021
Germany	9.807	11.828	16.343	248.031	09/22/2021
The Netherlands	10.175	12.187	16.468	152.034	07/28/2021
Norway	12.546	16.145	34.246	127.330	11/26/2021

Table 3: Critical values, test statistic and corresponding break date for the U.S. and European electricity price series.

series possess long memory properties. The characteristic weekly seasonality of electricity prices, which arises from the change in electricity demand and generation during the week as opposed to the weekend, is not evident in the data set, as weekends are not included. The seasonal component of electricity prices has already been analyzed by [Haldrup and Nielsen \(2006\)](#) in a long memory framework.

For the application of our test, we select the smallest in-sample period ending on 01/01/2018 and the largest in-sample period ending on 12/31/2019. The critical values for each time series are obtained by simulating 5,000 ARFIMA paths with  $T = 1,000$  observations,  $m_0 = \lfloor 0.2T \rfloor$ ,  $\bar{\mu} = 0.3$ ,  $\tau = 1$  and  $\varepsilon = 0.1$ . Table 3 shows the critical values, the test statistic and the break date if the test statistic is significant at any significance level. The European price series are subject to a break in forecast accuracy that is between the end

of July and the end of November 2021. Thus, the breaks are before the Russian invasion of Ukraine on 02/24/2022 and the corresponding European Union sanctions against Russia. The increase in global electricity prices in 2021 is due to increased demand for energy following the pandemic, which could not be met by supply. Another factor that may have had an impact on electricity prices in the European Union is the start of Phase IV of the EU ETS in 2021. Phase IV involves a higher, linearly increasing reduction factor for the EU-wide cap on emission allowances. As a result of the cap, the price per ton of CO<sub>2</sub> and the price of gas have risen sharply, while high gas prices drive up electricity prices through the merit order effect. For the U.S. price series, a break is identified in 2021 for Mass Hub and PJM West. The break in August 2020 for SP-15 is statistically significant at the 10% level.

## 6 Conclusion

In this paper, we propose a DSW test for detecting a forecast accuracy breakdown under a long memory time series setting. To account for the long memory properties, we standardize the test statistic with a MAC long-run variance estimate of the sample mean. The corresponding breakpoint is estimated with a long memory robust CUSUM test. The finite sample size and power properties of the test are derived in a Monte Carlo simulation. We show that our approach is superior in terms of size and power to a HAC estimate for the long-run variance if the underlying process exhibits long memory properties. The practical relevance of the method is demonstrated by applying the test to European and U.S. electricity prices. We find that European countries were more exposed to the global energy crisis that started in 2021.

There are ample opportunities to extend our flexible approach. The quadratic loss function may be replaced by any other loss function that the applied econometrician deems appro-

priate. Future research should also allow for multiple breaks, which may also occur in the in-sample period.

## Appendix

### Proofs

**Proof of Proposition 1.** The centered series  $a_t^* = a_t - \mu_a$  allows to reformulate the squared forecast error in Eq. (1).

$$\begin{aligned}
L_t &= (y_t - \hat{y}_t)^2 \\
&= (y_t^* + \mu_y)^2 + (\hat{y}_t^* + \mu_{\hat{y}})^2 - 2[(y_t^* + \mu_y)(\hat{y}_t^* + \mu_{\hat{y}})] \\
&= \underbrace{2[y_t^*(\mu_y - \mu_{\hat{y}}) - \hat{y}_t^*(\mu_y - \mu_{\hat{y}})]}_{\text{I}} - \underbrace{2y_t^*\hat{y}_t^*}_{\text{II}} + \underbrace{y_t^{*2} + \hat{y}_t^{*2}}_{\text{III}} + \text{const.}
\end{aligned} \tag{5}$$

We will now consider each component (I, II and III) of Eq. (5) individually. For I, recall that  $y_t^* \sim LM(d_y)$  and  $\hat{y}_t^* \sim LM(d_{\hat{y}})$ . For components II and III, Proposition 1 of [Leschinski \(2017\)](#) can be used, as it applies to products and squares of long memory time series. Thus, for component II we have

$$y_t^*\hat{y}_t^* \sim \begin{cases} LM(d_y + d_{\hat{y}} - 1/2), & \text{if } S_{y\hat{y}} = 0 \\ LM(\max\{d_y + d_{\hat{y}} - 1/2, 0\}), & \text{otherwise,} \end{cases} \tag{6}$$

where  $S_{y\hat{y}} = \sum_{k=-\infty}^{\infty} \gamma_y(k)\gamma_{\hat{y}}(k)$  and  $\gamma(k)$  is the autocovariance function. For component III we get

$$a_t^{*2} \sim LM(\max\{2d_a - 1/2, 0\}) \text{ for } a_t^* \in \{y_t^*, \hat{y}_t^*\}. \tag{7}$$

Based on  $0 \leq d_a < 1/2$  for  $a_t \in \{y_t, \hat{y}_t\}$ , we can generally state that

$$d_y > d_y + d_{\hat{y}} - 1/2, \quad d_{\hat{y}} > d_y + d_{\hat{y}} - 1/2 \tag{8}$$

and

$$d_a > 2d_a - 1/2. \tag{9}$$

The following case-by-case analysis for the expected values derives the memory order of the squared forecast error in Eq. (5). First, if  $\mu_y \neq \mu_{\hat{y}}$ , the order of the original terms dominates due to Eq. (8) and (9), and we get

$$d_L = \max\{d_y, d_{\hat{y}}\}.$$

Second, if  $\mu_y = \mu_{\hat{y}}$ , component I disappears and the linear combination of components II and III remains. It follows from Eq. (7) that the lower bound for the memory order of a squared long memory time series is zero and therefore, according to Proposition 3 of [Chambers \(1998\)](#), the lower bound for the remaining compound term is zero. In addition, the memory order of one of the squared series in Eq. (7) always dominates the memory of the product series in Eq. (6) and therefore,

$$d_L = \max\{2\max\{d_y, d_{\hat{y}}\} - 1/2, 0\}.$$

□

***Proof of Proposition 2.*** Under the mean relations  $\mu_y = \beta_y + \kappa_y \mu_x$  and  $\mu_{\hat{y}} = \beta_{\hat{y}} + \kappa_{\hat{y}} \mu_x$ , Assumption 3 and the centered series  $x_t^* = x_t - \mu_x$ , a similar expression as in Eq. (5) can

be obtained.

$$\begin{aligned}
L_t &= (y_t - \hat{y}_t)^2 \\
&= (\beta_y + x_t \kappa_y + \epsilon_{y,t})^2 + (\beta_{\hat{y}} + x_t \kappa_{\hat{y}} + \epsilon_{\hat{y},t})^2 - 2(\beta_y + x_t \kappa_y + \epsilon_{y,t})(\beta_{\hat{y}} + x_t \kappa_{\hat{y}} + \epsilon_{\hat{y},t}) \\
&= (\mu_y + x_t^* \kappa_y + \epsilon_{y,t})^2 + (\mu_{\hat{y}} + x_t^* \kappa_{\hat{y}} + \epsilon_{\hat{y},t})^2 - 2(\mu_y + x_t^* \kappa_y + \epsilon_{y,t})(\mu_{\hat{y}} + x_t^* \kappa_{\hat{y}} + \epsilon_{\hat{y},t}) \\
&= x_t^{*2} \kappa_y^2 + 2x_t^* \mu_y \kappa_y + 2x_t^* \kappa_y \epsilon_{y,t} + x_t^{*2} \kappa_{\hat{y}}^2 + 2x_t^* \mu_{\hat{y}} \kappa_{\hat{y}} + 2x_t^* \kappa_{\hat{y}} \epsilon_{\hat{y},t} \\
&\quad - 2 \left[ x_t^* \mu_y \kappa_{\hat{y}} + x_t^* \kappa_y \mu_{\hat{y}} + x_t^{*2} \kappa_y \kappa_{\hat{y}} + x_t^* \kappa_y \epsilon_{\hat{y},t} + x_t^* \kappa_{\hat{y}} \epsilon_{y,t} \right] \\
&\quad + \epsilon_{y,t}^2 + \epsilon_{\hat{y},t}^2 + 2\mu_y \epsilon_{y,t} + 2\mu_{\hat{y}} \epsilon_{\hat{y},t} - 2[\mu_y \epsilon_{\hat{y},t} + \mu_{\hat{y}} \epsilon_{y,t} + \epsilon_{y,t} \epsilon_{\hat{y},t}] + \underbrace{\mu_y^2 + \mu_{\hat{y}}^2 - 2\mu_y \mu_{\hat{y}}}_{\text{const.}} \\
&= 2 \left\{ x_t^* [\kappa_y (\mu_y - \mu_{\hat{y}}) - \kappa_{\hat{y}} (\mu_y - \mu_{\hat{y}})] + x_t^* \{ (\kappa_y - \kappa_{\hat{y}}) \epsilon_{y,t} - (\kappa_y - \kappa_{\hat{y}}) \epsilon_{\hat{y},t} \} \right. \\
&\quad \left. + x_t^{*2} \{ \kappa_y \kappa_{\hat{y}} + 0.5(\kappa_y^2 - \kappa_{\hat{y}}^2) \} + (\mu_y - \mu_{\hat{y}}) \epsilon_{y,t} - (\mu_y - \mu_{\hat{y}}) \epsilon_{\hat{y},t} - \epsilon_{y,t} \epsilon_{\hat{y},t} \right\} \\
&\quad + \epsilon_{y,t}^2 + \epsilon_{\hat{y},t}^2 + \text{const.}
\end{aligned} \tag{10}$$

To determine the memory order of the squared forecast error under fractional cointegration, first note that  $x_t^* [\kappa_y (\mu_y - \mu_{\hat{y}}) - \kappa_{\hat{y}} (\mu_y - \mu_{\hat{y}})]$  in Eq. (10) has a similar structure as  $y_t^* (\mu_y - \mu_{\hat{y}}) - \hat{y}_t^* (\mu_y - \mu_{\hat{y}})$  in Eq. (5). The remaining terms in Eq. (10) are products or squares of centered series and therefore their memory is reduced according to Eq. (8) and (9). It follows from Assumption 3 that  $x_t^*$  has the same memory order as the series that are fractionally cointegrated ( $y_t$  and  $\hat{y}_t$ ) and also that  $d_x > d_{\epsilon_{y,t}}, d_{\epsilon_{\hat{y},t}}$ . Consequently, the reduced memory order of all product and square series is always dominated by the memory of  $x_t^*$ . Finally, the memory order of the linear combination in Eq. (10) cannot be less than that of  $x_t^*$  whenever a bias term is non-zero. This describes the case of a biased forecast with  $\kappa_y \neq \kappa_{\hat{y}}$ .

To analyze the case of  $\kappa_y = \kappa_{\hat{y}}$  with a biased forecast, the term in square brackets in Eq. (10) is reformulated as  $\mu_y (\kappa_y - \kappa_{\hat{y}}) - \mu_{\hat{y}} (\kappa_y - \kappa_{\hat{y}}) = 0$  so that Eq. (10) simplifies to

$$L_t = 2 \{ x_t^{*2} \kappa_y \kappa_{\hat{y}} + (\mu_y - \mu_{\hat{y}}) \epsilon_{y,t} - (\mu_y - \mu_{\hat{y}}) \epsilon_{\hat{y},t} - \epsilon_{y,t} \epsilon_{\hat{y},t} \} + \epsilon_{y,t}^2 + \epsilon_{\hat{y},t}^2 + \text{const.}$$

We know from Eq. (7) and (9) and from Assumption 3 that the memory order is reduced

compared to the order of  $x_t^*$ , but it is not possible to derive the exact memory orders of  $\epsilon_{y,t}$  and  $\epsilon_{\hat{y},t}$ , and thus we cannot determine whether the memory order of  $x_t^{*2}$  or the errors dominate.

Finally, we consider the case of an unbiased forecast with  $\kappa_y \neq \kappa_{\hat{y}}$ . Eq. (10) then simplifies to

$$L_t = \underbrace{2x_t^{*2}\kappa_y\kappa_{\hat{y}}}_{\text{I}} - \underbrace{2\epsilon_{y,t}\epsilon_{\hat{y},t}}_{\text{II}} + \underbrace{\epsilon_{y,t}^2 + \epsilon_{\hat{y},t}^2}_{\text{III}} + \text{const.} \quad (11)$$

Again, the memory orders of the individual components in Eq. (11) are determined according to Proposition 1 of [Leschinski \(2017\)](#), and finally the memory of the entire series is given by the maximum of the memory orders of the individual components according to [Chambers \(1998\)](#). We get

$$\begin{aligned} \text{I} &\sim LM(\max\{2d_x - 1/2, 0\}) \\ \text{II} &\sim \begin{cases} LM(d_{\epsilon_{y,t}} + d_{\epsilon_{\hat{y},t}} - 1/2), & \text{if } S_{\epsilon_{y,t}\epsilon_{\hat{y},t}} = 0 \\ LM(\max\{d_{\epsilon_{y,t}} + d_{\epsilon_{\hat{y},t}} - 1/2, 0\}), & \text{otherwise} \end{cases} \\ \text{III} &\sim LM(\max\{2d_a - 1/2, 0\}), \text{ for } a \in \{\epsilon_{y,t}, \epsilon_{\hat{y},t}\}. \end{aligned}$$

Since  $d_x > d_{\epsilon_{y,t}}, d_{\epsilon_{\hat{y},t}}$  holds by Assumption 3, the memory order of component I dominates the memory order of component III and, for the same reason, I also dominates II. Consequently, with an unbiased forecast and  $\kappa_y \neq \kappa_{\hat{y}}$ , the memory order of the squared forecast error is equal to the order of component I

$$d_L = \max\{2d_x - 1/2, 0\}.$$

□

**Proof of Proposition 3.** We follow the proof of [Perron and Yamamoto \(2021\)](#), adapting it where necessary. The Wald test for a constant mean against a break at time  $T_b(m) = T_b = m_0 + \tau + \lfloor \lambda n_0 \rfloor$  for the loss series  $\{L_{t+\tau}^o\}_{t=m}^{T-\tau}$  is



$$W^m(T_b) = \frac{SSR_{L^o(m)} - SSR(T_b)_{L^o(m)}}{\hat{V}_{L^o(m)}}.$$

For a given  $m$ , the restricted SSR is

$$\begin{aligned} SSR_{L^o(m)} &= \sum_{t=m}^{T-\tau} L_{t+\tau}^{o^2} - \frac{1}{T-\tau-m+1} \left( \sum_{t=m}^{T-\tau} L_{t+\tau}^o \right)^2 \\ &= \sum_{t=m}^{T-\tau} L_{t+\tau}^{o^2} - \frac{n_0}{T-\tau-m+1} \left( n_0^{-1/2} \sum_{t=m}^{T-\tau} L_{t+\tau}^o \right)^2. \end{aligned}$$

The unrestricted SSR assuming a break at time  $T_b$  is given by

$$\begin{aligned} SSR(T_b)_{L^o(m)} &= \sum_{t=m}^{T_b-\tau} L_{t+\tau}^{o^2} - \frac{n_0}{T_b-\tau-m+1} \left( n_0^{-1/2} \sum_{t=m}^{T_b-\tau} L_{t+\tau}^o \right)^2 \\ &+ \sum_{t=T_b-\tau+1}^{T-\tau} L_{t+\tau}^{o^2} - \frac{n_0}{T-T_b} \left( n_0^{-1/2} \sum_{t=T_b-\tau+1}^{T-\tau} L_{t+\tau}^o \right)^2. \end{aligned}$$

We thus have

$$\begin{aligned} SSR_{L^o(m)} - SSR(T_b)_{L^o(m)} &= -\frac{n_0}{T-\tau-m+1} \left( n_0^{-1/2} \sum_{t=m}^{T-\tau} L_{t+\tau}^o \right)^2 \\ &+ \frac{n_0}{T_b-\tau-m+1} \left( n_0^{-1/2} \sum_{t=m}^{T_b-\tau} L_{t+\tau}^o \right)^2 \\ &+ \frac{n_0}{T-T_b} \left( n_0^{-1/2} \sum_{t=T_b-\tau+1}^{T-\tau} L_{t+\tau}^o \right)^2. \end{aligned}$$

Let  $\mu = \lim_{T \rightarrow \infty} \frac{m-m_0}{n_0}$ . With  $T = n_0 + m_0 + \tau - 1$  and  $T_b = m_0 + \tau + \lfloor \lambda n_0 \rfloor$ , we have

$$\begin{aligned} \frac{n_0}{T-\tau-m+1} &= \frac{n_0}{n_0 - (m - m_0)} \rightarrow \frac{1}{1-\mu}, \\ \frac{n_0}{T_b-\tau-m+1} &= \frac{n_0}{\lfloor \lambda n_0 \rfloor - (m - m_0) + 1} \rightarrow \frac{1}{\lambda - \mu}, \\ \frac{n_0}{T-T_b} &= \frac{n_0}{n_0 - \lfloor \lambda n_0 \rfloor - 1} \rightarrow \frac{1}{1-\lambda}. \end{aligned}$$

Since  $L_{t+\tau-1}^o = L_t$  for  $t = m+1, \dots, T-\tau+1$  and using Assumption 4, we get

$$\begin{aligned}
n_0^{-1/2+d_L} \sum_{t=m}^{T-\tau} L_{t+\tau}^o &= n_0^{-1/2+d_L} \sum_{t=m-\tau+1}^{T-2\tau+1} L_{t+\tau} \\
&= n_0^{-1/2+d_L} \sum_{t=m_0-\tau}^{T-2\tau+1} L_{t+\tau} - n_0^{-1/2+d_L} \sum_{t=m_0-\tau}^{m-\tau} L_{t+\tau} \\
&\Rightarrow \Omega^{1/2}[W_{d_L}(1) - W_{d_L}(\mu)],
\end{aligned}$$

$$\begin{aligned}
n_0^{-1/2+d_L} \sum_{t=m}^{T_b-\tau} L_{t+\tau}^o &= n_0^{-1/2+d_L} \sum_{t=m-\tau+1}^{T_b-2\tau+1} L_{t+\tau} \\
&= n_0^{-1/2+d_L} \sum_{t=m_0-\tau}^{T_b-2\tau+1} L_{t+\tau} - n_0^{-1/2+d_L} \sum_{t=m_0-\tau}^{m-\tau} L_{t+\tau} \\
&\Rightarrow \Omega^{1/2}[W_{d_L}(\lambda) - W_{d_L}(\mu)],
\end{aligned}$$

$$\begin{aligned}
n_0^{-1/2+d_L} \sum_{t=T_b-\tau+1}^{T-\tau} L_{t+\tau}^o &= n_0^{-1/2+d_L} \sum_{t=T_b-2\tau+2}^{T-2\tau+1} L_{t+\tau} \\
&= n_0^{-1/2+d_L} \sum_{t=m_0-\tau}^{T-2\tau+1} L_{t+\tau} - n_0^{-1/2+d_L} \sum_{t=m_0-\tau}^{T_b-2\tau+1} L_{t+\tau} \\
&\Rightarrow \Omega^{1/2}[W_{d_L}(1) - W_{d_L}(\lambda)].
\end{aligned}$$

Combining these results, we obtain

$$\begin{aligned}
&T^{2d_L} (SSR_{L^o(m)} - SSR(T_b)_{L^o(m)}) \Rightarrow \\
&\Omega \left[ 1 - \frac{[W_{d_L}(1) - W_{d_L}(\mu)]^2}{1 - \mu} + \frac{[W_{d_L}(\lambda) - W_{d_L}(\mu)]^2}{\lambda - \mu} + \frac{[W_{d_L}(1) - W_{d_L}(\lambda)]^2}{1 - \lambda} \right],
\end{aligned}$$

as under the null hypothesis  $\hat{V}_{L^o(m)} \xrightarrow{p} \Omega$ .

Note that  $\mu = \lim_{T \rightarrow \infty} \frac{m-m_0}{n_0} \leq \lim_{T \rightarrow \infty} \frac{m_1-m_0}{n_0} = \bar{\mu}$ , so that  $\mu \in [0, \bar{\mu}]$ . Further, we have

for the trimming parameter  $\varepsilon$ ,

$$\begin{aligned}
T_b &\in [m + \tau + \varepsilon n, m + \tau + (1 - \varepsilon)n], \\
\frac{T_b - m_0 - \tau}{n_0} &\in \left[ \frac{(m - m_0) + \varepsilon n}{n_0}, \frac{(m - m_0) + (1 - \varepsilon)n}{n_0} \right], \\
\frac{T_b - m_0 - \tau}{n_0} &\in \left[ \frac{(m - m_0) + \varepsilon(n_0 + m_0 - m)}{n_0}, \frac{(m - m_0) + (1 - \varepsilon)(n_0 + m_0 - m)}{n_0} \right].
\end{aligned}$$

Taking the limit implies  $\lambda \in [\mu + \varepsilon(1 - \mu), 1 - \varepsilon(1 - \mu)]$ , and the result follows.  $\square$

$\alpha/d$	$\phi = 0.1$				$\phi = 0.2$			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
10%	7.530	7.562	7.931	8.130	7.327	7.611	7.937	8.246
5%	9.147	9.280	9.875	10.058	8.970	9.287	9.916	10.528
1%	12.984	13.000	13.943	15.867	12.859	12.918	14.927	16.898
$\alpha/d$	$\phi = 0.3$				$\phi = 0.4$			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
10%	7.458	7.670	8.032	8.397	7.494	7.570	7.719	8.413
5%	8.964	9.118	9.771	10.483	9.032	9.216	9.482	10.339
1%	12.402	12.885	13.805	19.454	12.897	13.330	13.806	15.175

Table 4: Critical values for ARFIMA( $\phi, d, 0$ ) processes with 5,000 repetitions,  $T = 1,000$ ,  $m_0 = \lfloor 0.2T \rfloor$ ,  $\bar{\mu} = 0.3$ ,  $\tau = 1$  and  $\varepsilon = 0.1$ .

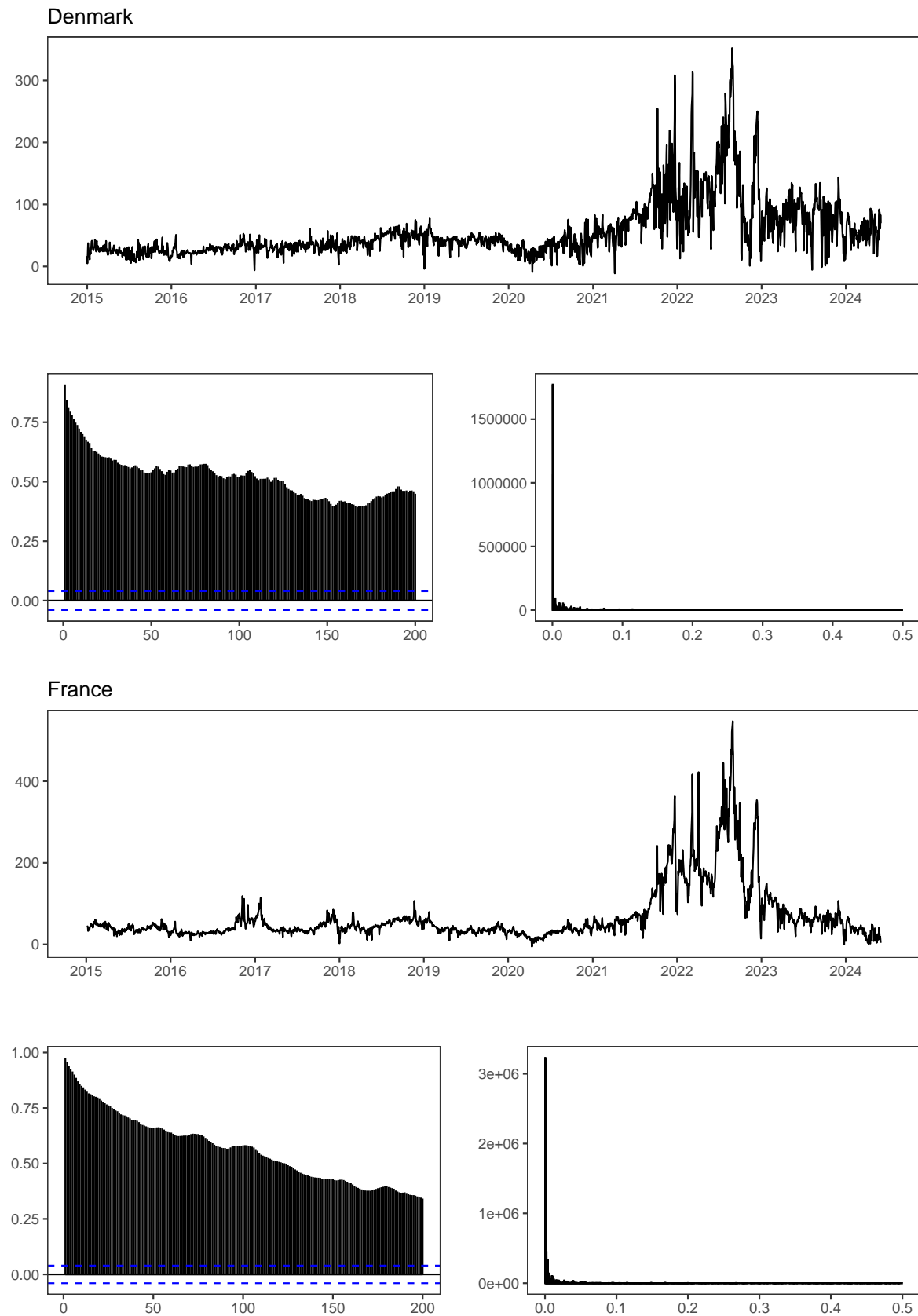


Figure 4: Price series, ACF and periodogram for Denmark and France.

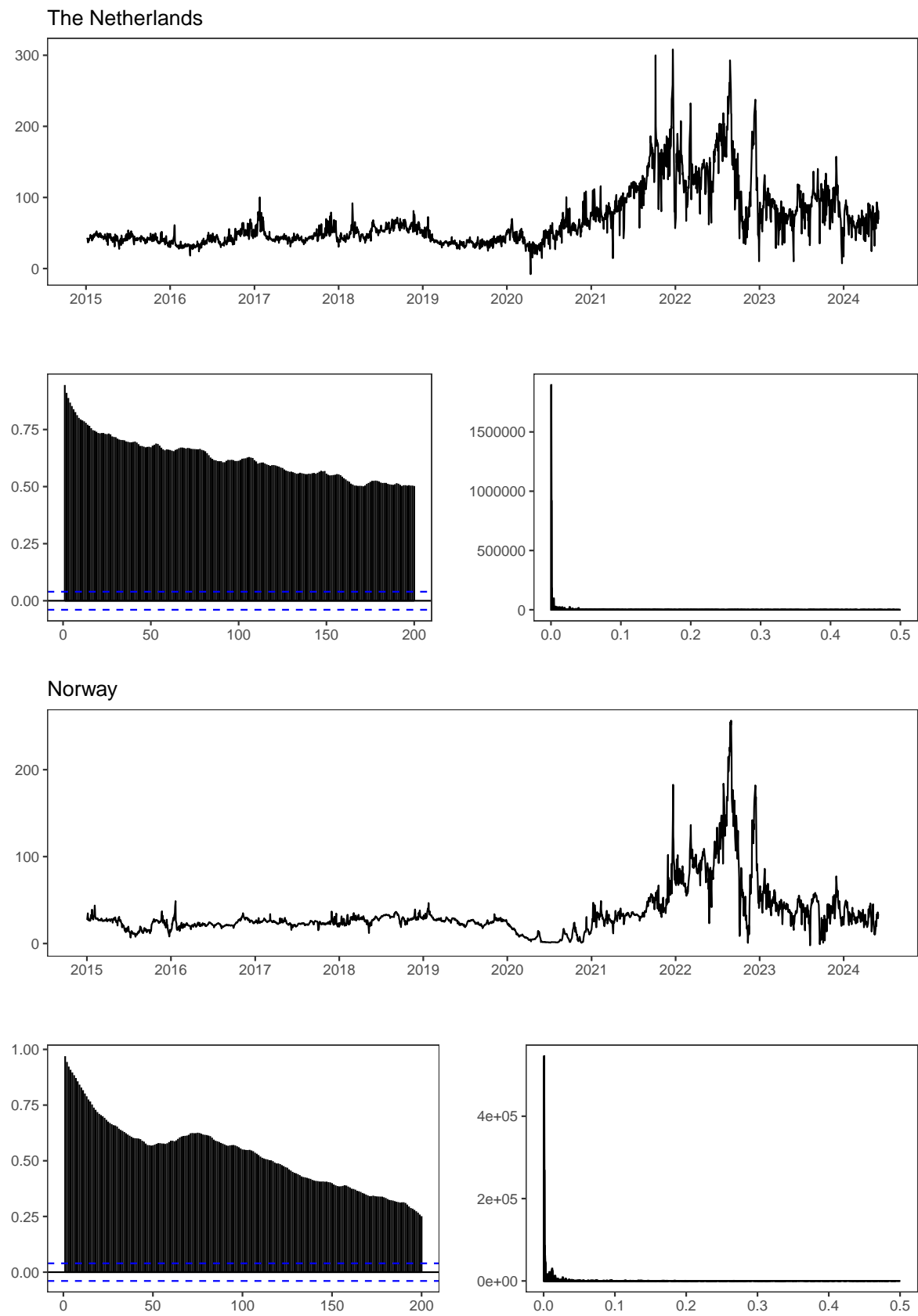


Figure 5: Price series, ACF and periodogram for The Netherlands and Norway.

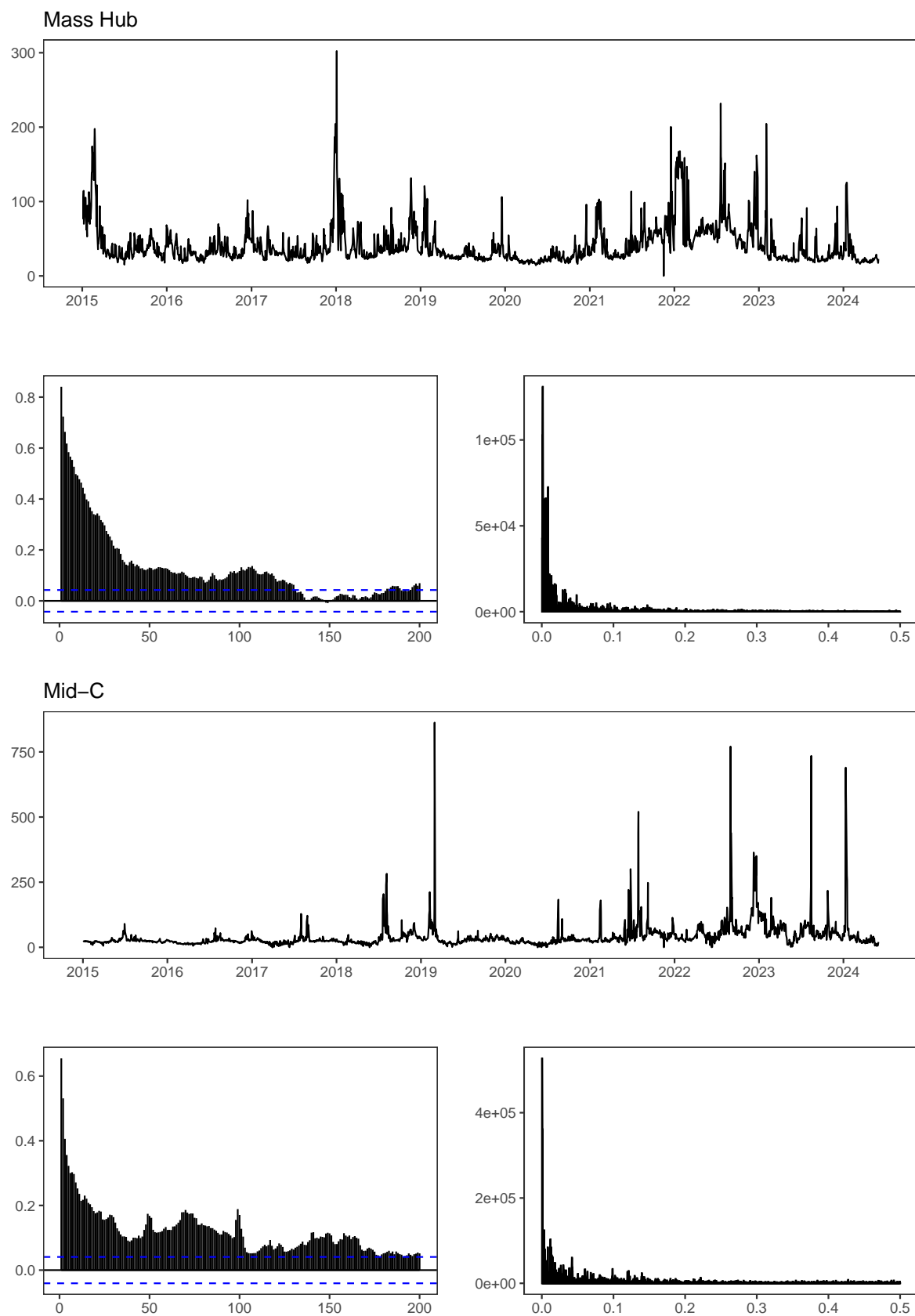


Figure 6: Price series, ACF and periodogram for Mass Hub and Mid-C.

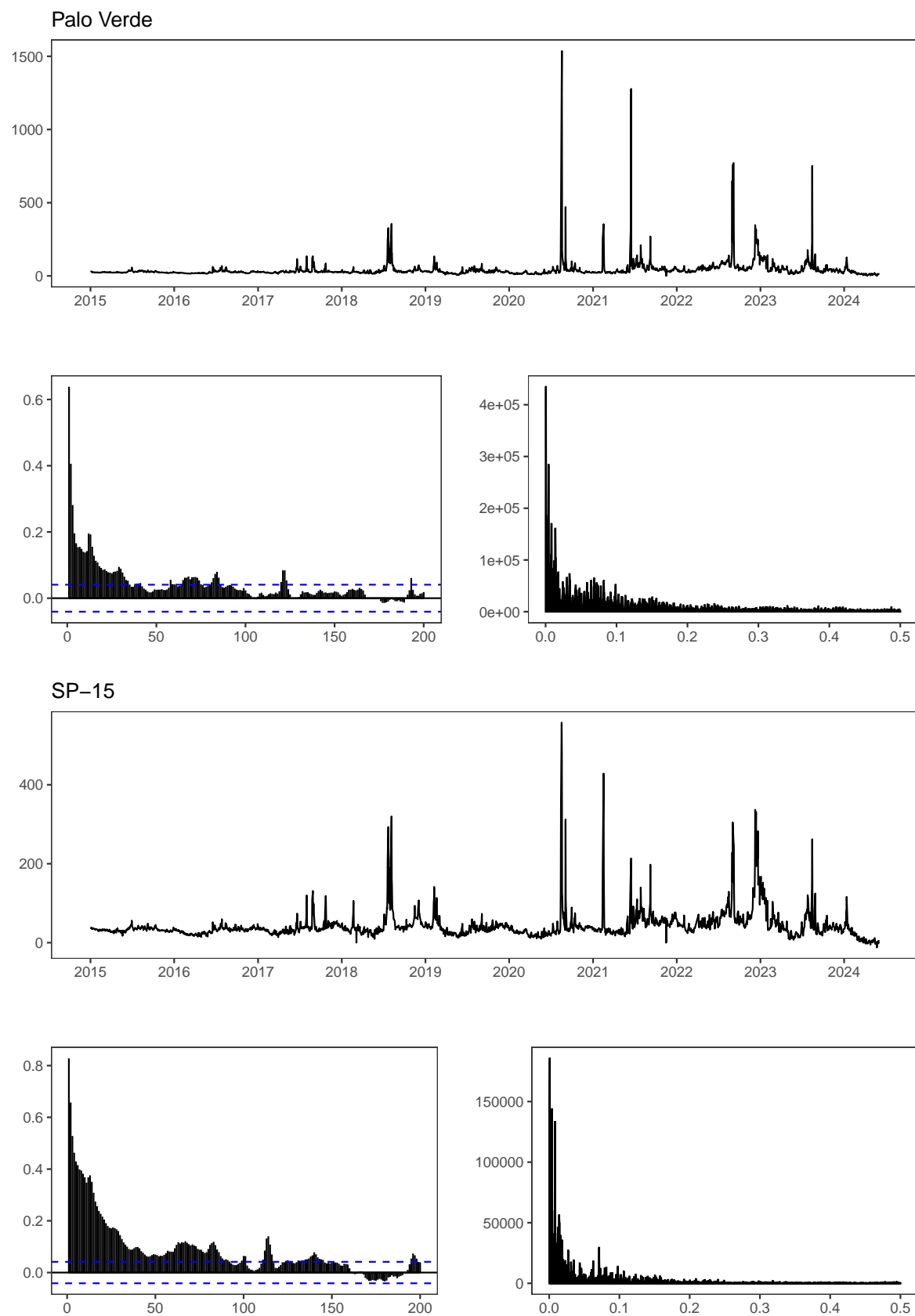


Figure 7: Price series, ACF and periodogram for Palo Verde and SP-15.

# References

- Abadir, K. M., W. Distaso, and L. Giraitis (2009). Two estimators of the long-run variance: beyond short memory. *Journal of Econometrics* 150(1), 56–70.
- Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 817–858.
- Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.
- Chambers, M. J. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review*, 1053–1072.
- Christoffersen, P. F. and F. X. Diebold (1997). Optimal prediction under asymmetric loss. *Econometric theory* 13(6), 808–817.
- Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of econometrics* 105(1), 131–159.
- Dittmann, I. and C. W. Granger (2002). Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics* 110(2), 113–133.
- Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies* 76(2), 669–705.
- Granger, C. W. and N. Hyung (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.
- Haldrup, N. and M. Ø. Nielsen (2006). A regime switching long memory model for electricity prices. *Journal of econometrics* 135(1-2), 349–376.



- Horváth, L. and P. Kokoszka (1997). The effect of long-range dependence on change-point estimators. *Journal of Statistical Planning and Inference* 64(1), 57–81.
- Kruse, R., C. Leschinski, and M. Will (2018). Comparing predictive accuracy under long memory, with an application to volatility forecasting. *Journal of Financial Econometrics* 17(2), 180–228.
- Künsch, H.-R. (1987). Statistical aspects of self-similar processes. *Proceedings of the First World Congress of the Bernoulli Society* 1, 67–74.
- Leschinski, C. (2017). On the memory of products of long range dependent time series. *Economics Letters* 153, 72–76.
- Mikosch, T. and C. Stărică (2004). Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *Review of Economics and Statistics* 86(1), 378–390.
- Paza Mboya, M. and P. Sibbertsen (2023). Optimal forecasts in the presence of discrete structural breaks under long memory. *Journal of Forecasting* 42(7), 1889–1908.
- Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics* 1(2), 278–352.
- Perron, P. and Y. Yamamoto (2021). Testing for changes in forecasting performance. *Journal of Business & Economic Statistics* 39(1), 148–165.
- Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, 1630–1661.
- Robinson, P. M. (2005). Robust covariance matrix estimation: Hac estimates with long memory/antipersistence correction. *Econometric Theory* 21(1), 171–180.
- Rossi, B. and A. Inoue (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics* 30(3), 432–453.

Wang, L. (2008). Change-in-mean problem for long memory time series models with applications. *Journal of Statistical Computation and Simulation* 78(7), 653–668.