On Long Memory Origins and Forecast Horizons

Vera-Valdés, J. Eduardo

Published in:
Journal of Forecasting

DOI (link to publication from Publisher):
10.1002/for.2651

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Vera-Valdés, J. E. (2020). On Long Memory Origins and Forecast Horizons. *Journal of Forecasting*, *39*(5), 811-826. https://doi.org/10.1002/for.2651

On Long Memory Origins and Forecast Horizons

J. Eduardo Vera-Valdés*

Department of Mathematical Sciences, Aalborg University, and CREATES.

Abstract

Most long memory forecasting studies assume that long memory is generated by the fractional difference operator. We argue that the most cited theoretical arguments for the presence of long memory do not imply the fractional difference operator and assess the performance of the autoregressive fractionally integrated moving average ($ARFIMA$) model when forecasting series with long memory generated by nonfractional models. We find that $ARFIMA$ models dominate in forecast performance regardless of the long memory generation mechanism and forecast horizon. Nonetheless, forecasting uncertainty at the shortest forecast horizon could make short memory models provide suitable forecast performance, particularly for smaller degrees of memory. Additionally, we analyse the forecasting performance of the heterogeneous autoregressive (HAR) model, which imposes restrictions on high-order AR models. We find that the structure imposed by the HAR model produces better short and medium horizon forecasts than unconstrained AR models of the same order. Our results have implications for, among others, Climate Econometrics and Financial Econometrics models dealing with long memory series at different forecast horizons.

JEL classification: C53, C22.

Keywords: forecasting, $ARFIMA$, long memory, model confidence set, HAR model.

*E-mail: eduardo@math.aau.dk Webpage: https://sites.google.com/view/veravaldes/
Address: Skjernvej 4A, 9220 Aalborg Øst, Denmark.

1. Introduction

Long memory analysis deals with the notion of series with long-lasting correlations; that is, series with autocorrelations that decay at a hyperbolic rate instead of the standard geometric one. One of the first works on long memory is due to Hurst (1956). He studied the long term capacity of reservoirs for the Nile and recommended to increase the height of a dam to be built given his observations on cycles of highs at the river. As found by Hurst, failing to account for the presence of long memory can lead to inaccurate forecasts. If the data is best modelled by a long memory process, then forecasts computed with standard models would be too optimistic, in the sense that they would predict a return to normal events faster than what we would observe in reality. A dam built based on a short memory forecast would be more prone to overflow that one built based on a long memory forecast, hence increasing the risk of a catastrophic event. Hurst's work highlights the importance of developing appropriate forecasting tools to deal with the presence of long memory.

In the time series literature, the $ARFIMA$ class of models remains to be the most popular given its appeal of bridging the gap between the stationary $ARMA$ models, and the nonstationary $ARIMA$ model by the use of the fractional difference operator. Moreover, some effort has been directed to assess the performance of the $ARFIMA$ type of models when forecasting fractional long memory processes. Nonetheless, not consensus has been formed.

Ray (1993) calculates the percentage increase in mean squared error (MSE) from forecasting fractionally integrated, $FI(d)$, series with AR models. She argues that the MSE may not increase significantly, particularly when we do not know the true long memory parameter, d. Crato and Ray (1996) compare the forecasting performance of $ARFIMA$ models against $ARMA$ alternatives and find that $ARFIMA$ models are in general outperformed by $ARMA$ alternatives for short forecast horizons. On real data, Martens et al. (2009) show that for daily realized volatility for forecast horizons of up to twenty days, it seems to be beneficial to use a flexible high-order AR model instead of a parsimonious but stringent fractionally integrated model.

On the other hand, Barkoulas and Baum (1997) find improvements in forecasting accuracy when fitting $ARFIMA$ models to Eurocurrency returns series, particularly for longer horizons. By allowing for larger data sets of both financial and macro variables, and considering larger forecast horizons, Bhardwaj and Swanson (2006) find that $ARFIMA$ processes generally outperform $ARMA$ alternatives in terms of forecasting performance.

One thing that most forecasting comparison studies have in common is the underlying assumption that long memory is generated by the fractional difference operator. There are two predominant theoretical explanations for the presence of long memory in the time series literature: cross-sectional aggregation of dynamic, persistent micro-units (Granger, 1980); and shocks of random duration (Parke, 1999). As argued in Section 2, neither of these sources of long memory implies an $ARFIMA$ specification and thus do not follow from the fractional difference operator. The question addressed in this paper is if an $ARFIMA$ model serves as a good approximation for forecasting purposes when the long memory generating mechanism is different from the fractional difference operator.

Moreover, as argued by Baillie et al. (2012), a practitioner's goals will generally include making forecasts over both short and long horizons. As an example, the surge of Climate Econometrics as a way to address Climate Change relies on the construction of long horizon forecasts. Thus, we analyse the forecasting performance of short and long memory models at several forecast horizons. In this sense, we extend previous studies to larger forecast horizons relevant to Climate Change analysis. We find that $ARFIMA$ models achieve better forecasting performance than short memory alternatives for all long memory generating mechanisms and forecast horizons. Moreover, the superior performance of the $ARFIMA$ model gets exacerbated the higher the degree of memory of the processes. Nonetheless, forecast uncertainty at short horizons is such that short memory models achieve comparable forecast performance as $ARFIMA$ models, particularly for smaller degrees of memory. Finally, we find that the restrictions imposed by the HAR model produce better short and medium forecasting performance than unconstrained AR alternatives.

This paper proceeds as follows. In Section 2, we present the long memory generating processes considered, and show that the most cited theoretical explanations for the presence of long memory do not imply an $ARFIMA$ specification. Section 3 describes the design of the Monte Carlo analysis used for the forecasting study, while Section 4 presents the main results. Furthermore, Section 5 discusses the results from the forecasting exercise in a bias-variance trade-off context. Sections 6 and 7 show that the insights gained from the Monte Carlo simulations hold on real data and address some practical considerations. Finally, Section 8 presents the conclusions.

2. Long Memory Generating Processes

In this section, we present the long memory generating processes considered in this work. All processes studied are long memory in the covariance sense formalised below.

Definition. *Let x_t be a second-order stationary process with autocovariance function $\gamma_x(k)$. Then x_t is said to exhibit long memory in the covariance sense if*

$$\gamma_x(k) \approx C_1 k^{2d-1} \text{ as } k \to \infty, \tag{1}$$

with $d \in (0, 1/2)$, and C_1 a constant. Above, for $h(x) \neq 0$, $g(x) \approx h(x)$ as $x \to \infty$ denotes that $g(x)/h(x)$ converges to 1 as x tends to ∞.

Note from the definition that long memory in the covariance sense relates to the rate of decay of the autocorrelations, see Haldrup and Vera-Valdés (2017) for a discussion on other definitions. For applied purposes, the fitted models try to mimic the rate of decay of the autocorrelations to assess the importance that past observations have on future realizations. In this context, the models use this information to produce better forecasts. Thus, we deem the covariance sense to the appropriate definition of long memory for this work.

2.1. Fractional Difference Operator

We include fractionally integrated processes in the analysis as a benchmark. Granger and Joyeux (1980), and Hosking (1981) proposed to use the standard binomial expansion to decompose the fractional difference operator, $(1-L)^d$, in a series with coefficients $\pi_j = \Gamma(j+d)/[\Gamma(d)\Gamma(j+1)]$ for $j \in \mathbb{N}$. That is, they propose to study a series given by

$$(1-L)^d x_t = \epsilon_t, \tag{2}$$

where ϵ_t is a white noise process, $d \in (-1/2, 1/2)$, and L is the lag operator.

For $d \in (0, 1/2)$, it can be shown that these coefficients decay at a hyperbolic rate, which in turn translates to hyperbolic decaying autocorrelations. Hence, fractionally differenced processes have long memory, Equation (1).

The case $d \in (-1/2, 0)$, typically denoted antipersistent in the literature, while also producing hyperbolic decaying autocorrelations, the autocorrelations are all negative by construction. Moreover, as argued in Section 6, the spectrum of antipersistent processes collapses to zero at the origin and thus have different behaviour than for the positive case and other long memory generating mechanisms. Hence, we will focus on the positive range for the rest of the analysis.

Using the fractional difference operator, Granger and Joyeux (1980), and Hosking (1981) extended the $ARMA$ model to include long memory dynamics resulting in the $ARFIMA$ model. An $ARFIMA$ process is given by

$$\phi(L)(1-L)^d x_t = \theta(L)\epsilon_t, \tag{3}$$

where ϵ_t is a white noise process, d is defined as before, and $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator with no common roots, all outside the unit circle.

The properties of the $ARFIMA$ model have been well documented in, among others, Baillie (1996), and Beran et al. (2013). Furthermore, $ARFIMA$ processes are relatively easy to estimate by maximum likelihood, see Sowell (1992). Thus, the $ARFIMA$ model has become the canonical construction for modelling and forecasting long memory processes in the time series literature.

For simulation purposes, we use the algorithm developed by Jensen and Nielsen (2014) based on the fast Fourier transform to generate fractionally differenced series.

2.2. Cross-Sectional Aggregation

Granger (1980), in line with the work of Robinson (1978) on autoregressive processes with random coefficients, showed that aggregating $AR(1)$ processes with coefficients sampled from a Beta distribution can produce long memory. He considered N random coefficient AR processes generated as

$$x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t} \tag{4}$$

where $\varepsilon_{i,t}$ is a white noise process with $E[\varepsilon_{i,t}^2] = \sigma_\varepsilon^2 \ \forall i \in \{1, 2, \cdots, N\}$, $\forall t \in \mathbb{Z}$, and α_i^2 is sampled from a Beta distribution, $\mathcal{B}(\alpha; p, q)$, with density given by

$$\mathcal{B}(\alpha; p, q) = \frac{1}{B(p,q)} \alpha^{p-1}(1-\alpha)^{q-1} \quad \text{for} \quad \alpha \in (0,1),$$

where $p, q > 1$, and $B(\cdot, \cdot)$ is the Beta function. Furthermore, define the cross-sectional aggregated series as

$$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}. \tag{5}$$

Granger showed that as $N \to \infty$, the autocorrelations of x_t in (5) decay at a hyperbolic rate, Equation (1). Thus, x_t has long memory in the covariance sense.

We denote $x_t \sim CSA(p, q; \sigma^2)$ to a series generated by cross-sectional aggregation of autoregressive parameters sampled from the Beta distribution, $B(p, q)$, and where the variance of the noise terms is σ^2. The notation makes explicit the origin of the memory by cross-sectional aggregation and its dependence on the parameters of the Beta distribution and the variance of the noise.

The cross-sectional aggregation result has been extended in several directions, including to allow for general $ARMA$ processes, as well as to other distributions. See for instance, Linden (1999), Oppenheim and Viano (2004), and Zaffaroni (2004). As argued by Haldrup and Vera-Valdés (2017), maintaining the Beta distribution allows us to have closed-form representations.

In applied work, cross-sectional aggregation has been cited as the source of long memory for many time series. For example, it has been argued for inflation, output, and volatility; see Balcilar (2004), Diebold and Rudebusch (1989), and Altissimo et al. (2009).

Haldrup and Vera-Valdés (2017) showed that the long memory generated by cross-sectional aggregation does not correspond to the one associated with the $ARFIMA$ model. In particular, the authors showed that although the long memory by cross-sectional aggregation can be removed by fractional differencing, the resulting series does not belong to the class of linear $ARMA$ processes. The question addressed in this paper is whether an $ARFIMA$ specification remains useful for forecasting purposes.

For simulation purposes, we use the algorithm developed by Vera-Valdés (2019) based on the fast Fourier transform. The algorithm is exact in the sense that it generates the limiting process in (5), and thus, no approximation is needed. Moreover, we follow the author's recommendation to choose the first parameter of the Beta distribution to mimic a fractionally differenced process and make fair comparisons between long memory generating processes.

2.3. Error Duration Model

Parke (1999) introduced the error duration model as one possible source of the long memory in the data. He showed that if a series is the result of the sum of shocks of stochastic duration, then it would exhibit hyperbolic decaying autocorrelations.

Let ε_s be a series of independent and identically distributed shocks with mean zero and finite variance σ^2. Assume that the shock ε_i has a stochastic duration of $n_i \geq 0$ time periods, surviving from period i until period $i + n_i$. Let p_k be the probability that event ε_i survives until period $i + k$, and take $g_{i,t}$ to be the indicator function for the event that the error ε_i survives until period t. Furthermore, define x_t as

$$x_t = \sum_{s=-\infty}^{t} g_{s,t}\varepsilon_s. \qquad (6)$$

By cleverly selecting the error survival probabilities, Parke showed that the autocorrelation function of x_t in (6) will decay at a rate similar to $FI(d)$ processes. That is, if $p_k \approx k^{-2+2d}$ as $k \to \infty$, x_t will have hyperbolic decaying autocorrelations and thus long memory in the covariance sense, Equation (1).

We denote $x_t \sim EDM(d;\sigma^2)$ to a series generated by the error duration model with $p_k \approx k^{-2+2d}$, and where the variance of the noise terms is σ^2. The notation makes explicit the origin of the memory by the error duration model and its dependence on the error survival probabilities and the variance of the noise.

The error duration model has been suggested as the source of long memory for realized volatility, Corsi et al. (2008), temperature, Gil-Alana (2005), gender gaps, Box-Steffensmeier et al. (2004), to name a few.

Even though the autocorrelation function of a series generated by the error duration model shows similar dynamics as the one due to the fractional difference operator, note that the error duration model does not imply the latter. In the error duration model, previous errors have a random probability of disappearing from one period to the next, in contrast to the fractional difference operator where they are always present. This work will analyse the forecasting capabilities of the $ARFIMA$ model when dealing with long memory processes generated by the error duration model.

For simulation purposes, we follow the algorithm from Parke (1999). Furthermore, we follow Parke's recommendation and consider error survival probabilities that mimic those of the $FI(d)$ model to make equitable comparisons between the long memory generating processes.

2.4. Long Memory Models

Table 1 summarizes the long memory generating mechanisms to be analysed.

Table 1: Long Memory Generating Processes

Fractional Difference Operator $FI(d;\sigma^2)$	$(1-L)^d x_t = \varepsilon_t$ $(1-L)^d = \sum_{s=0}^{\infty} \frac{\Gamma(s-d)}{\Gamma(-d)\Gamma(s+1)} L^s$
Cross-Sectional Aggregation $CSA(p,q;\sigma^2)$	$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}$ $x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t}$ $\alpha_i \sim \mathcal{B}(\alpha; p, q);\ p, q > 1$
Error Duration Model $EDM(d;\sigma^2)$	$x_t = \sum_{s=-\infty}^{t} g_{s,t}\varepsilon_s$ $g_{s,s+k} = \begin{cases} 0 & \text{with probability } 1-p_k \\ 1 & \text{with probability } p_k \end{cases}$ $p_k = k^{2d-2}$
All	$\varepsilon_t \sim \mathcal{N}(0,\sigma^2), \forall t$ $\sigma^2 = 1$

3. Monte Carlo Design

In this section, we describe the Monte Carlo analysis designed to compare the forecasting performance of $ARFIMA$ models against $ARMA$, and high-order AR models on long memory series generated by the processes described in Section 2.

3.1. Model Selection

This section presents the models considered for the forecasting analysis. Model selection was based on two criteria.

As a first criterion, we use the Bayesian Information Criterion, BIC, to select the number of lags to include in both the $ARFIMA$ and $ARMA$ models in an independent Monte Carlo analysis. The usefulness of the BIC for the class of processes with fractional differencing was proven by Beran et al. (1998). The authors show that for this class of processes, the penalty term must tend to infinity simultaneously with the sample size; thus, the Akaike Information Criterion is not consistent while the BIC is. Note that we made the lag selection exercise independent from the forecasting analysis to avoid the multiple testing problem.

We allow for a maximum of two lags at both components of the $ARFIMA$ model, while the maximum was set to four for the $ARMA$ model. We estimate both types of models using maximum likelihood.

Results from the lag selection exercise are presented in Figures 1 and 2.

Figure 1 shows that few lags, if any, are selected for the $ARFIMA$ specification for either component for both long memory generating mechanisms. In particular, the pure fractional difference model, $FI(d)$, is the preferred model in most cases. Moreover, the number of lags selected does not seem to be overly affected by the long memory parameter.

For the $ARMA$ specification, Figure 2 shows that, perhaps not surprisingly, more lags are selected. This of course can be due to the fact that we are not controlling for the long memory behaviour by the inclusion of the long memory parameter, d. Moreover, the number of lags selected seems to be affected by the long memory parameter. The BIC selects more lags the larger the long memory parameter is. Nonetheless, the maximum number of lags selected by the BIC is two for all cases considered.

Figure 1: Proportion of times the Bayesian Information Criterion, BIC, selects each long memory model at different degrees of memory for long memory generated by cross-sectional aggregation and the error duration model. Sample size and number of replications are set to 1,000.

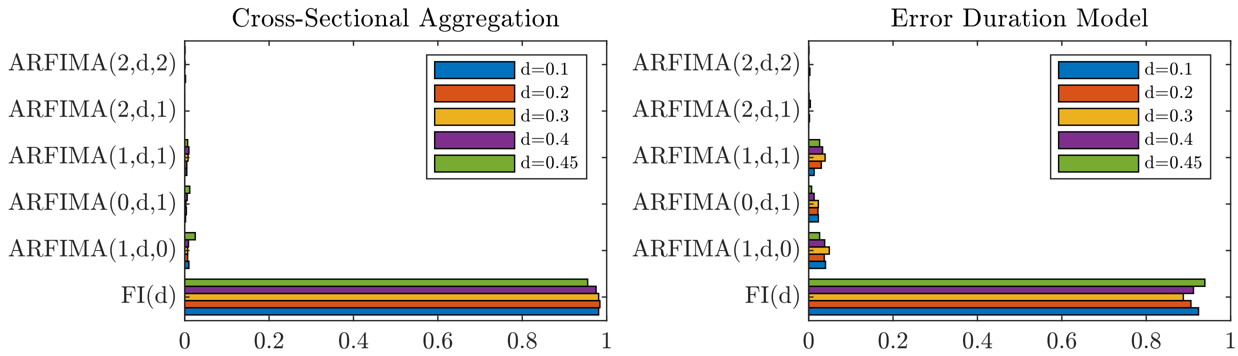
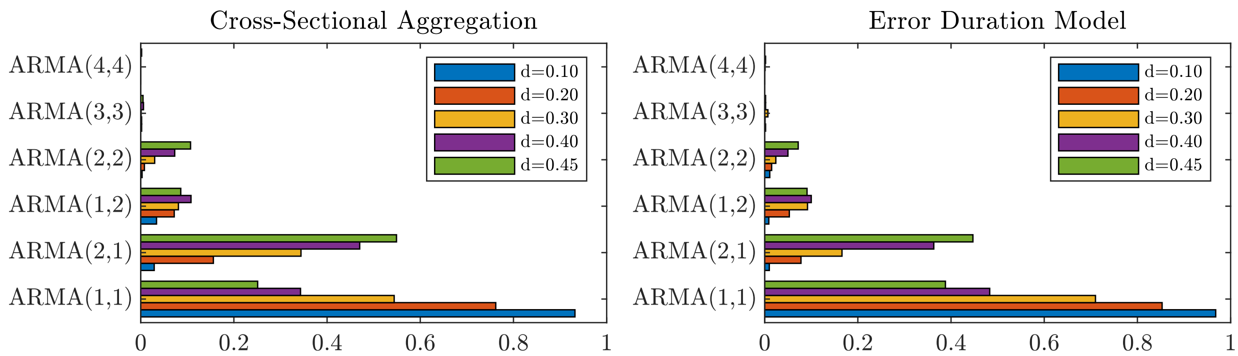


Figure 2: Proportion of times the Bayesian Information Criterion, BIC, selects each short memory model at different degrees of memory for long memory generated by cross-sectional aggregation and the error duration model. Sample size and number of replications are set to 1,000.



As a second criterion, we follow previous works on long memory forecasting exercises and consider high-order AR processes, $AR(22)$ and $AR(50)$. The intuition behind the inclusion of high-order AR processes in the forecasting exercise follows the standard bias-variance trade-off argument. High-order AR processes are more flexible given that they are free to select each coefficient in the autoregressive representation independently, while the fractional difference operator imposes a general structure to them. This enhanced flexibility may allow the high-order AR processes to be closer to the realized future values at the expense of greater variability, resulting in the standard trade-off.

Moreover, we include the $HAR(3)$ model of Corsi (2009) give its ability to mimic long memory behaviour, see for instance Andersen et al. (2007) and Chiriac and Voev (2011). The $HAR(3)$ model is a constrained $AR(22)$ given by

$$x_t = a_0 + a_1 x_{t-1}^{(f)} + a_2 x_{t-1}^{(w)} + a_3 x_{t-1}^{(m)} + \epsilon_t, \tag{7}$$

where $x_{t-1}^{(f)} = x_{t-1}$, $x_{t-1}^{(w)} = \frac{1}{5}\sum_{i=1}^{5} x_{t-i}$ and, $x_{t-1}^{(m)} = \frac{1}{22}\sum_{i=1}^{22} x_{t-i}$. Following the bias-variance trade-off discussion above, note that the $HAR(3)$ model imposes a compromise between the highly flexible high-order AR processes and the constrained fractional differenced ones.

The HAR specification has been used to model financial data. In that context, the model reflects the fact that different agents respond to uncertainty at distinct horizons. The three components of the model seek to capture the daily ($x_t^{(f)}$), weekly ($x_t^{(w)}$), and monthly ($x_t^{(m)}$) levels of uncertainty.

Note that including the $HAR(3)$ model allows us to extend Corsi's (2009) results in several directions. We make comparisons against a larger set of models, we include larger forecast horizons, and we remove the uncertainty regarding the presence of long memory in the data by comparing the performance of the HAR model in simulated long memory series, whereas Corsi used real data.

Following the two criteria, Table 2 presents the models considered for the forecasting exercise.

Table 2: Competing Models

$FI(d)$	$ARMA(1,1)$	$AR(22)$
$ARFIMA(1,d,0)$	$ARMA(2,1)$	$HAR(3)$
$ARFIMA(0,d,1)$	$ARMA(1,2)$	$AR(50)$
$ARFIMA(1,d,1)$	$ARMA(2,2)$	$I(1)$

3.2. Forecast Evaluation

For each long memory generating process, we fit all the competing models for a sample size T. The models are indexed by $i \in \{1, 2, \ldots, m\}$, and the out-of-sample forecast from model i is denoted by $\hat{y}_{T+k}^{i}$, $\forall k \in \{1, \ldots, h\}$.

Given the rise of Climate Econometrics studies keen on producing long horizon forecasts, we consider it relevant to evaluate forecast performances to horizons as far as $h = 100$ periods ahead. The forecast horizons considered are thus $h \in \{1, 5, 10, 30, 50, 100\}$.

As measures of forecast performance, we compute the out-of-sample mean absolute deviation (MAD), and the out-of-sample root mean square error ($RMSE$) given by

$$MAD_h^i = \frac{1}{h}\sum_{k=1}^{h}\left|y_{T+k} - \hat{y}_{T+k}^i\right| \quad RMSE_h^i = \left(\frac{1}{h}\sum_{k=1}^{h}\left(y_{T+k} - \hat{y}_{T+k}^i\right)^2\right)^{1/2},$$

where h and $\hat{y}_s^i$ are defined as above. We report the mean of both MAD and $RMSE$ across all replications.

3.3. Monte Carlo Design

All models were estimated by maximum likelihood (MLE) following the work of Baillie et al. (2012). The authors find that long memory forecasts based on MLE are superior to the ones obtained from local Whittle estimators. Moreover, throughout, we use a large sample size of $T = 1,000$ to reduce the estimation error.

Table 3 presents the Monte Carlo design for the forecasting analysis.

Table 3: Monte Carlo Design

• Generate series of size $T + h_{max}$, where h_{max} is the largest forecast horizon, using the long memory generating processes in Section 2, Table 1.
• Fit by maximum likelihood the competing models shown in Table 2, for a subsample of size T.
• Construct forecasts from each model for horizons $h \in \{1, 5, 10, 30, 50, 100\}$.
• Compute the MAD and $RMSE$.
• Repeat the steps above R times, the number of replications.
• After the R replications, report different statistics from the values of MAD and $RMSE$ obtained for each forecast horizon.

4. Monte Carlo Results

In this section, we present the results from the Monte Carlo simulations. Throughout, for reasons of space, we focus on the MAD loss function given that it is less sensitive to large mispredictions. Nonetheless, results from the $RMSE$ loss function are similar. Moreover, we will show results for a subset of the long memory parameters considered, additional figures for results with other long memory parameters are available upon request.

4.1. Fractional Difference Operator

As a benchmark, we present in Table 4 the results from the Monte Carlo analysis for long memory generated using the fractional difference operator with long memory parameter $d = 0.4$. The table presents the results normalized by the performance of the pure fractional difference model, $FI(d)$, to ease interpretation. Thus, numbers bigger/smaller than 1 represent that the given model shows inferior/superior performance than the $FI(d)$ model for the given loss function and forecast horizon.

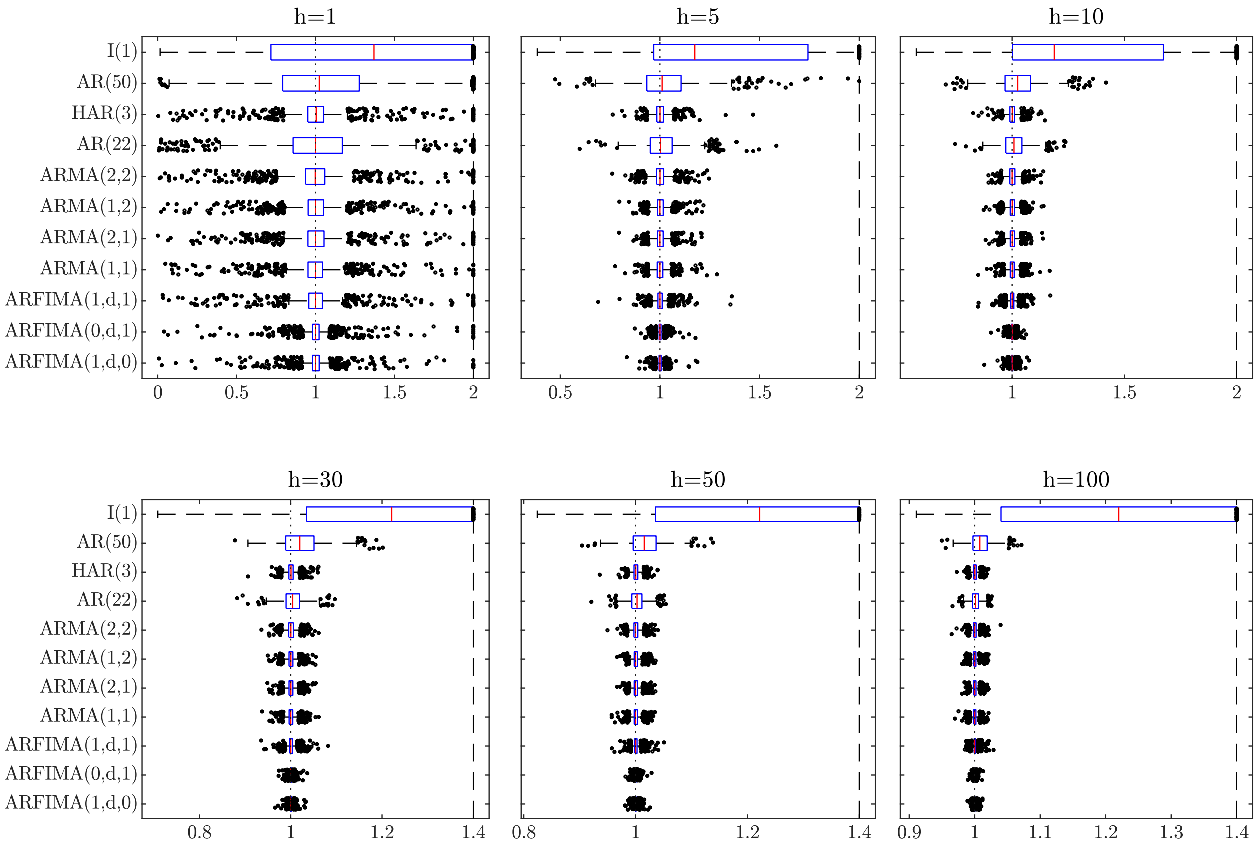
$FI(d;\sigma^2)$	h=1		5		10		30		50		100	
$d = 0.4$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$
$FI(d)$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
$ARFIMA(1,d,0)$	1.001	1.001	1.001	1.001	**1.000**	**1.000**	**1.000**	**1.000**	1.001	**1.000**	**1.000**	**1.000**
$ARFIMA(0,d,1)$	1.001	1.001	1.001	1.001	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
$ARFIMA(1,d,1)$	1.004	1.004	1.002	1.002	1.001	1.001	1.002	1.002	1.003	1.003	1.002	1.002
$ARMA(1,1)$	1.013	1.013	1.021	1.017	1.023	1.020	1.036	1.030	1.037	1.032	1.030	1.026
$ARMA(2,1)$	1.009	1.009	1.010	1.008	1.010	1.008	1.019	1.016	1.023	1.020	1.022	1.019
$ARMA(1,2)$	1.010	1.010	1.014	1.011	1.014	1.012	1.024	1.020	1.027	1.024	1.024	1.021
$ARMA(2,2)$	1.010	1.010	1.011	1.008	1.011	1.009	1.021	1.018	1.025	1.022	1.022	1.019
$AR(22)$	1.016	1.016	1.014	1.012	1.012	1.011	1.014	1.013	1.018	1.016	1.018	1.016
$HAR(3)$	1.009	1.009	1.008	1.006	1.008	1.007	1.012	1.010	1.016	1.014	1.017	1.015
$AR(50)$	1.025	1.025	1.024	1.025	1.023	1.023	1.027	1.026	1.026	1.025	1.020	1.018
$I(1)$	1.193	1.193	1.244	1.216	1.265	1.231	1.308	1.260	1.326	1.276	1.344	1.291

Table 4: Mean of the MAD and $RMSE$ loss functions relative to the $FI(d)$ model when forecasting long memory generated by the fractional difference operator with long memory parameter $d = 0.4$ and all forecast horizons.

Note that the $ARFIMA$ class of models is the preferred specification for all forecast horizons, which is not surprising given that the series are $ARFIMA$ processes. Nonetheless, we find that short memory models perform relatively well when forecasting a fractionally differenced series at short forecast horizons. In particular, the $ARMA(2,1)$ model achieves satisfactory performance for short forecast horizons. Thus, a practitioner could satisfactorily use short memory models to make short horizon forecasts of long memory processes. However, note that the performance of short memory models declines as the forecast horizon increases. Thus, the table shows that long memory dynamics should be included in the specification if the goal of the analysis is to make forecasts at medium or large horizons.

Moreover, it is interesting to contrast the performance of the $HAR(3)$ model against the unconstrained $AR(22)$ in the context of the bias-variance trade-off. The table shows that the added restrictions imposed by the $HAR(3)$ model seem to increase its forecasting performance at short and medium horizons. At larger horizons, the added restrictions seem to lose its added performance.

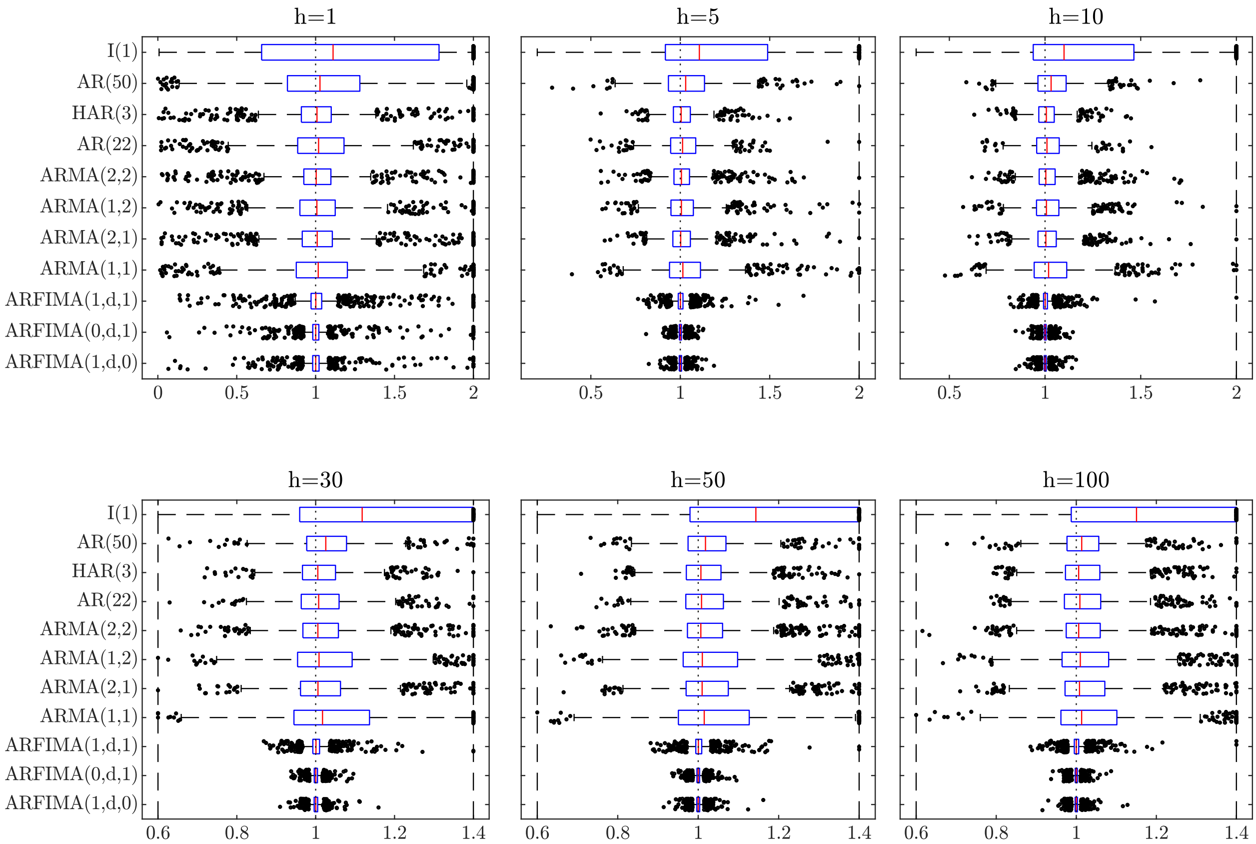
Figure 3: Boxplots for the MAD loss measure for competing models relative to the $FI(d)$ model when forecasting a long memory series generated by the fractional difference operator with long memory parameter $d = 0.1$ for all forecast horizons.



To get a better understanding of the bias-variance trade-off across models, Figures 3 and 4 present boxplots[1] for the MAD loss functions achieved for all models and for all forecast

[1] On each box, the central mark indicates the median, and the left and right edges of the box indicate the 25th, q_1, and 75th, q_3, percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually. Outliers are defined if they are greater than $q_3 + 1.5(q_3 - q_1)$ or less than $q_1 - 1.5(q_3 - q_1)$.

Figure 4: Boxplots for the MAD loss measure for competing models relative to the $FI(d)$ model when forecasting a long memory series generated by the fractional difference operator with long memory parameter $d = 0.45$ for all forecast horizons.



horizons. Figure 3 shows the results for long memory parameter $d = 0.1$, while Figure 4 for $d = 0.45$.

The boxplots show the uncertainty regarding the forecasting performance across models. Moreover, the figures allow us to collect three key insights.

First, the boxplots show that the models' forecast uncertainty, measured as the width of the boxes, is quite considerable for small forecast horizons. This suggests that short memory models may provide suitable alternatives for a small forecast horizon, $h = 1$ or $h = 5$, given the bias-variance trade-off. A practitioner may argue that a relatively simpler short memory model achieves similar performance than a stricter long memory model and prefer the former.

Second, the figures show that long memory models start to dominate in forecast performance as the forecast horizon increases. This in the sense that the boxes associated to $ARFIMA$ models and medium and long horizons are thinner than for short memory models. Moreover, looking at the different results between Figure 3 and Figure 4, the superior performance of $ARFIMA$ models is amplified the larger the long memory parameter is.

Finally, the figure allows us to further compare the $HAR(3)$ model against the unconstrained $AR(22)$ model. The figure shows that for short forecast horizons, the $HAR(3)$ model achieves better forecasting performance than the unconstrained $AR(22)$, particularly for the smaller degree of memory, $d = 0.1$. For larger forecast horizons and larger degrees of memory, the added restrictions lose some of their superior forecasting performance.

Overall, Table 4, and Figures 3 and 4 extend the findings of previous studies on forecasting long memory when the long memory is generated by the fractional difference operator. They show that short memory models may be sensible alternatives for short forecast horizons and smaller degrees of memory while extending the analysis to show that $ARFIMA$ models are the superior class of models for medium and large forecast horizons. Moreover, we find that the constraints imposed by the HAR model improve forecasting performance over the unconstrained AR model.

4.2. Cross-Sectional Aggregation

Results from the Monte Carlo analysis for cross-sectional aggregated processes are presented in Table 5, and Figures 5 and 6. The table presents the results for long memory parameter $d = 0.4$ normalized by the performance of the pure fractional difference model, $FI(d)$.

Note that the results for processes generated by cross-sectional aggregation are quite similar to the benchmark case using the fractional difference operator. That is, the $ARFIMA$ class of models give the best performance for all forecast horizons, while short memory models may provide a valid alternative at short forecast horizons. Moreover, the table shows the superior forecasting performance of the $HAR(3)$ model against the unconstrained $AR(22)$.

$CSA(p,q;\sigma^2)$	h=1		5		10		30		50		100	
$d=0.4$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$
$FI(d)$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
$ARFIMA(1,d,0)$	1.001	1.001	1.001	1.001	1.002	1.001	1.001	1.001	1.001	1.001	1.001	1.001
$ARFIMA(0,d,1)$	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001	**1.000**	1.001	1.001
$ARFIMA(1,d,1)$	1.002	1.002	1.001	1.001	1.003	1.002	1.003	1.002	1.002	1.001	1.003	1.002
$ARMA(1,1)$	1.011	1.011	1.023	1.021	1.034	1.028	1.050	1.043	1.047	1.040	1.040	1.036
$ARMA(2,1)$	1.004	1.004	1.007	1.008	1.010	1.010	1.023	1.020	1.025	1.022	1.027	1.024
$ARMA(1,2)$	1.005	1.005	1.012	1.013	1.018	1.016	1.033	1.028	1.034	1.029	1.033	1.029
$ARMA(2,2)$	1.004	1.004	1.008	1.008	1.014	1.013	1.025	1.022	1.025	1.022	1.024	1.021
$AR(22)$	1.012	1.012	1.008	1.010	1.012	1.012	1.016	1.015	1.018	1.015	1.020	1.018
$HAR(3)$	1.004	1.004	1.004	1.006	1.007	1.007	1.013	1.012	1.016	1.014	1.019	1.017
$AR(50)$	1.026	1.026	1.025	1.028	1.027	1.028	1.031	1.029	1.026	1.024	1.020	1.019
$I(1)$	1.128	1.128	1.191	1.170	1.218	1.186	1.265	1.224	1.285	1.241	1.299	1.252

Table 5: Mean of the $RMSE$ and MAD loss functions relative to the $FI(d)$ model when forecasting a long memory series generated by cross-sectional aggregation with parameter $d=0.4$ and all forecast horizons.

Figure 5: Boxplots for the MAD loss measure for competing models relative to the $FI(d)$ model when forecasting a long memory series generated by cross-sectional aggregation with long memory parameter $d=0.1$ for all forecast horizons.

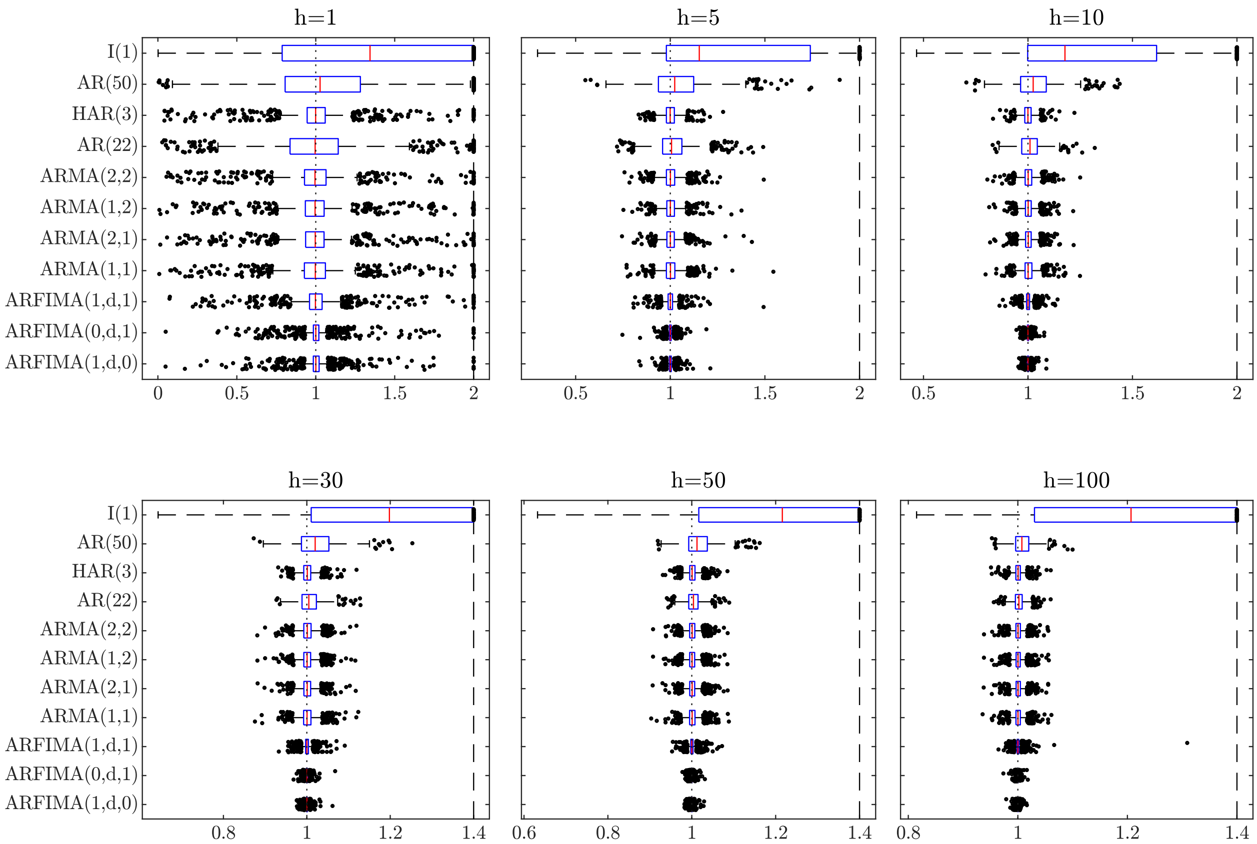
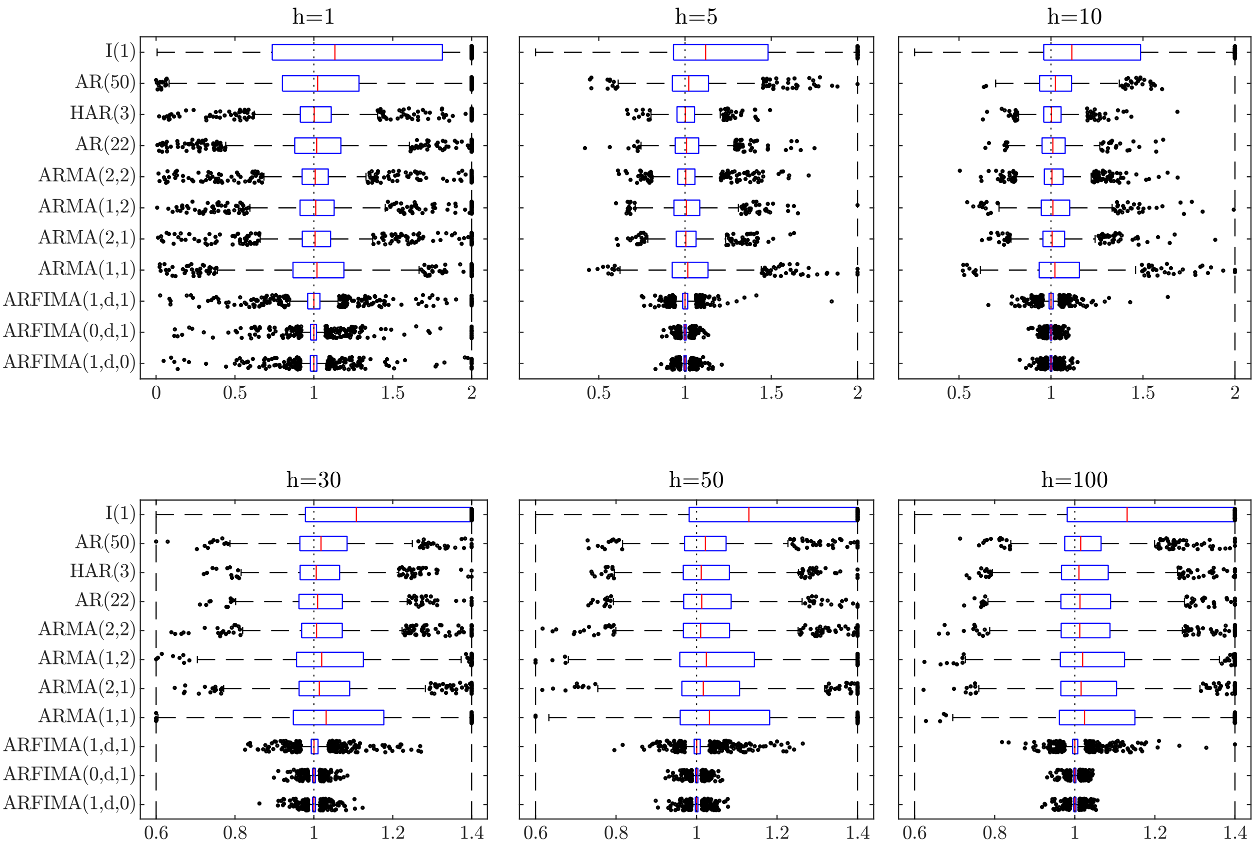


Figure 6: Boxplots for the MAD loss measure for competing models relative to the $FI(d)$ model when forecasting a long memory series generated by cross-sectional aggregation with long memory parameter $d = 0.45$ for all forecast horizons.



We can study the effect that the degree of long memory has on the results in Figures 5 and 6. The figures show boxplots for the Monte Carlo results for all competing models and forecast horizons and two long memory parameters, $d = 0.1$ in Figure 5, and $d = 0.45$ in Figure 6.

The figures extend the findings in Table 5, and are in line with the results for the fractional difference operator case. They show the relatively good performance of short memory models at short forecast horizons, although achieving slightly worse performance than for the fractional difference operator case. For $h = 1$ and $h = 5$, short memory models provide similar results to $ARFIMA$ models, measured by the width of the boxes, particularly for small degrees of memory.

For larger forecast horizons, $ARFIMA$ models are the unsurpassed forecasters for long memory processes generated by cross-sectional aggregation, and their dominance widens as the degree of long memory increases.

Furthermore, the figures show the increase in forecast performance of the $HAR(3)$ model against the unconstrained $AR(22)$ for small and medium forecast horizons. As the forecast horizon increases, the increased performance of the constrained $HAR(3)$ starts to dilute.

Overall, the results from the Monte Carlo analysis for long memory generated by cross-sectional aggregation are in line with the ones due to the fractional difference operator. They show that the $ARFIMA$ class of models are the preferred specification for forecast construction, while short memory models could provide sensible alternatives for smaller forecasting horizons and smaller degrees of memory. Finally, the added restrictions imposed by the $HAR(3)$ model show superior performance than the unconstrained $AR(22)$ alternative, particularly at smaller horizons and smaller degrees of memory.

4.3. Error Duration Model

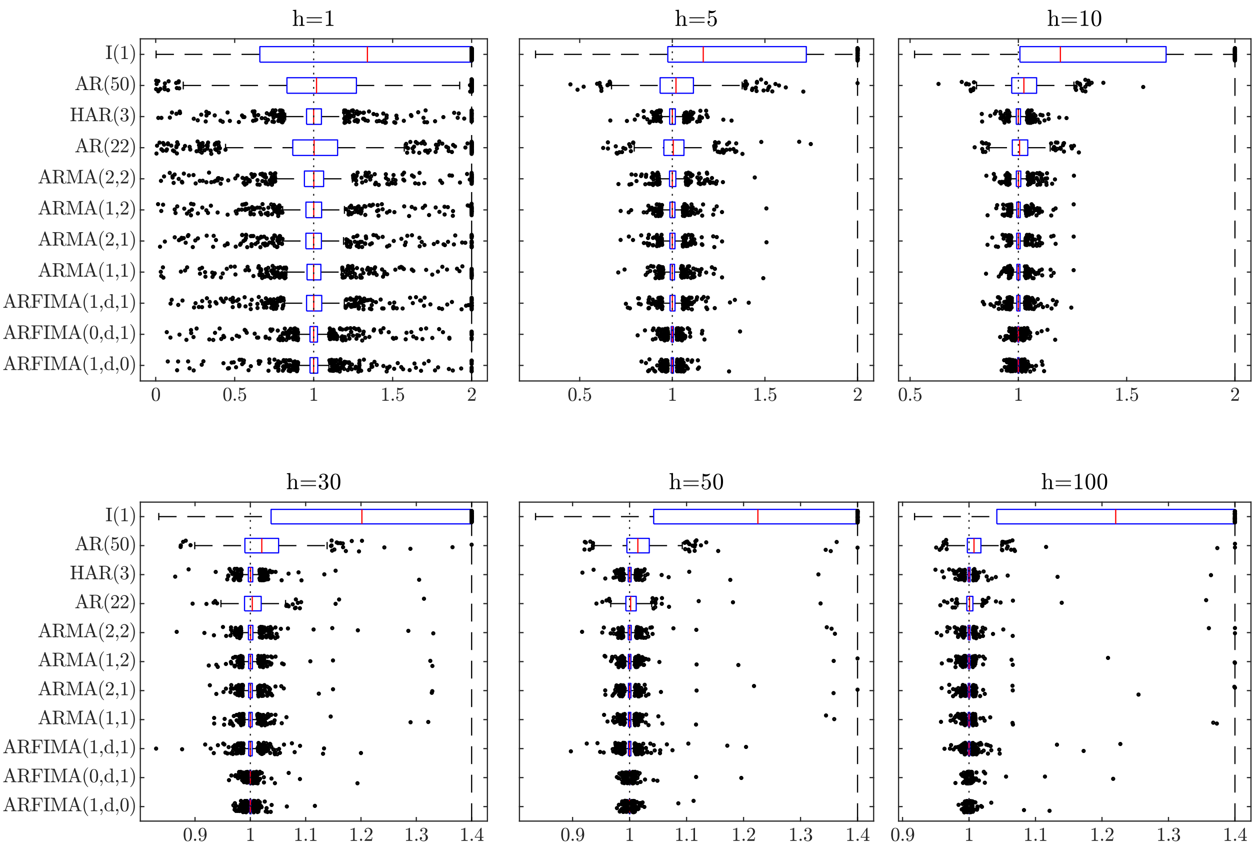
Table 6 and Figures 7 and 8 present the results from the Monte Carlo analysis for long memory processes generated by the error duration model, $EDM(d;\sigma^2)$. The table present the results for long memory parameter $d = 0.4$ normalized by the performance of the $FI(d)$ model. Figure 7 presents boxplots for long memory parameter $d = 0.1$, while Figure 8 presents boxplots for long memory parameter $d = 0.45$.

$EDM(d;\sigma^2)$	h=1		5		10		30		50		100	
$d = 0.4$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$
$FI(d)$	1.000	1.000	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	1.000	1.000	1.000	1.000
$ARFIMA(1,d,0)$	**0.999**	**0.999**	**1.000**	**1.000**	1.001	1.001	**1.000**	**1.000**	1.000	1.000	1.000	1.000
$ARFIMA(0,d,1)$	**0.999**	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	1.000	1.000	1.000	1.000
$ARFIMA(1,d,1)$	1.000	1.000	**1.000**	**1.000**	1.001	1.001	**1.000**	**1.000**	**0.999**	**0.999**	**0.999**	**0.999**
$ARMA(1,1)$	1.011	1.011	1.013	1.011	1.018	1.016	1.024	1.021	1.024	1.022	1.022	1.019
$ARMA(2,1)$	1.001	1.001	1.003	1.004	1.007	1.006	1.010	1.009	1.011	1.009	1.011	1.010
$ARMA(1,2)$	1.002	1.002	1.005	1.005	1.010	1.008	1.014	1.013	1.015	1.014	1.016	1.014
$ARMA(2,2)$	**0.999**	**0.999**	1.005	1.006	1.009	1.009	1.011	1.009	1.009	1.008	1.010	1.009
$AR(22)$	1.014	1.014	1.012	1.013	1.013	1.013	1.010	1.010	1.009	1.009	1.010	1.009
$HAR(3)$	1.002	1.002	1.004	1.004	1.005	1.005	1.006	1.005	1.007	1.006	1.009	1.008
$AR(50)$	1.039	1.039	1.024	1.029	1.023	1.028	1.021	1.023	1.018	1.018	1.012	1.012
$I(1)$	1.186	1.186	1.238	1.219	1.271	1.237	1.308	1.257	1.327	1.270	1.344	1.281

Table 6: Mean of the $RMSE$ and MAD loss functions relative to the $FI(d)$ model when forecasting a long memory series generated by the error duration model with parameter $d = 0.4$ and all forecast horizons.

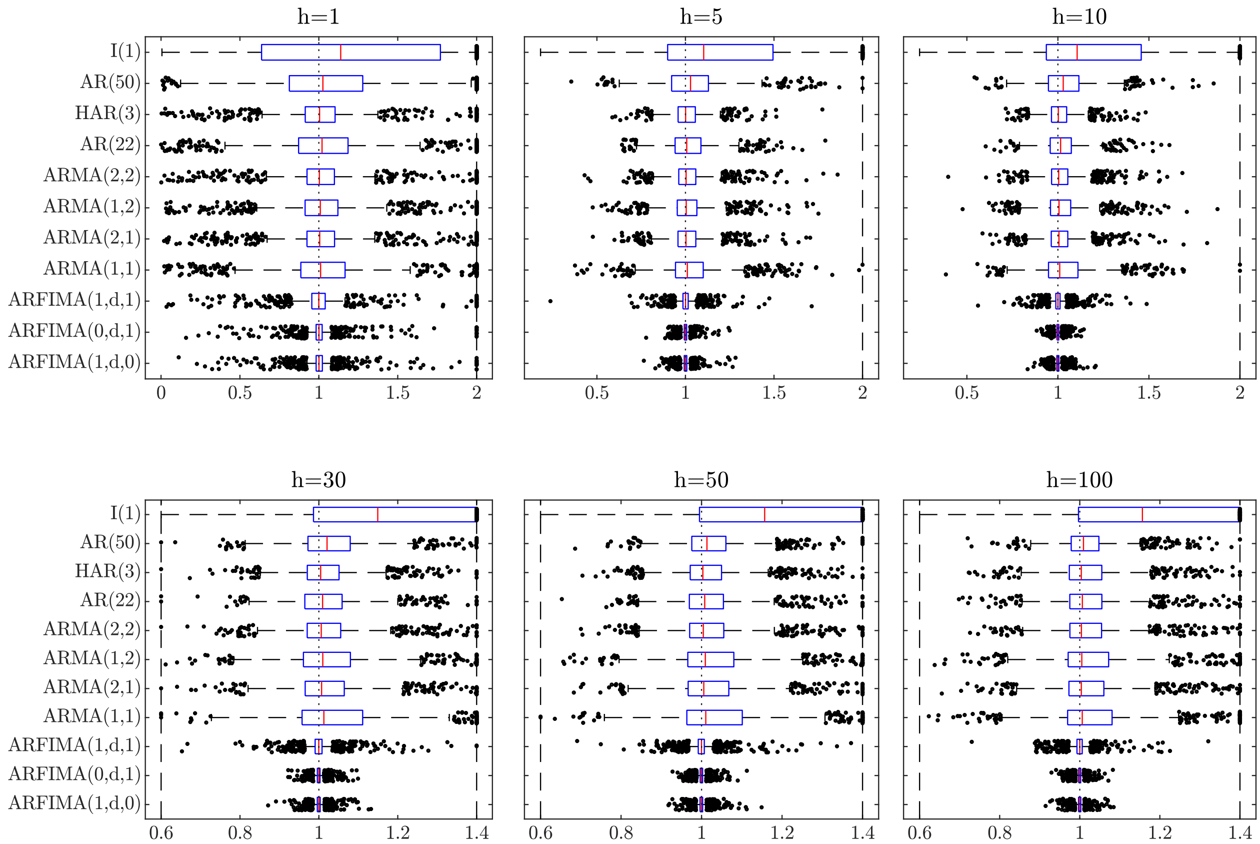
We can see from Table 6 that, even though the $ARFIMA$ class of models provide the best performance for all forecast horizons, short memory models can provide sensible alternatives for short forecast horizons. Notably, the $ARMA(2,2)$ achieves better performance than the pure $FI(d)$, in line with more complex $ARFIMA$ specifications, for $h = 1$. Nonetheless, as the forecast horizon increases, short memory models are dominated by long memory alternatives.

Figure 7: Boxplots for the MAD loss measure for competing models relative to the $FI(d)$ model when forecasting a long memory series generated by the error duration model with long memory parameter $d = 0.1$ for all forecast horizons.



Figures 7 and 8 present boxplots for the MAD loss for all competing models when forecasting processes generated by the error duration model for different degrees of memory. They show that the results from the forecasting exercise for long memory processes generated

Figure 8: Boxplots for the MAD loss measure for competing models relative to the $FI(d)$ model when forecasting a long memory series generated by the error duration model with long memory parameter $d = 0.45$ for all forecast horizons.



by the error duration model are in line with previous results from the other long memory generating mechanisms. The figures show the superior performance of $ARFIMA$ models. Notably, they show the increase in forecasting performance of the long memory models over short memory alternatives as the degree of memory and the forecast horizon increase.

Moreover, contrasting the performance of the $HAR(3)$ model against the $AR(22)$ model, the table and figures show the gains in performance of imposing some structure into the higher-order AR models when forecasting processes generated by the error duration model. The $HAR(3)$ model achieves similar or better performance for all forecast horizons than the unconstrained $AR(22)$ model.

4.4. Overall

The results from the Monte Carlo analysis for all long memory generating mechanisms can be conveyed in the main insights from this paper. Long memory models are the preferred alternative to forecasting long memory processes regardless of the long memory generating mechanism. Furthermore, the relative superior forecasting performance of $ARFIMA$ models against short memory alternatives increases with the long memory parameter. Nonetheless, forecast uncertainty at shorter horizons shows that short memory models may provide sensible forecasting alternatives, particularly for smaller degrees of memory. Finally, the structure imposed by the HAR model seems to provide superior forecasting performance than unconstrained alternatives regardless of the long memory generating mechanism.

5. Discussion

In this section, we explore further the relative performance between an unconstrained AR model and a constrained HAR one in the context of the bias-variance trade-off.

All processes considered in this paper are long memory in the covariance sense. Hence, all models are fitted to capture the information contained in the autocorrelation function and use it for forecasting purposes. In other words, the models select $\{a_i\}_{i=0}^{T}$ in the $x_t = a_0 + \sum_{i=1}^{T} a_i x_{t-i}$ representation, with the aim of mimicking the autocorrelation function.

The $ARFIMA$ and $ARMA$ models differ in terms of the way to select the a_i coefficients. $ARFIMA$ models impose a hyperbolic rate by the fractional differencing operator, $(1-L)^d$, see Equation 2, while high-order AR models are more flexible by selecting each coefficient individually.

In this sense, fractional models need just one parameter to establish the infinite list of coefficients and are hence of low variance. Nonetheless, the uncertainty surrounding the estimation of the long memory parameter may introduce some bias.

As an alternative, high-order AR models are more flexible by using several parameters for modelling. Hence, they can reduce the bias of the forecast, but suffer from increased variance given the number of estimated parameters.

As the Monte Carlo analysis showed, the increased flexibility of high-order AR models can produce good forecast performance at short horizons, mainly when the degree of memory is small. Nonetheless, AR models lose forecasting performance as the forecast horizon increases. We could increase the order of the autoregressive process to increase the forecasting performance at long horizons, but the estimation becomes unstable as the results for the $AR(50)$ showed.

In this context, HAR models are a compromise between the rigid $ARFIMA$ and flexible high-order AR model specifications. They incorporate high-order autoregressive specifications while greatly restricting the number of parameters to be estimated. This arrangement allows the HAR model to provide better forecast performance at small and medium forecast horizons as same-order unrestricted AR models while providing similar long horizon forecasts.

To further illustrate this point, we compare the forecasting performance of additional constrained AR models in the spirit of the HAR model against their unconstrained specifications when forecasting long memory processes generated by all processes considered. In particular, in addition to the $HAR(3)$ and $AR(22)$ models, we show an unrestricted $AR(50)$ model and a $HAR(4)$ given by

$$x_t = a_0 + a_1 x_{t-1}^{(f)} + a_2 x_{t-1}^{(w)} + a_3 x_{t-1}^{(m)} + a_4 x_{t-1}^{(b)} + \epsilon_t,$$

where $x_{t-1}^{(f)} = x_{t-1}$, $x_{t-1}^{(w)} = \frac{1}{5}\sum_{i=1}^{5} x_{t-i}$, $x_{t-1}^{(m)} = \frac{1}{22}\sum_{i=1}^{22} x_{t-i}$, and $x_{t-1}^{(b)} = \frac{1}{50}\sum_{i=1}^{50} x_{t-i}$. Note that the $HAR(4)$ model just described is a constrained $AR(50)$. Moreover, note that the addition of the fourth component depending on lag 50 is arbitrary, and it is only shown for illustrative purposes. A complete analysis of the best $HAR(4)$ specification is outside of the scope of this paper. Such an exercise would evaluate all possible lag combinations and would thus be highly computationally demanding. We leave this area of inquiry open for future research.

Furthermore, to analyse the effects of the particular type of restrictions imposed by the HAR model, we further include restricted versions of $AR(22)$ and $AR(50)$ models given by

$$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-5} + a_3 x_{t-22} + \epsilon_t,$$

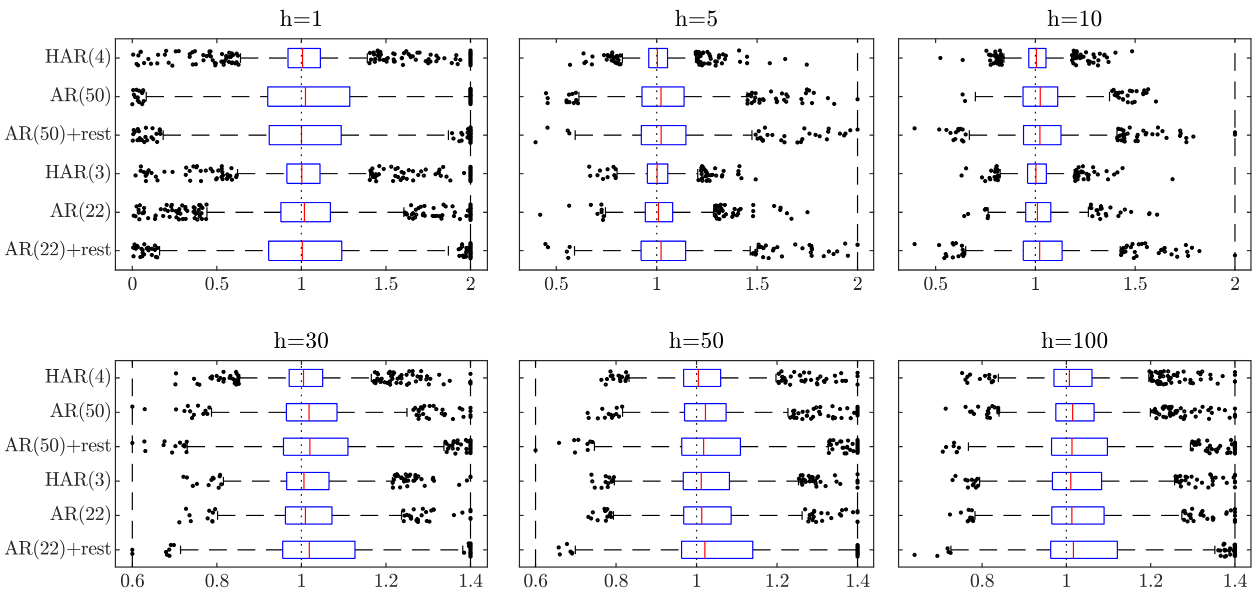
and

$$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-5} + a_3 x_{t-22} + a_4 x_{t-50} + \epsilon_t,$$

respectively. Note that the restricted AR processes only include non-zero values for a few lags, while the HAR specification allows for more non-zero coefficients although restricting them to a particular value.

Figure 9 shows boxplots for the MAD loss for all models and forecast horizons when forecasting long memory generated by cross-sectional aggregation[2] with parameter $d = 0.45$. The figure shows the increase in forecast performance achieved by the restrictions imposed by the HAR model, particularly against the constrained alternatives. The HAR models are the ones that achieve the best forecasting performance across the three different specifications.

Figure 9: Boxplots for the MAD loss measure for competing models relative to the $FI(d)$ model when forecasting a long memory series generated by cross-sectional aggregation with long memory parameter $d = 0.45$ for all forecast horizons.



Moreover, note the increase in forecasting performance of the $HAR(4)$ model relative to the $HAR(3)$ one. The figure shows that superior forecasting performance can be achieved

[2]For reasons of space, we only present the results for long memory generated by cross-sectional aggregation. The results for the other long memory generating mechanisms are qualitatively the same and are available upon request.

by expanding the standard HAR specification to include more lags while still imposing some structure.

6. Considerations

Throughout this study, we have restricted the analysis to the simplest specification of the long memory processes to ease the exposition. That is, we have generated long memory processes without additional short term dynamics. This consideration allowed more direct comparisons between the long memory generating mechanisms. Nonetheless, all mechanisms can be extended in the same manner than the $ARFIMA$ model extends the fractional difference operator by including additional AR and MA components.

In principle, the additional short term dynamics introduced by the extra AR and MA components can be captured by increasing the number of lags included in the forecasting models. A broader set of models would be considered in this regard.

Moreover, we have abstained from studying the role that the sample size has on estimation. We have assumed a relatively large sample of $T = 1,000$ to get reliable estimators. For smaller sample sizes, the uncertainty regarding the model estimates will add to the bias-variance trade-off. Given the uncertainty associated with estimating an increasing number of parameters, we would expect the performance of high-order AR models to deteriorate in short series.

Finally, we have abstracted from analysing processes with negative long memory parameter. Vera-Valdés (2019) show that, in contrast to $ARFIMA$ processes, cross-sectionally aggregated processes do not possess the antipersistence property. In particular, the spectrum of cross-sectionally aggregated processes with negative degrees of memory does not collapse at the origin. Thus, the behaviour between long memory generating processes with negative degrees of memory are qualitatively different, which could have significant repercussions for forecasting purposes. Further analysis of antipersistent processes is left for future work.

7. Illustrative Example

In this section, we evaluate the forecasting performance of the competing models on real data. We select the Realized Volatility (RV) of the S&P 500 for illustration. We obtain the RV series from the Oxford-Man Institute's "Realised Library" computed based on intradaily observations spaced into 5-minute intervals and subsampled at a 1-minute frequency. The sample runs from January 4, 2010, until December 31, 2018.

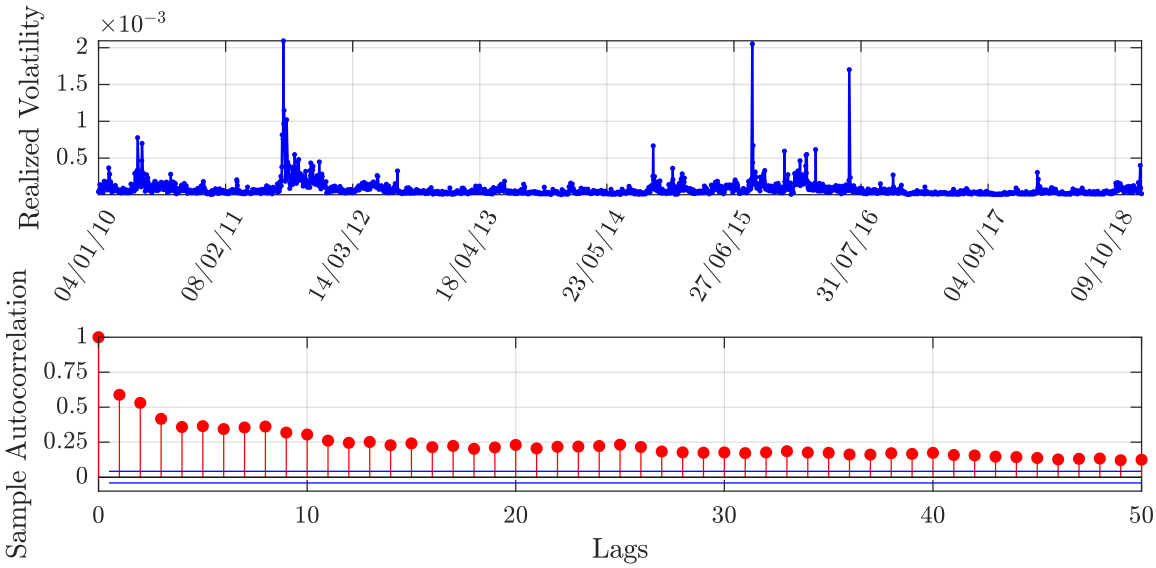
The RV has been proven to have long memory by, among others, Martens et al. (2009) and Andersen et al. (2003). In particular, notice that by construction, the S&P 500 series is an aggregated measure; thus, it is in line with the cross-sectional argument for long memory. Furthermore, Parke (1999) argues that the difference between information quality among agents makes the error duration model capable of explaining the long memory in volatility. Hence, the realized volatility can be argued to have long memory by the theoretical explanations considered in this work, making it a good fit for the exercise.

Figure 10 presents the RV series and its autocorrelation function. The autocorrelation function shows the typical behaviour of long memory processes, remaining significant at large lags. The estimates for the long memory parameter are 0.4315, 0.4161, and 0.4529 for the semi-parametric estimator of Geweke and Porter-Hudak (1983), the bias-corrected version of Andrews and Guggenberger (2003), and the maximum likelihood estimator of Sowell (1992), respectively.

For estimation, we use a rolling window of seven years starting from January 4, 2010, to December 29, 2017, which translates to 2049 observations. We then construct forecasts for $1, 5, 10, 22, 66,$ and 120 days periods ahead and compute the MAD and $RSME$ losses for each estimation window and forecast horizon. We report the mean values of 200 windows.

Table 7 presents the results from the forecasting exercise. As the table shows, short memory models are among the top-performing for horizons up to 10 periods ahead. In particular, both the unconstrained $AR(50)$ and $HAR(3)$ are among the ones with minimum loss functions for short forecast horizons, in line with the results of Martens et al. (2009) and Corsi (2009). Nonetheless, the $ARFIMA$ type of models tends to dominate

Figure 10: Realized volatility and its autocorrelation function.



in forecast performance as the forecast horizon increases, in line with our results from the Monte Carlo analysis. The $ARFIMA(0,d,1)$ is the one with the lowest MAD for horizons $h \in \{10, 22, 66, 122\}$, representing 2 weeks, and 1, 3, and 6 months.

RV	h=1		5		10		22		66		120	
	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$	MAD	$RMSE$
$FI(d)$	0.022	0.022	0.025	0.029	0.028	0.033	0.031	0.039	0.037	0.047	0.042	0.056
$ARFIMA(1,d,0)$	0.093	0.093	0.082	0.091	0.089	0.100	0.113	0.132	0.201	0.240	0.264	0.321
$ARFIMA(0,d,1)$	0.019	0.019	0.021	0.025	**0.022**	**0.028**	**0.024**	**0.032**	**0.026**	**0.036**	**0.028**	**0.041**
$ARFIMA(1,d,1)$	0.829	0.829	0.795	0.807	0.815	0.833	0.852	0.879	0.750	0.804	0.638	0.737
$ARMA(1,1)$	0.019	0.019	0.023	0.027	0.028	0.033	0.033	0.040	0.038	0.045	0.039	0.049
$ARMA(2,1)$	0.021	0.021	0.028	0.031	0.033	0.038	0.038	0.043	0.040	0.046	0.040	0.049
$ARMA(1,2)$	0.021	0.021	0.028	0.032	0.034	0.039	0.038	0.044	0.040	0.046	0.040	0.049
$ARMA(2,2)$	0.020	0.020	0.027	0.031	0.032	0.037	0.037	0.043	0.040	0.046	0.040	0.049
$AR(22)$	**0.018**	**0.018**	0.021	0.025	0.024	0.029	0.027	0.034	0.032	0.040	0.035	0.046
$HAR(3)$	**0.018**	**0.018**	0.021	0.025	0.023	0.029	0.027	0.034	0.032	0.040	0.035	0.045
$AR(50)$	**0.018**	**0.018**	0.021	0.025	0.023	**0.028**	0.026	0.033	0.030	0.038	0.033	0.044
$I(1)$	**0.018**	**0.018**	**0.020**	**0.024**	0.023	0.029	0.026	0.034	0.030	0.040	0.032	0.046

Table 7: Mean of the $RMSE$ and MAD loss functions when forecasting the S&P realized volatility on a rolling window at all forecast horizons.

Overall, the example with real data confirms our results from the simulations. It shows that for short horizons, short memory models achieve good forecasting performance. As we increase the forecasting horizon, the $ARFIMA$ alternatives start to dominate. These results are of significant relevance for fund managers interested in forecasting at medium and long horizons, some of the types that serve as inspiration for the HAR model.

8. Conclusions

This paper argues that the most cited theoretical arguments behind the presence of long memory in the data do not imply the fractional difference operator. In this context, it evaluates the forecasting performance of $ARFIMA$ models when the long memory is generated from nonfractional sources.

We find that $ARFIMA$ models achieve better forecasting performance than short memory alternatives for all long memory generating mechanisms and forecast horizons. Nonetheless, forecast uncertainty at short horizons is such that short memory models may achieve comparable forecast performance as $ARFIMA$ models.

Moreover, the superior performance of the $ARFIMA$ model gets exacerbated, the higher the degree of long memory of the processes. Thus, $ARFIMA$ models are the preferred alternative for forecasting long memory processes regardless of the generating mechanism; while short memory models may provide suitable alternatives for forecasts at short horizons and lower degrees of memory.

Additionally, by making a compromise between flexibility and complexity, we find that the structure imposed by the HAR model results on better forecasting performance than unconstrained alternatives and unstructured constrained ones. Furthermore, these results are robust to the long memory generating mechanism.

Our results are of use for Climate Econometrics and Financial Econometrics models dealing with forecasts at different horizons. They show that even when the practitioner may not be confident regarding the origin of the long memory in the data, the $ARFIMA$ class of models gives the best forecasting performance, notably at medium and long forecast horizons. As an illustrative example, we show for the Realized Volatility of the S&P 500 that while short memory models are well suited for forecasts up to a week ahead, the $ARFIMA$ class of models dominates in forecast performance for longer horizons.

Acknowledgements

The author would like to thank Niels Haldrup, and two anonymous referees for all the insightful comments, the paper greatly improved because of them. A previous version of the paper was partly written while on a visiting stay at Erasmus University Rotterdam. The author would like to thank the Faculty and Administrative Staff at the Erasmus School of Economics for all their help; in particular to Michel van der Wel and Dick van Dijk for the interesting conversations.

References

Altissimo, F., Mojon, B., Zaffaroni, P., 2009. Can aggregation explain the persistence of inflation? Journal of Monetary Economics 56, 231–241. doi:10.1016/j.jmoneco.2008.12.013.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. Review of Economics and Statistics 89, 701–720. doi:10.1162/rest.89.4.701.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625. doi:10.1111/1468-0262.00418.

Andrews, D.W., Guggenberger, P., 2003. A Bias-Reduced Log-Periodogram Regression Estimator For The Long-Memoty Parameter. Econometrica 71, 675–712.

Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59. doi:10.1016/0304-4076(95)01732-1.

Baillie, R.T., Kongcharoen, C., Kapetanios, G., 2012. Prediction from ARFIMA models: Comparisons between MLE and semiparametric estimation procedures. International Journal of Forecasting 28, 46–53. URL: http://dx.doi.org/10.1016/j.ijforecast.2011.02.012, doi:10.1016/j.ijforecast.2011.02.012.

Balcilar, M., 2004. Persistence in inflation: does aggregation cause long memory? Emerging Markets Finance and Trade 40, 25–56. URL: http://mesharpe.metapress.com/index/61tjjgncdlktvdrh.pdf.

Barkoulas, J.T., Baum, C.F., 1997. Fractional Differencing Modeling and Forecasting of Eurocurrency Deposit Rates. The Journal of Financial Resarch XX, 355–372.

Beran, J., Bhansali, R.J., Ocker, D., 1998. On unified model selection for stationary and non stationary short and long memory autoregressive processes. Biometrika 85, 921–934.

Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long-Memory Processes: Probabilistic Theories and Statistical Methods. Springer. URL: http://onlinelibrary.wiley.com/doi/10.1002/9780470131466.ch3/summary, doi:10.1007/978-3-642-35512-7.

Bhardwaj, G., Swanson, N.R., 2006. An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. Journal of Econometrics 131, 539–578. doi:10.1016/j.jeconom.2005.01.016.

Box-Steffensmeier, J.M., De Boef, S., Lin, T.M., 2004. The dynamics of the partisan gender gap. American Political Science Review 98, 515–528. doi:10.1017/S0003055404001315.

Chiriac, R., Voev, V., 2011. Modelling and Forecasting Multivariate Realized Volatility. Journal of Applied Econometrics 26, 922–947.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196. doi:10.1093/jjfinec/nbp001.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econometric Reviews 27, 46–78. doi:10.1080/07474930701853616.

Crato, N., Ray, B.K., 1996. Model selection and forecasting for long-range dependent processes. Journal of Forecasting 15, 107–125. doi:10.1002/(SICI)1099-131X(199603)15:2<107::AID-FOR612>3.3.CO;2-4.

Diebold, F.X., Rudebusch, G.D., 1989. Long Memory and Persistence in Agregate Output. Journal of Monetary Economics 24, 189–209. URL: http://scholar.google.com.br/scholar_url?hl=pt-BR&q=http://www.ssc.upenn.edu/$\sim$fdiebold/papers/paper88/DieboldRudebusch1989.pdf&sa=X&scisig=AAGBfm29cdDPDH0vZTbRWl2x6I3_zBiN5A&oi=scholarr&ei=JywWU_fGBdD6kQfslID4AQ&ved=0CC0QgAMoADAA, doi:http://dx.doi.org/10.1016/0304-3932(89)90003-2.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.

Gil-Alana, L.A., 2005. Statistical modeling of the temperatures in the Northern Hemisphere using fractional integration techniques. Journal of Climate 18, 5357–5369. doi:10.1175/JCLI3543.1.

Granger, C.W., 1980. Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227–238. doi:10.1016/0304-4076(80)90092-5.

Granger, C.W., Joyeux, R., 1980. An Introduction to Long Memory Time Series Models and Fractional Differencing. Journal of Time Series Analysis 1, 15–29. doi:10.1111/j.1467-9892.1980.tb00297.x.

Haldrup, N., Vera-Valdés, J.E., 2017. Long memory, fractional integration, and cross-sectional aggregation. Journal of Econometrics 199, 1–11. URL: http://dx.doi.org/10.1016/j.jeconom.2017.03.001, doi:10.1016/j.jeconom.2017.03.001.

Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68, 165–176. doi:10.1093/biomet/68.1.165.

Hurst, H.E., 1956. The Problem of Long-Term Storage in Reservoirs. International Association of Scientific Hydrology. Bulletin 1, 13–27. URL: http://www.tandfonline.com/doi/abs/10.1080/02626665609493644, doi:10.1080/02626665609493644.

Jensen, A.N., Nielsen, M.Ø., 2014. A Fast Fractional Difference Algorithm. Journal of Time Series Analysis 35, 428–436. URL: http://doi.wiley.com/10.1111/jtsa.12074, doi:10.1111/jtsa.12074.

Linden, M., 1999. Time series properties of aggregated AR(1) processes with uniformly distributed coefficients. Economics Letters 64, 31–36. URL: http://ideas.repec.org/a/eee/ecolet/v64y1999i1p31-36.html.

Martens, M., van Dijk, D., 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, 282–303. URL: http://dx.doi.org/10.1016/j.ijforecast.2009.01.010, doi:10.1016/j.ijforecast.2009.01.010.

Oppenheim, G., Viano, M.C., 2004. Aggregation of random parameters ornstein-uhlenbeck or ar processes: Some convergence results. Journal of Time Series Analysis 25, 335–350. doi:10.1111/j.1467-9892.2004.01775.x.

Parke, W., 1999. What is fractional integration? Review of Economics and Statistics 81, 632–638. URL: http://www.mitpressjournals.org/doi/abs/10.1162/003465399558490, doi:10.1162/003465399558490.

Ray, B.K., 1993. Modeling Long Memory Processes for Optimal Long Range Prediction. Journal of Time Series Analysis 14, 511–525. doi:10.1111/j.1467-9892.1993.tb00161.x.

Robinson, P.M., 1978. Statistical Inference for a Random Coefficient Autoregressive Model. Scandinavian Journal of Statistics 5, 163–168. doi:10.2307/4615707.

Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165–188. doi:10.1016/0304-4076(92)90084-5.

Vera-Valdés, J.E., 2019. Nonfractional Long Memory: Generation, Antipersistence, and Estimation. Working Paper .

Zaffaroni, P., 2004. Contemporaneous aggregation of linear dynamic models in large economies. Journal of Econometrics 120, 75–102. doi:10.1016/S0304-4076(03)00207-0.