Exponential Smoothing, Long Memory and Volatility Prediction

Tommaso Proietti[1]
Department of Economics and Finance, University of Rome Tor Vergata, Italy, and CREATES, Aahrus University, Denmark

[1]Address for correspondence: Dipartimento di Economia e Finanza, Università di Roma Tor Vergata, Via Columbia 2, 00133 Rome, Italy, email: tommaso.proietti@uniroma2.it, tel. +39 06 7259 5941. The author thanks Federico Carlini for discussion on the fractional lag operator and Niels Haldrup for discussion concerning Whittle estimation of the fractionally integrated models. The author gratefully acknowledges financial support by the Italian Ministry of Education, University and Research (MIUR), PRIN Research Project 2010-2011 - prot. 2010J3LZEN, "Forecasting economic and financial time series" and from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation..

Abstract

Extracting and forecasting the volatility of financial markets is an important empirical problem. The paper provides a time series characterization of the volatility components arising when the volatility process is fractionally integrated, and proposes a new predictor that can be seen as extension of the very popular and successful forecasting and signal extraction scheme, known as exponential smoothing (ES). First, we derive a generalization of the Beveridge-Nelson result, decomposing the series into the sum of fractional noise processes with decreasing orders of integration. Secondly, we consider three models that are natural extensions of ES: the fractionally integrated first order moving average (FIMA) model, a new integrated moving average model formulated in terms of the fractional lag operator (FLagIMA), and a fractional equal root integrated moving average (FerIMA) model, proposed originally by Hosking. We investigate the properties of the volatility components and the forecasts arising from these specification, which depend uniquely on the memory and the moving average parameters. For statistical inference we show that, under mild regularity conditions, the Whittle pseudo-maximum likelihood estimator is consistent and asymptotically normal. The estimation results show that the log-realized variance series are mean reverting but nonstationary. An out-of-sample rolling forecast exercise illustrates that the three generalized ES predictors improve significantly upon commonly used methods for forecasting realized volatility, and that the estimated model confidence sets include the newly proposed fractional lag predictor in all occurrences.

Keywords: Realized Volatility. Volatility Components. Fractional lag models. Fractional equal-root IMA model. Model Confidence Set.

JEL codes: C22, C53, G17.

1 Introduction

Volatility is an important characteristic of financial markets. Its measurement and prediction has attracted a lot of interest, being quintessential to the assessment of market risk and the pricing of financial products. A possible approach is to adopt a conditionally heteroscedastic or a stochastic volatility model for asset returns; see Engle (1995) and Shephard (2005) for a collection of key references in these areas. Alternatively, we can provide a statistical model for a time series proxy of volatility, such as squared returns or daily realized volatility (RV) measures (possibly after a logarithmic transformation of the series). RV is a nonparametric estimate of the volatility constructed from intraday price movements[1], introduced in the seminal contributions by Merton (1980), and characterized by Comte and Renault (1998), Andersen et al. (2001a, 2001b), and Barndorff-Nielsen and Shephard (2002a, 2002b), as an estimator of the quadratic variation process of the logarithmic price of a financial asset.

This paper is based on the latter approach and takes as a well established fact the presence of long range dependence as a characteristic feature of volatility, see Ding, Granger and Engle (1993), Comte and Renault (1998), Bollerslev and Wright (2000), Andersen et al. (2003), Hurvich and Ray (2003).

The mechanism originating the long memory phenomenon is not without controversy. One view is that it is spuriously generated by stochastic regime switching and occasional breaks (see, among others, Diebold and Inoue, 2001, Granger and Hyung, 2004, and Mikosh and Starica, 2004).

An alternative view considers long memory as an intrinsic feature of volatility, arising from its multicomponent structure, that is, from the aggregation of various components, characterised by different degrees of persistence and predictability. This view is taken in the *mixture-of-distributions hypothesis* by Andersen and Bollerslev (1997), and in Gallant, Hsu and Tauchen (1999). It embeds Granger's (1980) seminal result, showing that a long memory process results from the contemporaneous aggregation of infinite first order AR processes[2].

[1]We refer, in particular, to the realized variance, obtained as the sum of squared intraday returns computed over small time intervals.

[2]In the GARCH framework, Engle and Lee (1999) introduced a component GARCH model such that the conditional variance is the sum of two AR processes for the short and long run components. Adrian and Rosenberg (2008) formulate a log-additive model of volatility where the long run component, a persistent AR(1) process with non-zero mean, is related to business cycle conditions, and the short run component, a zero mean AR(1), is related to the tightness of

Müller et al. (1993) propose the *heterogeneous market hypothesis*, according to which the components reflect the different time horizons expressed by the market participants. This hypothesis is also at the foundations of the heterogeneous autoregressive (HAR) model by Corsi (2009), which identifies three volatility components over different horizons (daily, weekly and monthly). The HAR model has become a discipline standard, as it reproduces quite effectively the long memory feature and has proved extremely practical and successful in predicting realized volatility. Corsi and Renò (2012) have recently extended the heterogenous structure to model the leverage effect, while Corsi et al. (2012) and Majeski et al. (2015) formulate the HAR Gamma model in an option pricing context. Calvet and Fisher (2001) propose a multifractal approach which models the volatility process as the multiplicative product of an infinite number of Markov-switching components with heterogeneous frequencies. In Bandi et al. (2015) a multicomponent model is presented, where the components are defined at different scales, $j = 1, \dots, J$, corresponding to horizons of $h_j = 2^j$ time units; each scale is characterized by different levels of predictability and is impacted by specific shocks that vary in size and persistence with j.

With respect to the available literature, the present paper does a reverse operation: starting from the long memory representation of the volatility process, we derive components that are uniquely identified by their time series properties and are characterised by a different degree of memory. More specifically, we provide a decomposition of a process integrated of order d into the sum of fractional noise processes of decreasing orders, plus a stationary remainder term. Our result generalizes the popular Beveridge and Nelson (1981, BN henceforth) decomposition to any positive fractional integration order.

Secondly, we focus on a set of long memory models that can be used to extract the volatility components and produce forecasts that generalize the exponential smoothing (ES) predictor; ES is a very popular and successful forecasting scheme among practitioners, as well as a filter for extracting the long run component, or underlying level, of a time series. The models that we consider share the common feature of depending on two parameters, in addition to the one-step-ahead prediction error variance: the memory parameter and a moving average (MA) parameter, which regulates the mean reversion property and the size of the short-run component. We

financial conditions. Engle, Ghysels and Sohn (2013) have recently introduced multiplicative and log-additive GARCH-MIDAS model, where the long run component is a weighted average of past realized volatilities, where the weights follow a beta distribution.

focus in particular on three models: the first is the well known fractionally integrated moving average (FIMA) process. The second is a new proposal, an integrated moving average model formulated in terms of the fractional lag operator (FLagIMA). Finally, we consider the fractional equal-root integrated moving average (FerIMA) model, which according to Hosking (1981, p. 175) "as a forecasting model it corresponds to fractional order multiple exponential smoothing".

For the three specifications we present the time series properties of the components, the filters for their extraction and the forecast function, which encompasses ES in all cases when the integration order is 1. The characteristic property of the FLagIMA decomposition is that the short-run components is purely unpredictable from its past, i.e. it is white noise (WN).

The models can be estimated in the frequency domain by Whittle pseudo-maximum likelihood, and we show that, under regularity conditions, the maximum likelihood estimators are consistent and asymptotically normal. We finally address the empirical relevance of the volatility predictors arising from the three models by performing a rolling forecasting experiment, using the realized variance time series of 21 stock indices, and estimating the model confidence sets, when the prediction methods under comparison includes variants of ES and the HAR predictor. The overall conclusion is that the generalized long memory predictors are a useful addition to the model set, and that in particular the newly proposed fractional lag predictor belongs to the estimated model confidence sets in all occurrences.

The paper is structured in the following way. Section 2 reviews the ES predictor and the signal extraction filter associated with the integrated MA model. In section 3 we present the decomposition of a fractionally integrated process into components with decreasing level of persistence. Section 4 discusses the three extensions of ES for fractionally integrated processes. In section 5 we deal with maximum likelihood estimation of the parameters. In section 6 we present the empirical study and discuss the estimation and forecasting results. Section 7 concludes the paper.

2 Exponential smoothing

Let y_t denote a time series. Given the availability of a sample stretching back to the indefinite past, ES yields the l-steps-ahead predictor of y_t

$$\tilde{y}_{t+l|t} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j y_{t-j}, \quad l = 1, 2, \dots. \tag{1}$$

The predictor depends on the smoothing constant, λ, which takes values in the range (0,1]. The above expression is an exponentially weighted moving average (EWMA) of the available observations, with weights declining according to a geometric progression with ratio $1-\lambda$. The eventual forecast function is a horizontal straight line drawn at $\tilde{y}_{t+1|t}$. The predictor is efficiently computed using either one of the two following equivalent recursions:

$$\tilde{y}_{t+1|t} = \lambda y_t + (1-\lambda)\tilde{y}_{t|t-1}, \qquad \tilde{y}_{t+1|t} = \tilde{y}_{t|t-1} + \lambda(y_t - \tilde{y}_{t|t-1}). \tag{2}$$

If a finite realisation is available, $\{y_t, t = 1, 2, \dots n\}$, and we denote by $\tilde{y}_{1|0}$ the initial value of (2), then we may initialise (2) by $\tilde{y}_{1|0} = y_1$, or by the average of the first t observations, $\tilde{y}_{1|0} = \sum_{j=0}^{t-1} y_{t-j}/t$. Comprehensive reviews of exponential smoothing and its extensions are provided by Gardner (2006) and Hyndman et al. (2008).

It is well known that the ES predictor is the best linear predictor for the integrated moving average (IMA) process

$$(1-B)y_t = (1-\theta B)\xi_t, \;\; \xi_t \sim \mathrm{WN}(0, \sigma^2), \tag{3}$$

where B is the backshift operator, $B^k y_t = y_{t-k}$, $0 \leq \theta < 1$ is the moving average parameter. In particular,

$$\tilde{y}_{t+l|t} = y_t - \theta\xi_t = \frac{1-\theta}{1-\theta B} y_t \tag{4}$$

is the minimum mean square predictor of y_{t+l} based on the time series $\{y_j, j \leq t\}$. This is equivalent to the above EWMA with $\lambda = 1-\theta$.

The BN decomposition of the IMA(1,1) process $y_t = m_t + e_t$ is such that the permanent component is the random walk (RW) process, $m_t = m_{t-1} + (1-\theta)\xi_t$, and the transitory component is WN, $e_t = \theta\xi_t$. The permanent component is equal to the long run prediction of the series at time t and it is thus measurable at time t

as $m_t = (1-\theta)(1-\theta B)^{-1} y_t$.

A remarkable feature of the IMA(1,1) process and the associated ES forecasting scheme is that the one-step-ahead forecast is coincident with the long-run (eventual) forecast. This feature is only possessed by this model (which encompasses the RW).

Exponential smoothing has been used for forecasting and extracting the level of volatility according to the RiskMetrics methodology developed by J.P. Morgan. The Riskmetrics 1994 methodology (RM1994, see RiskMetrics Group, 1996) is based on an EWMA with parameter $\lambda = 0.06$, or, equivalently, $\theta = 0.94$. For long memory time series Riskmetrics (see Zumbach, 2007) has proposed a new methodology, referred to as RM2006, which aims at mimicking a filter with weights decaying at a hyperbolic, rather than geometric, rate. This is achieved by combining several ES filters with different smoothing constants. See Appendix A for more details.

3 Fractionally integrated time series models and their decomposition

Let $\{y_t\}$ denote the fractionally integrated random process

$$(1-B)^d y_t = \psi(B)\xi_t, \quad \xi_t \sim \text{WN}(0, \sigma^2),$$

where $\psi(B) = 1 + \psi_1 B + \cdots$ is such that $\sum_j |\psi_j| < \infty$. For non-integer $d > -1$, the fractional differencing operator is defined according to the binomial expansion as

$$(1-B)^d = 1 - \sum_{j=1}^{\infty} \beta_j B^j, \quad \beta_j = \frac{d(1-d)\cdots(j-1-d)}{j!} = (-1)^{j+1}\frac{(d)_j}{j!}, \qquad (5)$$

where $(d)_j = d(d-1)\cdots(d-j+1)$ is the Pochhammer symbol. Notice that β_j is given recursively as $\beta_j = \beta_{j-1}\frac{j-d-1}{j}, j \geq 1$. When $\psi(B) = 1$, y_t is a fractional noise process and will be denoted $y_t \sim \text{FN}(d)$. For $d \in (0, 0.5)$ the process is stationary and its properties can be characterised by the autocorrelation function, which decays hyperbolically to zero, and its spectral density, which is unbounded at the origin.

In this section we provide a decomposition of a fractionally integrated process of order $d > 0$, which can be viewed as a generalization of the Beveridge and Nelson (1981) decomposition, which arises for $d = 1$. We denote by $\lfloor x \rfloor$, the largest integer not greater than x.

Proposition 1. *Let y_t be generated as $(1-B)^d y_t = \psi(B)\xi_t$, $\psi(B) = 1+\psi_1 B+\psi_2 B^2+\cdots$, with $d > 0$ and $\xi_t \sim WN(0,\sigma^2)$. Assume that $\psi(z)$ converges absolutely for $|z| < 1+\epsilon, \epsilon > 0$, and let $r = \lfloor d+1/2 \rfloor$. Then y_t admits the following decomposition:*

$$\begin{aligned} y_t &= m_t + e_t, \\ m_t &= u_{0t} + u_{1t} + \cdots u_{r-1,t}, \;\; u_{jt} = \tfrac{\psi_j(1)}{(1-B)^{d-j}}\xi_t, \quad j = 0,\ldots,r-1, \\ e_t &= \tfrac{\psi_r(B)}{(1-B)^{d-r}}\xi_t, \end{aligned} \tag{6}$$

where $\psi_0(B) = \psi(B)$, and

$$\psi_{j-1}(B) = \psi_{j-1}(1) + (1-B)\psi_j(B), \quad j = 1, 2, \ldots.$$

Proof. See Appendix B. □

The process u_{jt} is integrated of order $d-j > 0.5$, $j = 1,\ldots,r-1$, and thus it is nonstationary. The component m_t is the nonstationary component, determining the behaviour of the forecast function for long multistep horizons. The component e_t is stationary, featuring long memory ($d > r$) or antipersistence ($d < r$), its integration order being $d-r \in [-0.5, 0.5)$.

If $d > 0.5$, m_t has an ARFIMA$(0,d,r-1)$ representation, $(1-B)^d m_t = \left[\sum_{j=0}^{r-1}\psi_j(1)(1-B)^j\right]\xi_t$. In terms of the observations, $\psi(B)m_t = \left[\sum_{j=0}^{r-1}\psi_j(1)(1-B)^j\right]y_t$.

It is perhaps useful to highlight some particular cases:

- The case $d = 1$ gives the well-known BN decomposition of y_t into a RW and a short memory transitory component.
- When $d \in (0, 0.5)$ the component m_t is identically equal to zero and all the series is transitory. We could obviously extract a fractional noise component from y_t, e.g. writing $y_t = (1-B)^{-d}\psi(1)\xi_t + (1-B)^{-d+1}\psi_1(B)\xi_t$, but it would still be a stationary component.
- When $d \in [0.5, 1.5)$, y_t admits the following nonstationary-stationary decomposition:

  $$y_t = m_t + e_t, \quad m_t = u_{0t}, \quad e_t = \frac{\psi_1(B)}{(1-B)^{d-1}}\xi_t$$

  Notice that for $d \in [0.5, 1)$ the long run forecast of the series is zero and the shocks $(1-\theta)\xi_t$ have long lasting, but transitory, effects. Hence m_t is not equivalent its long run prediction, i.e. the value the series would take if it were on its long run path. It should also be noticed that the above decomposition differs from the one proposed by Ariño and Marmol (2004). The latter is based on a different interpolation argument and decomposes $y_t = m_t^* + e_t^*$, where, in terms of our notation, $e_t^* = \frac{\psi_1(B)}{\Gamma(d)}\xi_t$, where $\Gamma(\cdot)$ is the gamma function. As a result, their permanent component is

$$m_t^* = y_t - e_t^* = z_{0t} + \frac{1}{\Gamma(d)}\left[\Gamma(d)(1-B)^{1-d} - 1\right]\psi_1(B)\xi_t.$$

Thus, it contains a purely short memory component.

- When $d \in [1.5, 2.5)$, y_t admits the following nonstationary-stationary decomposition:

$$y_t = m_t + e_t, \quad m_t = u_{0t} + u_{1t}, \quad e_t = \frac{\psi_2(B)}{(1-B)^{d-2}}\xi_t$$

  If $d = 2$ and $\psi(B) = (1 - \theta_1 B - \theta_2 B^2)$, $\theta_2 < 0$, then m_t can be computed according to the Holt-Winters recursive formulae (see Hyndman et al., 2008) and e_t is white noise. If further $\theta_1 = -4\theta_2/(1-\theta_2)$, m_t is the real time Hodrick-Prescott trend.

- If $\psi(B) = 1 + \psi_1 B + \cdots + \psi_q B^q$, an MA($q$) polynomial, then, if $r = q$, $\psi_r(B)$ is a zero degree polynomial. If $r < q$, then $\psi_r(B)$ is MA($q - r$).

- If $\psi(B)$ is an ARMA(p, q) polynomial, then $\psi_r(B)$ is an ARMA(p, $\min\{(p-r), (q-r)\}$) polynomial. Consider for simplicity the case when $q \geq p$, so that, defining $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ and $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, we can write

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = \frac{\theta_0(1) + \theta_1(1)(1-B) + \cdots + \theta_q(1)(1-B)^q}{\phi(B)};$$

  we have that for $r \leq q$,

$$\psi_r(B) = \theta_r(1) + \theta_{r+1}(1-B) + \cdots + \theta_q(1)(1-B)^{q-r},$$

  i.e. an MA($q - r$) lag polynomial.

- For integer $d = r > 0$, e_t is a stationary short memory process, with ARMA(p, $\min\{(p-r), (q-r)\}$) representation, if $\psi(B)$ is ARMA(p, q), and can be referred to as the BN transitory component, as its prediction converges geometrically fast to zero as the forecast horizion increases; m_t is the long run or permanent component, as it represents the value that the series would take if it were on the long run path, i.e. the value of the long run forecast function actualised at time t.

Finally, it is important to notice that the predictions of y_t coincide with those of m_t if either $\psi_r(B) = (1-B)^{d-r}$, or $d = r$ and $\psi_r(B) = c$, for a constant c, i.e. when e_t is a WN process.

4 Fractional Generalizations of the Exponential Smoothing Filter and Predictor

We consider three extensions of the IMA(1,1) model, that are valid in the fractionally integrated case and which yield predictors that generalize ES and that allow to define volatility components encompassing its BN decomposition. The first is a specific model of the ARFIMA class, and it is well known. The second is a novel proposal that is formulated in terms of the fractional lag operator, while the third was suggested by Hosking (1980), but to our knowledge has never been applied in practice. The three models are all particular cases of the fractional model $(1-B)^d y_t = \psi(B)\xi_t$ considered in the previous section and differ for the specification of $\psi(B)$, which will depend on d and an MA parameter.

4.1 Fractional Integrated Moving Average (FIMA) Process

The fractional integrated moving average process is

$$(1-B)^d y_t = (1-\theta B)\xi_t, \;\; \xi_t \sim \text{WN}(0, \sigma^2),$$

where we assume that $0 \leq \theta < 1$.

By Proposition 1, if $d \in [0.5, 1.5)$, $y_t = m_t + e_t$, where $(1-B)^d m_t = (1-\theta)\xi_t$ and $e_t = \theta\xi_t/(1-B)^{d-1}$ is a stationary fractional noise process. In terms of the observations, $m_t = [(1-\theta)/(1-\theta B)]y_t$, which is an EWMA of the current and past observations. Hence, the peculiar trait of this model is that signal extraction takes place by ES, regardless of the d parameter.

The forecasts are dependent on the long memory parameter: the one-step-ahead predictor can be written as

$$\tilde{y}_{t+1|t} = (d-\theta)y_t + \theta\tilde{y}_{t|t-1} + \sum_{j=0}^{\infty}(-1)^{j+1}\frac{(d)_{j+2}}{(j+2)!}y_{t-j-1}.$$

This formula encompasses ES for $d = 1$, but except for this case, the predictions of m_t are not coincident with those of y_t, unless $\theta = 0$. In particular, $\tilde{m}_{t+1|t} = \theta m_t + (1-\theta)\tilde{y}_{t+1|t}$, a weighted average of the current level and the one-step prediction of the series. In other words, the prediction of y_{t+1} is shrunk towards m_t to form the prediction of m_{t+1}. Finally, e_t is predictable from its past, namely, $\tilde{e}_{t+1|t} = \theta(\tilde{y}_{t+1|t} - m_t)$.

4.2 Fractional Lag Integrated Moving Average (FLagIMA) Process

Let B_d denote the fractional lag operator, defined as

$$B_d = 1 - (1-B)^d. \tag{7}$$

This is a generalization of the lag operator to the long memory case originally proposed by Granger (1986) and recently adopted by Johansen (2008) to define a new class of vector autoregressive models. Obviously, $d = 1$ gives $B_d = B$, the usual lag operator.

Consider the following fractional lag IMA (FLagIMA) model, obtained by replacing the lag operator by B_d in the IMA(1,1) model:

$$(1 - B_d)y_t = (1 - \theta B_d)\xi_t, \quad 0 \le \theta < 1.$$

We can rewrite the model in terms of the usual lag operator as

$$(1-B)^d y_t = \psi(B)\xi_t, \;\; \psi(B) = (1-\theta) + \theta(1-B)^d,$$

and, using (5), it is immediate to show that $\psi(B) = \sum_{j=0}\psi_j B^j$ has coefficients

$$\psi_0 = 1, \quad \psi_j = (-1)^j\theta\frac{(d)_j}{j!}, \quad j \ge 1. \tag{8}$$

We will restrict $0 \leq \theta < 1$, although, following a similar argument as Johansen (2008, section 3.1), it can be shown that the invertibility region of the MA polynomial $(1 - \theta B_d)$ is $(1 - 2^d)^{-1} < \theta < 1$.

The above specification provides an interesting extension of ES to the fractional case. When $d \in [0.5, 1.5)$, the process admits the decomposition into a fractional noise component and a WN component:

$$y_t = m_t + e_t, \quad m_t = \frac{1-\theta}{(1-B)^d}\xi_t, \quad e_t = \theta\xi_t.$$

In terms of the observations,

$$m_t = \frac{1-\theta}{1-\theta B_d} y_t = \sum_{j=0}^{\infty} w_j^* y_{t-j},$$

with weights given by

$$w_j^* = (1-\theta)u_j, \quad u_j = \theta \sum_{k=1}^{j} (-1)^{k+1} \frac{(d)_k}{k!} u_{j-k}, \quad j > 0, \quad u_0 = 1,$$

by application of Gould's formula (Gould, 1974), which yields ES in the case $d = 1$.

Interestingly, $m_t = y_t - \theta\xi_t$ is a weighted linear combination of y_t and its one-step-ahead prediction:

$$m_t = (1-\theta)y_t + \theta\tilde{y}_{t|t-1}.$$

The one-step-ahead prediction is in turn a weighted average of the current and past values of the component m_t:

$$\begin{aligned}
\tilde{y}_{t+1|t} &= \tilde{m}_{t+1|t} \\
&= \textstyle\sum_{j=1}^{\infty} \beta_j m_{t+1-j} \\
&= \textstyle\sum_{j=1}^{\infty} (-1)^{j+1} \frac{(d)_j}{j!} [(1-\theta)y_{t+1-j} + \theta\tilde{y}_{t+1-j|t-j}].
\end{aligned} \tag{9}$$

This expression encompasses the ES formula, which arises for $d = 1$, in which case $\tilde{y}_{t+1|t} = m_t$. Moreover, for any d the forecasts of the series are coincident with those of the component m_t.

4.3 Fractional Equal Root Integrated Moving Average (FerIMA) Process

In his seminal Biometrika paper, Hosking (1981) introduced the fractional equal–root integrated moving average process

$$(1-B)^d y_t = (1-\theta B)^d \xi_t, \;\; \xi_t \sim \mathrm{WN}(0,\sigma^2). \tag{10}$$

To our knowledge, this process has never had time series applications and has never been investigated in detail. Notice that the model is not identifiable if either $d = 0$ or $\theta = 1$. If $\theta = 1$, and $d \neq 0$, y_t is white noise. We will restrict θ to lie in the region $0 \leq \theta < 1$.

The coefficients of the infinite lag polynomial $\psi(B) = (1-\theta B)^d = \sum_{j=0} \psi_j B^j$ are obtained from applying Gould's formula (Gould 1974), giving

$$\psi_0 = 1, \quad \psi_j = (-1)^j \theta^j \frac{(d)_j}{j!}. \tag{11}$$

It is interesting to compare this formula with (8). It is argued that *coeteris paribus* the coefficients of (11) converge more rapidly to zero, due to the geometric decay of θ^j.

Applying Proposition 1 with $\psi(B) = (1-\theta B)^d$, we have $\psi_j(B) = (-1)^j\theta^j(1-\theta B)^{d-j}\frac{(d)_j}{j!}$, so that, for $j = 0, \dots, r-1$, the components of the decomposition are:

$$\begin{array}{rcl} u_{jt} & = & (-1)^j \theta^j (1-\theta)^{d-j} \frac{(d)_j}{j!} \frac{\xi_t}{(1-B)^{d-j}} \\ & = & (-1)^j \frac{(d)_j}{j!} \left(\frac{\theta}{(1-\theta}\right)^j \frac{(1-\theta)^d}{(1-\theta B)^d} (1-B)^j y_t. \end{array} \tag{12}$$

If $d \in [0.5, 1.5)$, the above decomposition reduces to $y_t = m_t + e_t$, where $m_t = u_{0t}$ is the following weighted moving average of the observations available at time t:

$$m_t = \left(\frac{1-\theta}{1-\theta B}\right)^d y_t = \sum_{j=0}^{\infty} w_j^* y_{t-j}, \tag{13}$$

with weights given recursively by

$$w_j^* = \theta \frac{j+d-1}{j} w_{j-1}^*, \; j \geq 1, \; w_0^* = (1-\theta)^d,$$

again by application of Gould's formula (Gould, 1974), or

$$w_j^* = (1-\theta)^d \frac{\theta^j}{j!} d^{(j)}, \tag{14}$$

where $d^{(j)}$ denotes the rising factorial $d^{(j)} = d(d+1)\cdots(d+j-1)$.

In terms of $\lambda = 1-\theta$,

$$w_j^* = w_j c_j, \;\; w_j = \lambda(1-\lambda)^j, \;\; c_j = \frac{\lambda^{d-1} d^{(j)}}{j!}, \tag{15}$$

which shows that the weights result from correcting the geometrically declining weights, $w_j = (1-\lambda)w_{j-1}, w_0 = 1$, by a factor depending on d, decreasing hyperbolically with j:

$$c_j = \left(1 - \frac{1-d}{j}\right) c_{j-1}, \quad c_0 = \lambda^{d-1}.$$

The importance of the correction term is larger, the smaller is λ, due to the factor λ^{d-1}. Moreover, if $d=1$, $\frac{\lambda^{d-1}d^{(j)}}{j!} = 1$.

The one-step-ahead forecast can be derived by taking the conditional mean of $y_{t+1} = \sum_{j=1}^{\infty} \beta_j y_{t+1-j} + \xi_{t+1} - \theta \sum_{j=1}^{\infty} \theta^j \beta_j \xi_{t+1-j}$. Hence,

$$\tilde{y}_{t+1|t} \;\; = \;\; \textstyle\sum_{j=1}^{\infty} (-1)^{j+1} \frac{(d)_j}{j!} \left[(1-\theta^j) y_{t+1-j} + \theta^j \tilde{y}_{t+1-j|t-j}\right], \tag{16}$$

which yields ES for $d=1$. It is also interesting to compare the form of this predictor with (9).

5 Whittle Pseudo-Maximum Likelihood Estimation

Let $\delta_0 = (d_0, \theta_0)$ and σ_0^2, denote the true parameters of the FIMA, FLagIMA and FerIMA models; their estimation is carried out in the frequency domain by Whittle pseudo-maximum likelihood estimation. The spectral density of the three models is written

$$f(\omega; \delta_0, \sigma_0^2) = \frac{\sigma^2}{2\pi} k(\omega; \delta), \omega \in [-\pi, \pi],$$

where

- for the FIMA model,

$$k(\omega;\delta) = \frac{1+\theta^2-2\theta\cos\omega}{[2(1-\cos\omega)]^d}; \tag{17}$$

- for the FLagIMA model,

$$k(\omega;\delta) = \frac{(1-\theta)^2+[2(1-\cos\omega)]^d\theta^2+2[2(1-\cos\omega)]^{d/2}\theta(1-\theta)\cos(d\varphi(\omega))}{[2(1-\cos\omega)]^d}, \tag{18}$$

with $\varphi(\omega) = \arctan\left(\frac{\sin(\omega)}{1-\cos(\omega)}\right)$;

- for the FerIMA model,

$$k(\omega;\delta) = \left(\frac{1+\theta^2-2\theta\cos\omega}{2(1-\cos\omega)}\right)^d. \tag{19}$$

Notice that $\int_{-\pi}^{\pi}\ln k(\omega;\delta)\mathrm{d}\omega = 0$, which implies that σ^2 is functionally independent of δ and that σ^2 is the prediction error variance of y_t, by the Kolmogorov-Szegö formula.

Given a time series realisation $\{y_t, t = 1, 2, \ldots, n\}$, and letting $\omega_j = \frac{2\pi j}{n}$, $j = 1, \ldots, n-1$, denote the Fourier frequencies (we exclude the zero frequency), the tapered periodogram is defined as

$$I_p(\omega_j) = \frac{1}{2\pi\sum_{t=1}^{n} h_t^2}\left|\sum_{t=1}^{n} h_t(y_t-\bar{y})e^{-\imath\omega_j t}\right|^2,$$

where $\bar{y} = n^{-1}\sum_{t=1}^{n} y_t$ and $\{h_t\}_{t=1}^{n}$ is a taper sequence. As in Velasco and Robinson (2000), $\{h_t\}_{t=1}^{n}$ is obtained from the coefficients of the polynomial

$$\left(\frac{1-z^{[n/p]}}{1-z}\right)^p.$$

This is know as a Zhurbenko data taper of order p.

The vector δ_0 is estimated by minimising the profile pseudo-likelihood

$$Q(\delta) = \frac{p}{n}\sum_{j=1}^{\lfloor n/p\rfloor}\frac{2\pi I_p(\omega_{pj})}{k(\omega_{pj};\delta)},$$

$\tilde{\delta} = \text{argmin}_{\delta \in \Delta} Q(\delta)$; the estimate of the prediction error variance is then $\tilde{\sigma}^2 = Q(\tilde{\delta})$. Tapering aims at reducing the estimation bias that characterises the periodogram ordinates in the nonstationary case. However, to reduce the undesirable dependence it induces between adjacent periodogram ordinates, only the Fourier frequencies ω_j such that j is a multiple of p are used.

The following theorem proves the consistency and the asymptotic normality of the Whittle Estimator.

Theorem 1. *Let* $\delta_0 = (d_0, \theta_0)$*; assume* $d_0 \in (0, 1.5)$, $\theta_0 \in (0, 1)$, $\sigma_0^2 \in (0, \infty)$ *and* ξ_t *be a martingale difference sequence with* $E(\xi_t^2 | \mathcal{F}_{t-1}) = \sigma_0^2 < \infty$. *Let* $f(\omega) = (2\pi)^{-1}\sigma_0^2 k(\omega; \psi_0)$, *where* $k(\omega; \delta)$ *is given either by (17), (18), or (19). Then,*

$$\tilde{\delta} \rightarrow_p \delta_0, \quad \tilde{\sigma}^2 \rightarrow_p \sigma_0^2. \tag{20}$$

If we further assume $E(\xi_t^s | \mathcal{F}_{t-1}) < \infty, s = 3, 4$, *and in the FLagIMA case we restrict* d_0 *in the range* $(0, 3/4) \bigcup (1, 3/2)$,

$$\sqrt{n}(\tilde{\delta} - \delta_0) \rightarrow_d N(0, pK_p V),$$

with K_p *being the variance inflation factor due to tapering and*

$$V^{-1} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \delta} \ln k(\omega; \delta_0) \frac{\partial}{\partial \delta'} \ln k(\omega; \delta_0) \mathrm{d}\omega.$$

Proof. Appendix C provides a detailed proof, based on Velasco and Robinson (2000, VR henceforth). □

For the Zhurbenko data tapers, $K_2 = 1.05; K_3 = 1.00354$. The specialty of the result for the FLagIMA model arises from the fact that the slowly varying factor of its spectral density, $|\psi(e^{-\imath\omega})|^2$, is bounded and continuous in ω, but it is not differentiable at ω, and it satisfies a local Lipschitz condition of order d_0 around the zero frequency, making it unsufficiently smooth for $\tilde{\delta}$ to converge in distribution. Notice that we restrict the range of d_0 to a plausible range. For our volatility series we will see that we need to consider the nonstationary but mean reverting case.

For the analytical evaluation of the variance covariance matrix of the Whittle estimators and for comparing the spectral fit, the following expressions holding for the FIMA model,

$$\ln k(\omega; \delta) = -2 \sum_{k=1}^{\infty} \frac{d + \theta^k}{k} \cos(\omega k),$$

and the FerIMA model,

$$\ln k(\omega; \delta) = -2d \sum_{k=1}^{\infty} \frac{1 + \theta^k}{k} \cos(\omega k),$$

can be used.

6 Empirical Analysis: the Components of Realized Volatility and its Prediction

We consider the empirical problem of forecasting one-step-ahead the daily asset returns volatility, using realized measures constructed from high frequency data. The volatility proxy is the Realized Variance (5-minute) of 21 stock indices extracted from the database "OMI's realized measure library" version 0.2, produced by Heber, Lunde, Shephard, and Sheppard (2009). The series are available from 03/01/2000 to 1/10/2015, for a total of 4,115 daily observations.

6.1 Parameter estimates and components

Denoting by RV_t a generic realized volatility series, we focus on its logarithmic transformation, that is we take $y_t = \ln RV_t$. The Whittle likelihood estimates of the parameters of the three fractionally integrated models (FIMA, FLagIMA and FerIMA) are presented in table 1. The estimates are based on the untapered periodogram ($p = 1$ in section 5; using $p = 2$ did not change the estimates). In most occurrences, the estimates of the memory parameter $\tilde{d}$ lie in the nonstationary region and are systematically higher for the FLagIMA specification, whereas they are very similar for the FIMA and FerIMA models. The MA parameter θ is estimated to be positive and significantly different from zero at the 5% level for all specifications and all occurrences, with two notable exceptions, the *Nikkei 225* and the *Nasdaq 100* series, for which the estimates $\tilde{d}$ are also indistinguishable. The last three columns report the estimated innovations variance, which is very similar across the three specifications, so that the fit can be expected to be similar.

Table 1: Log-Realized volatility series. Whittle maximum likelihood estimates of the parameters of the FIMA, FLagIMA and FerIMA models using the full sample (4,115 daily observations from 03/01/2000 to 1/10/2015). The standard error are presented in parenthesis.

	$\tilde{d}$			$\tilde{\theta}$			$\tilde{\sigma}^2$		
Series	FIMA	FLagIMA	FerIMA	FIMA	FLagIMA	FerIMA	FIMA	FLagIMA	FerIMA
S&P 500	0.565	0.627	0.572	0.138	0.313	0.254	0.343	0.343	0.343
	(0.022)	(0.031)	(0.024)	(0.028)	(0.046)	(0.044)			
FTSE 100	0.572	0.643	0.587	0.130	0.321	0.256	0.241	0.240	0.241
	(0.021)	(0.030)	(0.025)	(0.029)	(0.046)	(0.047)			
Nikkei 225	0.520	0.547	0.521	0.031	0.113	0.063	0.241	0.241	0.241
	(0.020)	(0.030)	(0.020)	(0.026)	(0.063)	(0.050)			
DAX	0.574	0.641	0.585	0.126	0.308	0.241	0.257	0.257	0.257
	(0.021)	(0.029)	(0.024)	(0.028)	(0.046)	(0.046)			
Russel 2000	0.526	0.590	0.533	0.123	0.310	0.243	0.359	0.359	0.359
	(0.022	(0.032)	(0.024)	(0.028)	(0.051)	(0.047)			
All Ord.	0.538	0.720	0.608	0.303	0.649	0.608	0.363	0.362	0.362
	(0.027	(0.037)	(0.037)	(0.036)	(0.031)	(0.037)			
DJIA	0.552	0.636	0.566	0.166	0.386	0.319	0.374	0.374	0.373
	(0.023	(0.032)	(0.026)	(0.029)	(0.044)	(0.045)			
Nasdaq 100	0.547	0.569	0.548	0.025	0.091	0.047	0.250	0.250	0.250
	(0.019	(0.027)	(0.019)	(0.025)	(0.057)	(0.045)			
CAC 40	0.569	0.624	0.577	0.103	0.256	0.197	0.235	0.235	0.235
	(0.021	(0.030)	(0.023)	(0.028)	(0.050)	(0.047)			
Hang Seng	0.537	0.619	0.547	0.165	0.397	0.318	0.233	0.233	0.233
	(0.021)	(0.029)	(0.024)	(0.027)	(0.043)	(0.043)			
KOSPI C. I.	0.582	0.634	0.587	0.120	0.274	0.215	0.208	0.208	0.208
	(0.021)	(0.030)	(0.024)	(0.028)	(0.049)	(0.047)			
AEX Index	0.582	0.638	0.591	0.104	0.256	0.198	0.243	0.243	0.243
	(0.023)	(0.030)	(0.026)	(0.031)	(0.042)	(0.043)			
Swiss M. I.	0.621	0.690	0.637	0.162	0.334	0.284	0.181	0.181	0.181
	(0.020)	(0.030)	(0.023)	(0.028)	(0.054)	(0.050)			
IBEX 35	0.545	0.594	0.552	0.077	0.221	0.158	0.225	0.225	0.225
	(0.021)	(0.031)	(0.023)	(0.028)	(0.054)	(0.049)			
S&P CNX Nifty	0.546	0.597	0.553	0.086	0.234	0.173	0.295	0.294	0.295
	(0.021)	(0.034)	(0.027)	(0.029)	(0.046)	(0.051)			
IPC Mexico	0.485	0.605	0.512	0.162	0.470	0.381	0.405	0.404	0.404
	(0.021)	(0.034)	(0.023)	(0.027)	(0.065)	(0.054)			
Bovespa Index	0.492	0.539	0.496	0.064	0.210	0.140	0.259	0.259	0.259
	(0.021)	(0.034)	(0.023)	(0.027)	(0.065)	(0.054)			
S&P/TSX C. I.	0.533	0.610	0.546	0.126	0.340	0.261	0.274	0.273	0.274
	(0.020)	(0.030)	(0.024)	(0.027)	(0.048)	(0.048)			
Euro STOXX 50	0.514	0.579	0.524	0.075	0.256	0.172	0.348	0.347	0.348
	(0.020)	(0.032)	(0.023)	(0.028)	(0.057)	(0.056)			
FT Straits T. I.	0.545	0.639	0.567	0.135	0.376	0.294	0.158	0.158	0.158
	(0.021)	(0.031)	(0.026)	(0.029)	(0.047)	(0.050)			
FTSE MIB	0.568	0.629	0.577	0.121	0.290	0.229	0.246	0.246	0.246
	(0.021)	(0.030)	(0.024)	(0.028)	(0.048)	(0.046)			

For illustration purposes we have computed the volatility components arising from the decomposition $y_t = m_t + e_t$ implied by the three models for the logarithm of the SP500 realized volatility series. These are presented in figure 1. The left plots display the original time series and the level or long-run component m_t, respectively for the FIMA, FLagIMA and FerIMA specifications, whereas the short-run component is plotted on the right. While the FIMA and FerIMA models lead to virtually the same decomposition (the correlation between the short-run components is 0.99), the FLagIMA model has a stationary component with higher variance; recall that $e_t = \theta\xi_t$, and thus $e_t \sim \mathrm{WN}(0, \theta^2\sigma^2)$. As a consequence, the long-run shocks have smaller size than for the other models. The comparison of $\tilde{\sigma}^2 = Q(\tilde{\delta})$ in the last three columns of table 1 highlights that we are unable to select what representation is most suitable on the basis of the likelihood or an information criterion.

6.2 Predictive performance and model confidence set

A fundamental aim of the analysis of realized volatility series is the prediction of future volatility. We thus turn our attention to the comparison of the performance of the alternative predictors discussed in the previous sections and to the estimation of the model confidence set (MCS), in the sense specified by Hansen, Lunde and Nason (2011), which contains the best models with a given confidence according to their ability to predict the log-realized volatility one-step-ahead. Our model set comprises the following elements: the ES predictor with parameter λ resulting from estimating an MA(1) model for the first differences, $(1-B)y_t$, by maximum likelihood; RM1994, which is the standard exponential smoothing predictor with $\lambda = 0.06$; the RM2006 predictor, described in appendix A, and the heterogeneous autoregressive (HAR) model proposed by Corsi (2009). The latter is specified as follows:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_5 \bar{y}_{t-5} + \phi_{22}\bar{y}_{t-22} + \xi_t, \quad \xi_t \sim \mathrm{WN}(0, \sigma^2),$$

where

$$\bar{y}_{t-5} = \frac{1}{5}\sum_{j=1}^{5} y_{t-j}, \bar{y}_{t-22} = \frac{1}{22}\sum_{j=1}^{22} y_{t-j},$$

are respectively the average realized variance over the previous trading week and over the previous month. We also consider the FN model, $(1-B)^d y_t = \xi_t, \xi_t \sim \mathrm{WN}(0, \sigma^2)$ along with the three fractional exponential smoothing models. Hence, the collection of models is $\mathcal{M}^0 = \{$ES, RM1994, RM2006, HAR, FN, FIMA, FLagIMA, FerIMA$\}$.

We perform a rolling forecasts experiment such that, starting from time $n_0 = 1000$ we compute the one-step-ahead volatility prediction according to the seven methods and we proceed by adding one future observation and removing the initial one, until the end of the sample is reached, so that each forecast is based on a fixed number of observations, and re-estimating the parameters for each rolling window, in the case of the HAR and the fractionally integrated models. The experiment yields 3,174 one step ahead prediction errors for each forecasting methodology used in the comparative assessment.

Denoting by $\tilde{y}_{k,t|t-1}$ the prediction arising from method $\mathcal{M}_k$, where $k = 1, 2, \ldots, K$, and by $v_{k,t} = y_t - \tilde{y}_{k,t|t-1}$, the corresponding prediction error, $t = n_0 + 1, \ldots, n$, we compare the mean square forecast error,

$$\text{MSFE}_k = \frac{1}{n - n_0} \sum_{t=n_0+1}^{n} v_{k,t}^2.$$

The results for the 21 realized volatility series are reported in table 2, which displays the MSFE's expressed as a percent of that characterising the FN predictor, which is assumed as a benchmark (its MSFE is reported in the last column of the table). The evidence is that the FLagIMA predictor outperforms the others in most occurrences, although the differences with the FIMA and FerIMA predictors are small. We also notice that the ES predictor is more accurate than both RM1994 and RM2006; the average estimated value for the smoothing parameter λ is 0.38 (0.62 for θ), with the estimates ranging from 0.52 and 0.78, which suggest that for optimizing the performance of the ES predictor we should learn from the series the value of the smoothing constant.

The prediction errors arising from the rolling forecasting experiment have then been used to estimate the model confidence set, according to the methodology proposed by Hansen, Lunde and Nason (2011).

Denote by $d_{kl,t} = v_{kt}^2 - v_{lt}^2$ the quadratic loss differential at time t between predictors k and l, for all $k, l \in \mathcal{M}^0$, and define

$$\bar{d}_{kl} = \frac{1}{n - n_0} \sum_{t=n_0+1}^{n} d_{kl,t}, \quad \bar{d}_k = \frac{1}{K} \sum_{l=1}^{K} \bar{d}_{kl}.$$

The t-statistics associated to the null $H_0 : \text{E}(\bar{d}_{k\cdot}) = 0$ (equal forecast accuracy)

Table 2: Log-Realized volatility series. Rolling forecasting exercise: comparison of one-step-ahead predictive performance (based on 3114 forecast errors). The columns are the percentage MSFE ratios $\frac{\text{MSFE}_k}{\text{MSFE}_{FN}}$, where the benchmark in the denominator is the MSFE of the Fractional Noise model, reported in the last column of the table and k indexes the seven alternative predictors considered in the first seven columns.

Series	ES	RM1994	RM2006	HAR	FIMA	FLagIMA	FerIMA	FN MSE
S&P 500	97.93	130.62	128.51	113.28	95.33	**95.32**	95.27	0.388
FTSE 100	101.12	135.77	136.16	121.35	98.09	**97.87**	98.03	0.239
Nikkei 225	104.60	142.93	140.31	126.40	99.70	**99.66**	99.72	0.259
DAX	99.23	132.46	132.64	117.33	96.40	**96.26**	96.34	0.287
Russel 2000	102.80	132.76	130.15	118.22	**98.91**	99.01	98.94	0.344
All Ordinaries	87.51	97.74	97.95	93.20	86.76	**86.45**	**86.45**	0.400
DJIA	95.95	120.19	118.48	107.70	93.71	93.71	**93.66**	0.433
Nasdaq 100	103.78	145.20	142.19	127.40	99.46	**99.39**	99.44	0.264
CAC 40	101.35	139.01	137.32	122.01	98.01	**97.88**	97.96	0.261
Hang Seng	101.53	121.51	122.54	113.78	97.94	**97.90**	97.95	0.249
KOSPI Composite Index	101.89	133.69	133.71	119.75	**98.02**	98.10	98.05	0.209
AEX Index	100.99	142.86	141.26	122.57	98.09	97.98	**98.06**	0.261
Swiss Market Index	97.41	138.60	138.65	117.50	95.17	**95.01**	95.11	0.192
IBEX 35	101.39	137.82	134.52	122.73	97.77	**97.62**	97.70	0.250
S&P CNX Nifty	101.47	125.07	122.60	118.23	97.74	**97.61**	97.69	0.328
IPC Mexico	95.25	108.01	106.70	102.09	93.05	**92.97**	93.03	0.427
Bovespa Index	101.17	130.62	127.01	116.30	**97.43**	97.50	**97.43**	0.268
S&P/TSX Composite Index	101.53	123.16	123.25	115.54	98.15	**98.06**	98.16	0.338
Euro STOXX 50	101.23	128.50	125.90	118.57	97.40	**97.32**	97.38	0.385
FT Straits Times Index	99.16	127.72	129.00	115.11	97.50	**97.13**	97.37	0.145
FTSE MIB	99.80	132.37	130.87	117.49	96.82	**96.67**	96.75	0.270

versus the alternative $H_0 : \mathrm{E}(\bar{d}_{k.}) > 0$, is $t_k = \frac{\bar{d}_{k.}}{\hat{\mathrm{Var}}(\bar{d}_{k.})}$, where $\hat{\mathrm{Var}}(\bar{d}_{k.})$ is an estimate of $\mathrm{Var}(\bar{d}_{k.})$. If $\mathcal{M}$ is the current set of model under assessment, to test the hypothesis $H_{0,\mathcal{M}} : \mathrm{E}(\bar{d}_{k.}) = 0, \forall k \in \mathcal{M}$, i.e. all the models have the same predictive accuracy, we use the test statistic $T_{\max} = \max_{k \in \mathcal{M}} t_k$. Following the MCS algorithm defined by Hansen, Lunde and Nason (2011), we initially set $\mathcal{M} = \mathcal{M}^0$, and test $H_{0,\mathcal{M}}$ using the statistic $T_{\max}$, at the significance level $\alpha = 0.10$. The critical values are obtain by the block-bootstrap method, using a block length of 20 observations and 10,000 replications; the bootstrap replications are also used to estimate $\mathrm{Var}(\bar{d}_k)$. If $H_{0,\mathcal{M}}$ is accepted, then the MCS at level $1 - \alpha$ is $\widehat{\mathcal{M}}^*_{1-\alpha} = \mathcal{M}$; otherwise, we proceed to eliminate from the set the predictor for which the $t_{k.}$ statistic is a maximum, $k \in \mathcal{M}$, and iterate the procedure with the surviving predictors.

Table 3 reports the MCS p-values, computed according to section 2.3 of Hansen, Lunde and Nason (2011), so that model k belongs to the MCS at the level $1 - \alpha$ if and only if the corresponding p-value is greater or equal to α. The values in bold correspond to predictors in $\widehat{\mathcal{M}}^*_{0.90}$. The main evidence is that the FLagIMA model belongs to the 90% MCS in all occurrences (this is robust to different implementations of the bootstrap). On the contrary, the FIMA and FerIMA predictors are in $\widehat{\mathcal{M}}^*_{0.90}$ respectively in 14 and 16 cases out of 21, and are in $\widehat{\mathcal{M}}^*_{0.95}$ respectively in 19 and 20 cases. The RM1994, RM2006 and HAR predictors are never in the MCS and the FN model belongs to $\widehat{\mathcal{M}}^*_{0.90}$ only once.

7 Conclusions

The long memory extensions of the popular exponential smoothing predictor that have been considered in this paper prove very effective for forecasting realized volatility. Among the three proposed models, the fractional lag integrated moving average model (FLagIMA) stands out prominently, as it outperforms its competitors and has a permanent seat in the model confidence set for the 21 realized variance series of the Oxford-Man Institute Realized library.

From the point of view of signal extraction, the FLagIMA has the unique feature of decomposing the series into a fractional noise component and a purely unpredictable one, whose standard deviation is proportional to the moving average parameter.

In conclusion, the FLagIMA model is a simple and parsimonious model, that can play a useful role in forecasting volatility and extracting its underlying level.

Table 3: MCS p-values for eight forecasting methods. The predictors in $\widehat{\mathcal{M}}^*_{90\%}$ are in bold.

Series	ES	RM1994	RM2006	HAR	FN	FIMA	FLagIMA	FerIMA
S&P 500	0.000	0.000	0.000	0.000	0.001	**0.800**	**0.800**	**1.000**
FTSE 100	0.001	0.000	0.000	0.000	0.005	0.051	**1.000**	0.051
Nikkei 225	0.000	0.000	0.000	0.000	0.507	**0.654**	**1.000**	**0.182**
DAX	0.000	0.000	0.000	0.000	0.000	**0.332**	**1.000**	**0.332**
Russel 2000	0.000	0.000	0.000	0.000	**0.132**	**1.000**	**0.341**	**0.599**
All Ordinaries	0.011	0.000	0.000	0.000	0.000	0.072	**1.000**	**0.996**
DJIA	0.000	0.000	0.000	0.000	0.000	**0.889**	**0.889**	**1.000**
Nasdaq 100	0.000	0.000	0.000	0.000	0.054	0.088	**1.000**	**0.139**
CAC 40	0.002	0.000	0.000	0.000	0.002	0.090	**1.000**	**0.137**
Hang Seng	0.005	0.000	0.000	0.000	0.066	**0.814**	**1.000**	**0.534**
KOSPI Composite Index	0.000	0.000	0.000	0.000	0.011	**1.000**	**0.437**	**0.437**
AEX Index	0.002	0.000	0.000	0.000	0.002	**0.212**	**1.000**	**0.212**
Swiss Market Index	0.000	0.000	0.000	0.000	0.000	**0.191**	**1.000**	**0.191**
IBEX 35	0.001	0.000	0.000	0.000	0.002	0.028	**1.000**	**0.093**
S&P CNX Nifty	0.000	0.000	0.000	0.000	0.000	**0.197**	**1.000**	**0.197**
IPC Mexico	0.001	0.001	0.001	0.001	0.001	**0.711**	**1.000**	**0.711**
Bovespa Index	0.000	0.000	0.000	0.000	0.000	**1.000**	**0.399**	**0.913**
S&P/TSX Composite Index	0.002	0.000	0.000	0.000	0.003	**0.341**	**1.000**	0.079
Euro STOXX 50	0.028	0.000	0.000	0.000	0.028	**0.720**	**1.000**	**0.720**
FT Straits Times Index	0.025	0.000	0.000	0.000	0.025	0.019	**1.000**	0.014
FTSE MIB	0.001	0.000	0.000	0.000	0.001	0.052	**1.000**	0.073

A The Riskmetrics 2006 methodology

The new RM methodology, referred to as RM2006 (see Zumbach, 2007), extends the 1994 methodology to the long memory case, by computing the one-step-ahead volatility prediction, denoted $\tilde{y}_{t+1|t}$, as a weighted sum of K exponentially weighted moving averages with smoothing constants $\lambda_k, k = 1, \dots, K$:

$$\tilde{y}_{t+1|t} = \sum_{k=1}^{K} w_k \tilde{y}_{t+1|t}^{(k)}, \qquad \tilde{y}_{t+1|t}^{(k)} = \lambda_k y_t + (1-\lambda_k)\tilde{y}_{t|t-1}^{(k)}.$$

The weights decay according to $w_k \propto 1 - \frac{\ln \tau_k}{\ln \tau_0}$, where $\tau_k, k = 1, \dots, K$, are time horizons, chosen according to the geometric sequence: $\tau_k = \tau_1 \rho^{k-1}, k = 2, \dots, K, \ \ \tau_0 = 1560, \ \ \tau_1 = 4, \ \ \rho = \sqrt{2}$. The smoothing constants are related to the time horizons via $\lambda_k = 1 - \exp(-\tau_k^{-1})$. In the current empirical implementation the value K is chosen so that $\tau_K = 512$, which gives $K = 15$.

The weights attached to past observations are an arithmetic weighted average of those arising from EWMAs with different smoothing constants and are no longer a geometric sequence.

B Proof of Proposition 1

Let

$$c_j = \frac{(-1)^j}{j!} \left.\frac{d^j \psi(z)}{dz^j}\right|_{z=1} = (-1)^j \sum_{k=j}^{\infty} \binom{k}{j} \psi_k,$$

and define the polynomial

$$\psi_k(z) = \sum_{s=0}^{\infty} c_{k+s}(1-z)^s,$$

so that $\psi_k(1) = c_k$. The absolute convergence of $\psi(z)$ in the region containing the point $z = 1$ implies that there exists a neibourhood of $z = 1$ in which $\psi(z)$ has the power series expansion $\psi(z) = \sum_{j=0}^{\infty} \psi_j(1)(1-z)^j$, and that the polynomials $\psi_j(z)$ have the same radius of convergence (see, e.g., Apostol, 1974, theorems 9.22 and 9.23), so that $\psi_j(1)$ is finite, for all j.

The above recursive formula for the polynomial $\psi_{j-1}(z)$ is proved as follows:

$$\begin{aligned}\psi_{j-1}(z) - \psi_{j-1}(1) &= \textstyle\sum_{s=1}^{\infty} c_{j-1+s}(1-z)^s \\ &= (1-z)\textstyle\sum_{s=0}^{\infty} c_{j+s}(1-z)^s \\ &= (1-z)\psi_j(z).\end{aligned}$$

Using these results, we write $(1-B)^d y_t = \psi(1)\xi_t + \psi_1(1)(1-B)\xi_t + \cdots + \psi_{r-1}(1)(1-B)^{r-1}\xi_t + \psi_r(B)(1-B)^r\xi_t$. The decomposition follows straightforwardly. The component e_t converges in mean square to a causal stationary random process by theorem 4.10.1 in Brockwell and Davies (1991). Note that $-0.5 \leq d - r < 0.5$.

C Proof of Theorem 1

Let $h(\omega; s) = (1 - e^{-\imath\omega})^s$, so that $|h(\omega; s)| = O(|\omega|^s)$ and $\frac{d}{d\omega}h(\omega; s) = O(|\omega|^{s-1})$, as $\omega \to 0$. Notice that $k(\omega; \delta) = |h(\omega; -d)|^2|\psi(\omega)|^2$, $\psi(\omega) = \sum_j \psi_j e^{-\imath\omega j}$, $\sum_j \psi_j^2 < \infty$, for the three specifications considered. Under the stated conditions, assumptions A1-A4 in VR are verified. In particular, (A1): δ_0 is an interior point of Δ; for $\omega \to 0$, $k(\omega, \delta) \sim G_\delta|\omega|^{-2d}$, with $0 < G_\delta < \infty$, continuous and positive at all $\omega \neq 0$ and $\delta \in \Delta$; $\delta_1 \neq \delta_2$ implies $k(\omega, \delta_1) \neq k(\omega, \delta_2)$ on a set of positive Lebesue measure. (A2): $\partial k(\omega, \psi)/\partial\omega$ exists and it is continuous for all $\omega \neq 0, \delta$, and $\partial k(\omega; \delta)/\partial\omega = O(|\omega|^{-2d-1})$, as $\omega \to 0$. (A4): the process $u_t^{(r)} = (1-B)^r y_t$, $r = \lfloor d_0 + 1/2 \rfloor$, has the representation $u_t^{(r)} = \alpha(B)\xi_t$, with $\alpha(B) = \sum_0^\infty \alpha_j B^j$ given by $\alpha(B) = (1-B)^{r-d}\psi(B)\sigma$, with $\sum_j \alpha_j^2 < \infty$, and we have assumed $\mathrm{E}(\xi_t|\mathcal{F}_{t-1}) = 0$ and $\mathrm{E}(\xi_t^2|\mathcal{F}_{t-1}) = \sigma^2$. Hence, the convergence in probability of the Whittle estimators (20) follows from Theorem 1 in VR.

For proving the asymptotic normality in the untapered case ($p = 1$), the tranfer function $\alpha(\omega) = \sum_j \alpha_j e^{-\imath\omega j}$, which has the representation $\alpha(\omega) = h(\omega; r-d)\psi(\omega)$, satisfies assumption A5 of VR: in particular, it is differentiable in ω at all $\omega \neq 0$ and δ, it is such that $|\alpha(\omega)| = O(|\omega|^{r-d})$ as $\omega \to 0$, and

$$\frac{\partial}{\partial\omega}\alpha(\omega) = \frac{\partial h(\omega; r-d)}{\partial\omega}\psi(\omega) + h(\omega; r-d)\frac{\partial\psi(\omega)}{\partial\omega}$$

is $O(|\omega|^{r-d-1})$, as $\psi(\omega)$ and its derivative are continuous and bounded and $\frac{\partial}{\partial\omega}h(\omega; r-d) = O(|\omega|^{r-d-1})$. When a taper is applied, the stronger smoothness condition in VR's assumption A6 is satisfied by the FIMA and FerIMA models. As it is evident from (17), (18), or (19), as $\omega \to 0$

$$k(\omega;\delta) = G_\delta |\omega|^{-2d} \left(1 + E_\delta |\omega|^\beta + o(|\omega|^\beta)\right)$$

for $0 < G_\delta, E_\delta < \infty$, with $\beta = 2$ for the FIMA and the FerIMA models. As a result, $\partial k(\omega;\delta)/\partial\omega$ is Lipschitz($\beta-1$). The FLagIMA model is such that the slowly varying component of the spectrum, $L(\omega) = |\omega|^{2d} k(\omega;\delta)$, is not differentiable at $\omega = 0$, and it satisfies a Lipschitz condition of degree d_0 around $\omega = 0$, i.e. $L(\omega) = O(|\omega|^{d_0})$. Hence, VR's assumption A6 holds in the FLagIMA case only for $d_0 > 1$.

Assumptions A7 in VR holds, as we have assumed $\mathrm{E}(\xi_t^s|\mathcal{F}_{t-1}) < \infty$, for $s = 3, 4$ and for all t. The function $k^{-1}(\omega;\delta) = |h(\omega;d)|^2 |\psi(\omega)|^{-2}$ is twice continuously differentiable with respect to $\delta \in \Delta$, for all $\omega \neq 0$ and δ, with $\frac{\partial}{\partial\delta\partial\delta'} k^{-1}(\omega;\delta) = O(|\omega|^{2d-\epsilon})$, for each $\epsilon \geq ||\delta - \delta_0||$, and these derivatives are continuously differentiable in $\omega \neq 0$, so that $\frac{\partial}{\partial\omega\partial\delta_k\partial\delta_l} k^{-1}(\omega;\delta) = O(|\omega|^{2d-1-\epsilon})$ (assumption A8). Finally, $\int_{-\pi}^{\pi} k^{-1}(\omega;\delta)\mathrm{d}\omega$ and $\int_{-\pi}^{\pi} \ln k(\omega;\delta)\mathrm{d}\omega$ are twice continuously differentiable in $\delta \in \Delta$ under the integral sign (assumption A9). Hence, all the regularity conditions for the asymptotic normality of $\tilde{\delta}$ are satisfied.

References

[1] Adrian, T., and Rosenberg, J. (2008). Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk. *The Journal of Finance*, 63, 2997-3030.

[2] Andersen, T.G., and Bollerslev, T. (1997). Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. *The Journal of Finance*, 52, 975–1005.

[3] Andersen, T.G., Bollerslev, T., Diebold, F.X., and Ebens, H (2001a). The distribution of stock returns volatilities. *Journal of Financial Economics* , 61, 43-76.

[4] Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (2001b). The Distribution of Exchange Rate Volatility. *Journal of the American Statistical Association*, 96, 42–55.

[5] Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 579-625.

[6] Apostol, T.M. (1974). *Mathematical Analysis, 2nd ed.*. Reading, Mass: Addison-Wesley.

[7] Ariño, M.A., and Marmol, F. (2004). A Permanent-Transitory Decomposition for ARFIMA Processes. *Journal of Statistical Planning and Inference*, 124, 87–97.

[8] Bandi, F., Perron, B., Tamoni, A., and Tebaldi, C. (2015). The scale of predictability. Cirano Séries Scientifique, n. 2015s-21, Montréal.

[9] Barndorff-Nielsen, O., and Shephard, N. (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64, 253-280.

[10] Barndorff-Nielsen, O., and Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17, 457-477.

[11] Beveridge, S., and Nelson, C.R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, 7, 151–174.

[12] Bollerslev, T. and Wright, J.H. (2000). Semiparametric Estimation of Long-Memory Volatility Dependencies: The Role of High Frequency Data. *Journal of Econometrics*, 98, 81–106.

[13] Brockwell, P.J., and Davis, R.A. (1991), *Time Series: Theory and Methods*, Springer-Verlag, New York.

[14] Calvet, L., and Fisher, A. (2001). Forecasting multifractal volatility. *Journal of Econometrics*, 105, 27–58.

[15] Comte, F., and Renault, E. (1998). Long memory in continuous time stochastic volatility models. *Mathematical Finance*, 8, 291-323.

[16] Corsi, F. (2009). A simple long memory model of realized volatility. *Journal of Financial Econometrics*, 7, 174-196.

[17] Corsi, F., Fusari, N., and La Vecchia, D. (2012). Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics*, 107, 284–304.

[18] Corsi, F., and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30, 368–380. Journal of Financial Econometrics 7, 174196.

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

[20] Ding, Z., and Granger, C. W. J. (1996). Modeling Volatility Persistence of Speculative Returns: A New Approach, *Journal of Econometrics*, 73, 185–215.

[21] Ding, Z., Granger, C. W. J., and Engle, R.F. (1993). A Long Memory Property of StockMarket Returns and a New Model. *Journal of Empirical Finance*, 1, 83–106.

[22] Engle, R.F. (1995). *ARCH: Selected Readings*. Oxford University Press, 1995.

[23] Engle, R.F., Ghysels, E., and Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economic and Statistics*, 95, 776–797.

[24] Engle, R.F., and Lee (1999). A Long-Run and Short-Run Component Model of Stock Return Volatility. In Engle, R.F., and White, A. (eds), *Cointegration, Causality, and Forecasting*, Oxford University Press. Oxford, UK.

[25] Gallant, A.R., Hsu, C.T. and Tauchen, G. (1999). Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance, *Review of Economics and Statistics*, 81, 617–631.

[26] Gardner, E.S. (2006). Exponential smoothing: the state of the art. Part II. *International Journal of Forecasting*, 22, 637–666.

[27] Giraitis, L., Koul, H.L., Surgailis, D. (2013), *Large Sample Inference for Long Memory Processes*, Imperial College Press.

[28] Gould, H. W. (1974). Coefficient Identities for Powers of Taylor and Dirichlet Series. *American Mathematical Monthly*, 81, 3–14.

[29] Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238.

[30] Granger, C.W.J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48, 213-228.

[31] Granger, C.W.J., and Hyung, N. (2004). Occasional Structural Breaks and Long Memory with an Application to the S&P 500 absolute returns. *Journal of Empirical Finance*, 11, 399 – 421.

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

[33] Hansen P.R., Lunde A. and Nason J.M. (2011). The model confidence set. *Econometrica*, 79, 453–497.

[34] Heber, G., Lunde, A., Shephard, N. and Sheppard, K. K. (2009). *OMIs realised measure library. Version 0.2*, Oxford-Man Institute, University of Oxford.

[35] Hyndman, R.J., Koehler, A.B., Ord, J.K., and Snyder, R.D. (2008). *Forecasting with Exponential Smoothing. The State Space Approach*. Springer Series in Statistics, Springer.

[36] Hosking, J.R.M. (2006). Fractional differencing. *Biometrika*, 88, 168–176.

[37] Hurvich, C.M., and Ray, B.K. (2003). The Local Whittle Estimator of Long-Memory Stochastic Volatility. *Journal of Financial Econometrics*, 1, 445–470.

[38] Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory*, 24, 651–676.

[39] Majewski, A. A., Bormetti, G., and Corsi, F. (2015). Smile from the past: A general option pricing framework with multiple volatility and leverage components. *Journal of Econometrics*, 187, 521-531.

[40] Merton, R. C. (1980). On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics*, 8, 323-61.

[41] Mikosch, T., and Starica, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics*, 86, 378–390.

[42] Müller, U., Dacorogna, M., Davé, R., Olsen, R., Pictet, O., and von Weizsacker, J. (1997). Volatilities of different time resolutions - analyzing the dynamics of market components. *Journal of Empirical Finance*, 4, 213-239.

[43] RiskMetrics Group (1996). RiskMetrics Technical Document, New York: J.P. Morgan/Reuters

[44] Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press.

[45] Velasco, C., and Robinson, P.M. (2000), Whittle pseudo-maximum likelihood estimation for non-stationary time series, *Journal of the American Statistical Association*, 95, 1229-1243.

[46] Zumbach, G.O. (2007). The Riskmetrics 2006 Methodology. Available at SSRN: http://ssrn.com/abstract=1420185 or http://dx.doi.org/10.2139/ssrn.1420185.

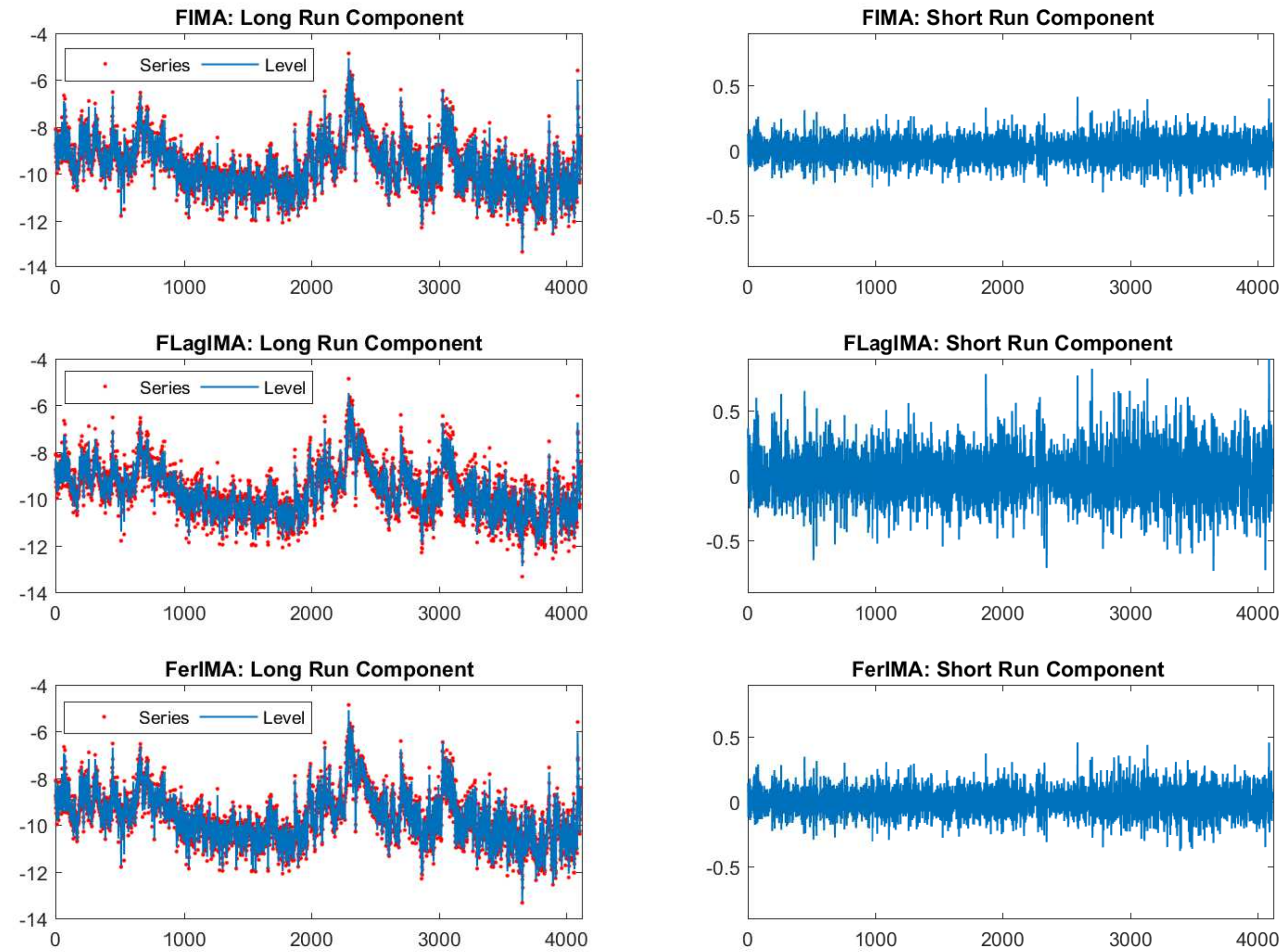

Figure 1: Logarithms of 5-minutes daily realized volatility for the SP500 index (red points) and its decomposition, $y_t = m_t + e_t$, into a long-run, or level component (m_t, solid line, left plots) and the short-run component (e_t, right plots), for the three specifications.