Exponential Smoothing, Long Memory and Volatility Prediction

Proietti, Tommaso

Dipartimento di Economia e Finanza

10 July 2014

Exponential Smoothing, Long Memory and Volatility Prediction

Tommaso Proietti

Department of Economics and Finance, University of Rome Tor Vergata, Italy, and CREATES, Aahrus University, Denmark

Abstract

Extracting and forecasting the volatility of financial markets is an important empirical problem. Time series of realized volatility or other volatility proxies, such as squared returns, display long range dependence. Exponential smoothing (ES) is a very popular and successful forecasting and signal extraction scheme, but it can be suboptimal for long memory time series. This paper discusses possible long memory extensions of ES and finally implements a generalization based on a fractional equal root integrated moving average (FerIMA) model, proposed originally by Hosking in his seminal 1981 article on fractional differencing. We provide a decomposition of the process into the sum of fractional noise processes with decreasing orders of integration, encompassing simple and double exponential smoothing, and introduce a low-pass real time filter arising in the long memory case. Signal extraction and prediction depend on two parameters: the memory (fractional integration) parameter and a mean reversion parameter. They can be estimated by pseudo maximum likelihood in the frequency domain. We then address the prediction of volatility by a FerIMA model and carry out a recursive forecasting experiment, which proves that the proposed generalized exponential smoothing predictor improves significantly upon commonly used methods for forecasting realized volatility.

Keywords: Realized Volatility. Signal Extraction. Permanent-Transitory Decomposition. Fractional equal-root IMA model.

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 daily realized volatility measures, see e.g. McAleer and Medeiros (2008), or squared returns (possibly after a logarithmic transformation of the series). This paper is based on the latter approach and takes as well established fact the presence of long range dependence as a characteristic feature of volatility, see Ding, Granger and Engle (1993), Bollerslev and Wright (2000), Taylor (2005, section 12.9), Andersen et al. (2001), Hurvich and Ray (2003), among others, though this point is not without controversy (see, e.g. Diebold and Inoue, 2001, and Granger and Hyung, 2004, who illustrate that stochastic regime switching and occasional breaks can mimic long range dependence).

Long memory is often modelled by a parametric model featuring fractional integration. Letting $\{y_t\}$ denote a univariate random process and B the backshift operator, $B^k y_t = y_{t-k}$, a basic model (Granger and Joyeux, 1980, and Hosking, 1981) is the fractional differenced noise,

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

where d is the memory parameter, $\text{WN}(0, \sigma^2)$ denotes white noise, a sequence of uncorrelated random variables with zero mean and variance σ^2, and, for non-integer $d > -1$, the fractional differencing operator is defined according to the binomial expansion as

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

where $\Gamma(\cdot)$ is the gamma function. We shall denote this by $y_t \sim \text{FN}(d)$. 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. The model can be extended so that ξ_t is replaced by a stationary short memory autoregressive moving average (ARMA), leading to the important class of ARFIMA (autoregressive, fractionally integrated, moving average) processes. For comprehensive treatments of long memory time series see Palma (2007), Giraitis, Koul and Surgailis (2012) and Beran et al. (2013).

The long-memory feature can also be approximated by the linear combination of short memory autoregressive processes, as in Gallant, Hsu and Tauchen (1999) (it should be considered that, according to Granger's (1980) seminal result, a long memory process can result from the contemporaneous aggregation of infinite first order AR processes). The heterogeneous autoregressive model (HAR) model by Corsi, based on a constrained long autoregressive model depending only on three parameters, associated to volatility components over different horizons (daily, weekly and monthly), can mimic the long memory feature and has proved extremely practical and successful in predicting realized volatility.

At the same time, there is growing interest in decomposing volatility into its short and long run components. Engle and Lee (1999) introduced a component GARCH model such that the conditional variance is the sum of two AR processes. 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 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. Colacito, Engle and Ghysels (2011) formulate a multivariate GARCH-MIDAS component model for dynamic correlations. Related recent papers in the multivariate framework dealing with volatility components are Hafner and Linton (2010), Bauwens, Hafner and Pierret (2013), and Amado and Teräsvirta (2013, 2014).

Exponential smoothing (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; it has also been extensively applied to forecasting volatility and value at risk. Its success is not only due to its simplicity, but also for constituting a remarkably close approximation to the volatility extracted by popular parametric methods such as GARCH models, when asset returns are modelled. It has thus become a reference and it has been incorporated in the widely popular RiskMetrics methodology. However, it is inadequate for handling long memory in volatility, which is an important feature of realized volatility series and other volatility proxies, such as squared or absolute returns. ES is a linear filter with weights declining according to a geometric progression with given ratio, usually fixed at 0.94, as advocated by Riskmetrics (see RiskMetrics Group, 1996). 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 weighs decaying at a hyperbolic, rather than geometric, rate, by combining several ES filters with different smoothing constants.

This paper aims at evaluating the RM2006 filter from the point of view of filtering theory and it will propose a sensible direction for extending ES to the class of long memory processes. We start from the consideration of several alternative models that generalize aspects of the ES predictor, and we end up by looking into the so called *fractional equal-root integrated moving average* (Fer-

IMA) model, originally proposed by Hosking (1981): in the closing of his seminal paper, Hosking mentions two fractionally integrated processes that can prove useful in applications: the first is the generalized fractional Gegenbauer process, see Gray, Woodward and Zhang (1989), which has found many applications in the modelling of stationary stochastic cycles with long range dependence. The second is the FerIMA process (to be introduced in a later section), which according to Hosking (p. 175, last paragraph) “as a forecasting model it corresponds to fractional order multiple exponential smoothing”. The FerIMA process is a particular case of a fractional power process which has the representation $(1-B)^d y_t = \left(\frac{\theta(B)}{\phi(B)}\right)^d \xi_t$, where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials in the lag operator B with roots outside the unit circle. To the author’s knowledge not much work has been done in this area, although there is some related work by Pillai, Shitan and Peiris (2012) and the class of Spectral ARMA models considered in Proietti and Luati (2014).

In the sequel we explore the characteristics of this process and propose a decomposition and corresponding filters that can be viewed as a generalization of exponential smoothing for fractionally integrated time series. The main result is the decomposition of a process integrated of order $d > 0$ into the sum of fractional noise processes of decreasing orders d, $d-1$, $d-2$, etc. plus a stationary remainder term. The first component generalizes the ES filter to any value of the fractional differencing parameter d and has the following features: (i) it encompasses the traditional ES filter, as well as double exponential smoothing, when the order of integration is 2. (ii) In the long memory case the filter weights have a interesting analytic form, resulting from the multiplication of coefficients decaying at a geometric rate (as in traditional ES) and correction factors that decline hyperbolically. (iii) For a FerIMA process it yields a fractional noise process with the same integration order. (iv) It captures the persistent (long-run if $d > 1$) behaviour of the series as it can be characterised as a low-pass filter.

We address the issue of the empirical relevance of the volatility predictor arising from the FerIMA model by performing a recursive forecasting experiment and documenting that it almost systematically outperforms the RiskMetrics predictor as well as th HAC model.

The paper is structured in the following way. Section 2 reviews the essential ES predictor and signal extraction filter. We next consider the extension proposed by RiskMetrics for long memory volatility proxies (section 3), as an attempt to accommodate the long memory feature by combining ES predictors with different smoothness. Section 4 discusses various possible extensions of ES for fractionally integrated processes: the fractional local level mode, the ARFIMA(0,d, 1) predictor, the fractional lag IMA(1,1) process, and it eventually concentrates on the FerIMA model, proposing a decomposition into fractional noise components of decreasing orders, the first of which yields a fractional ES (FES) filter. The nature of the decomposition is discussed in section 5, as a special case of a generalized Beveridge and Nelson (1981) decomposition, whereas section 6 analyses the properties of the FES filter as a low-pass filter in the frequency domain. In section 7 we briefly discuss maximum likelihood estimation of the parameters of the FerIMA model and prediction. Our empirical assessment of the FerIMA predictor and signal extraction filter is presented in section 8. Section 9 concludes the paper.

2 Exponential smoothing

Let y_t denote a time series stretching back to the indefinite past. The method known as exponential smoothing (ES) yields the l-steps-ahead predictor of y_t, $l = 1, 2, \ldots$,

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

The predictor depends on the smoothing constant, λ, which takes vales in the range (0,1). The above expression is an exponentially weighted moving average (EWMA) of the available observations. The weights received by past observations decline 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 (adaptive-expectations formulae) :

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

If a finite realisation is available, $\{y_t, t = 1, 2, \ldots n\}$, and we denote by $\tilde{y}_{1|0}$ the initial value of (2), then

$$\tilde{y}_{t+l|t} = \lambda \sum_{j=0}^{t-1} (1-\lambda)^j y_{t-j} + (1-\lambda)^t \tilde{y}_{1|0}. \tag{3}$$

If $\tilde{y}_{1|0} = y_1$ (in which case $\tilde{y}_{2|1} = y_1$), the weight of the first observation is increased, yielding $\tilde{y}_{t+l|t} = \lambda \sum_{j=0}^{t-2} (1-\lambda)^j y_{t-j} + (1-\lambda)^{t-1} y_1$, and the weights attached to the past and current observations sum up to unity. Alternatively, we could start the recursion with the average of the first t observations, $\tilde{y}_{1|0} = \sum_{j=0}^{t-1} y_{t-j}/t$. The same solution is obtained by setting $\tilde{y}_{1|0} = 0$ and rescaling the weights so that they sum to one: $\tilde{y}_{t+l|t} = (1-(1-\lambda)^{t-1})^{-1}\lambda \sum_{j=0}^{t-1}(1-\lambda)^j y_{t-j}$. Comprehensive reviews of exponential smoothing and its extensions are provided by Gardner (1985), 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

$$(1-B)y_t = (1-\theta B)\xi_t, \quad \xi_t \sim \mathrm{WN}(0, \sigma^2), \tag{4}$$

where B is the backshift operator, $0 < \theta < 1$ is the moving average parameter, and WN denotes white noise, a sequence of uncorrelated random variables with zero mean and variance σ^2. In particular

$$\tilde{y}_{t+l|t} = y_t - \theta\xi_t = \frac{1-\theta}{1-\theta B} y_t, \tag{5}$$

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 IMA(1,1) process admits the decomposition, known as the Beveridge and Nelson (1981) decomposition, into a permanent component m_t, represented by a random walk (RW) process, $m_t = m_{t-1} + (1-\theta)\xi_t$, and a transitory purely random component, $e_t = \theta\xi_t$, so that $y_t = m_t + e_t$. The permanent component is equal to the long run prediction of the series at time t and it is thus measurable at time t by $m_t = (1-\theta)(1-\theta B)^{-1} y_t$.

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).

The ES predictor is useful also in the presence of model misspecification: see Cox (1961) and Tiao and Xu (1993). Thus, it has potential also for a long memory process, if the parameter θ is estimated so as to minimise the multistep prediction error variance.

3 Riskmetrics 2006

Exponential smoothing has been used for forecasting and extracting the level of volatility from either squared returns or realized volatility measures according to the RiskMetrics methodology developed by J.P. Morgan. The Riskmetrics 1994 methodology (RM1994, see RiskMetrics Group, 1996) is based on a single EWMA with parameter $\lambda = 0.06$, or, equivalently, $\theta = 0.94$.

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 $\lambda_k, k = 1, \ldots, K$:

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

The weights decay according to

$$w_k \propto 1 - \frac{\ln \tau_k}{\ln \tau_0},$$

where $\tau_k, k = 0, 1, \ldots, K$, are time horizons, chosen according to the geometric sequence:

$$\tau_k = \tau_1 \rho^{k-1}, k = 2, \ldots, K, \quad \tau_0 = 1560, \quad \tau_1 = 4, \quad \rho = \sqrt{2}.$$

The smoothing constants are related to the time horizons via

$$\lambda_k = 1 - \exp(-\tau_k^{-1}).$$

Finally, in the current empirical implementation the value K is chosen so that $\tau_K = 512$, which gives $K = 15$. Table 1 reports the values of the time horizons, the smoothing constants of the K EWMAs, and the combination weights for the current implementation of RM2006.

Table 1: RM2006: values of the time horizons τ_k, the corresponding smoothing constants, $\lambda_k = 1 - \exp(-\tau_k^{-1})$ and the weight w_k, proportional to $1 - \frac{\ln \tau_k}{\ln \tau_0}$.

k	τ_k	λ_k	w_k
1	4.0000	0.2212	0.1124
2	5.6569	0.1620	0.1058
3	8.0000	0.1175	0.0993
4	11.3137	0.0846	0.0928
5	16.0000	0.0606	0.0862
6	22.6274	0.0432	0.0797
7	32.0000	0.0308	0.0732
8	45.2548	0.0219	0.0667
9	64.0000	0.0155	0.0601
10	90.5097	0.0110	0.0536
11	128.0000	0.0078	0.0471
12	181.0193	0.0055	0.0406
13	256.0000	0.0039	0.0340
14	362.0387	0.0028	0.0275
15	512.0000	0.0020	0.0210

Denoting $w_j^\dagger = \sum_{k=1}^K w_k \lambda_k (1 - \lambda_k)$, so that $\tilde{y}_{t+1|t} = \sum_j w_j^\dagger y_{t-j}$, 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. For $l \geq 1$, multistep volatility prediction is carried out by multiplying $\tilde{y}_{t+1|t}$ by the square root of the forecast horizon.

Recalling that the RM1994 volatility prediction is based on a single EWMA with parameter $\lambda = 0.06$, the forecast horizon corresponding to this smoothing constant is $\tau = -1/\ln(0.94) = 16.16$. It is not clear how to interpret this. In section 6 we consider the notion of a reference period for the EWMA filter based on its properties as a band-pass filter.

4 Exponential smoothing for long memory processes

In this section, after discussing several possible interesting extensions of ES that can be envisaged in the long memory case, some of which are based on the structural decomposition into an underlying level and a noise component, whereas others are based on the reduced form integrated moving average representation, we turn our attention to the FerIMA model.

4.1 Fractional Local Level Model

As it is well known (see, e.g. Harvey, 1989), ES provides the minimum mean square estimator (MMSE) of the level component of the unobserved components model $y_t = \mu_t + \epsilon_t$, where the level component, μ_t, evolves as a random walk, $\mu_t = \mu_{t-1} + \eta_t$, $\eta_t \sim$ IID $\mathrm{N}(0, \sigma_\eta^2)$, and $\epsilon_t \sim$ IID $\mathrm{N}(0, \sigma_\epsilon^2)$, with $\mathrm{E}(\eta_s \epsilon_t) = 0$, forall t, s.

Hence, we may think of replacing μ_t by a fractional noise process, giving the following fractional local level model (fLLM)

$$y_t = \mu_t + \epsilon_t, \quad (1-B)^d \mu_t = \eta_t, \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right], t = 1, \ldots, n.$$

Estimation of d and of the noise-signal variance ratio $\sigma_\epsilon^2/\sigma_\eta^2$ by frequency domain and wavelet methods has been considered in Tanaka (2004). Deo and Hurvich (2001) have investigated the estimation of d via log-periodogram regression, whereas Arteche (2004) has focused on local Whittle estimation.

The main difficulty with the fLLM is signal extraction. Applying the Wiener-Kolmogorov filter (see Whittle, 1983, chapter 5 and section 8.5), the MMSE of the underlying fractional noise, assuming the availability of a doubly infinite sample, is

$$\tilde{\mu}_{t|\infty} = \frac{1}{1 + \frac{\sigma_\epsilon^2}{\sigma_\eta^2}(1-B)^d(1-B^{-1})^d} y_t.$$

The above filter encompasses two-sided exponential smoothing ($d = 1$) and the Hodrick and Prescott (1991) filter ($d = 2$). The MMSE of the level based on a semi-infinite sample (also said the concurrent or real time estimator) is (Whittle, 1983, p. 58).

$$m_t = \frac{\varphi(1)}{\varphi(B)} \left[\frac{\varphi(1)}{\varphi(B^{-1})}\right]_+ y_t,$$

where $\varphi(B)\varphi(B^{-1})\sigma^2 = \sigma_\eta^2 + \sigma_\epsilon^2(1-B)^d(1-B^{-1})^d$ and the operator $[h(B)]_+$ defines a lag polynomial containing only nonnegative powers of B, i.e. if $h(B) = \sum_{-a}^{b} h_j B^j$, $a, b > 0$,, then $[h(B)]_+ = \sum_{j=0}^{b} h_j B^j$. For fractional d, the above expressions are complicated and do not provide the signal extraction weights in closed form. The easiest way to address signal extraction and forecasting with the fLLM is to approximate the process μ_t by a finite order Markovian process as in Chan and Palma (1998).

4.2 ARFIMA(0, d, 1) process

The ARFIMA(0,d,1) process $(1-B)^d y_t = (1-\theta B)\xi_t, \xi_t \sim$ IID $\mathrm{N}(0, \sigma^2)$, where we assume that $0 \leq \theta < 1$, admits the orthogonal decomposition into two orthogonal fractional noise processes integrated respectively of order d and $d-1$:

$$y_t = \frac{\eta_t}{(1-B)^d} + \frac{\epsilon_t}{(1-B)^{d-1}}, \quad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} (1-\theta)^2 & 0 \\ 0 & \theta \end{pmatrix}\right], t = 1, \ldots, n.$$

This results from writing $(1-\theta B)\xi_t = \eta_t + (1-B)\epsilon_t$.

If μ_t denotes the first component, then the MMSE based on a doubly infinite sample is

$$\tilde{\mu}_{t|\infty} = \frac{(1-\theta)}{(1-\theta B)} \frac{(1-\theta)}{(1-\theta B^{-1})} y_t.$$

This is a two-sided EWMA which depends only on the MA parameter. The second FN($d-1$) process is extracted by the filter $\frac{\theta(1-B)(1-B^{-1})}{(1-\theta B)(1-\theta B^{-1})}$.

The real time signal extraction filter is again ES: as a matter of fact, writing $(1-\theta B) = (1-\theta) + \theta\Delta$, $y_t = m_t + e_t$, where $(1-B)^d m_t = (1-\theta)\xi_t$ can be written in terms of the observations, $m_t = [(1-\theta)/(1-\theta B)]y_t$, i.e. an EWMA of the current and past observations. Moreover, $e_t = \theta\xi_t/(1-B)^d$, and in terms of the observed time series, $e_t = \frac{\theta}{1-\theta B}(1-B)y_t$, is proportional to the same EWMA filter applied to the first differences.

Hence, the peculiar trait of this model is that signal extraction takes place by ES, regardless of the d parameter. In the case $d = 1$ we obviously obtain ES, whereas for $d = 2$ we decompose an IMA(2,1) process into an integrated RW and a RW. The forecasts are however dependent on the long memory parameter and would not take the form of exponential smoothing. In particular, the one-step-ahead predictor is given by the recursive formula

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

Notice also that e_t is predictable from its past (it is an FN($d-1$) process). Also, if $d \in (0, 1)$ the long run prediction (which reverts to zero) is not coincident with m_t.

4.3 Fractional Lag IMA(1,1)

Consider the following fractional lag IMA(1,1) model:

$$(1-B)^d y_t = (1-\theta B_d)\xi_t, \quad 0 \leq \theta < 1,$$

where the fractional lag operator, B_d, is defined as $B_d = 1-(1-B)^d$. This is a generalization of the lag operator to the long memory case originally proposed by Granger (1986) and it has been recently adopted by Johansen (2008) to define a new class of vector autoregressive models.

The above specification provides an interesting extension of ES to the fractional case. In fact, 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,

$$m_t = \frac{1-\theta}{1-\theta B_d} y_t$$

which is the ES in the case $d=1$. In the case $d=2$, the model is the basis for the decomposition of y_t into an integrated random walk plus a noise component (the reduced form being $(1-B)^2 = (1-2\theta B + \theta^2 B^2)\xi_t$).

4.4 Fractional Equal Root Moving Average Process

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

$$(1-B)^d y_t = (1-\theta B)^d \xi_t, \quad \xi_t \sim \text{WN}(0,\sigma^2). \tag{6}$$

Writing $1-\theta B = (1-\theta)-\theta(1-B)$ and using the binomial expansion of $[(1-\theta)-\theta(1-B)]^d$, the process (6) admits the following decomposition:

$$y_t \;=\; \textstyle\sum_{j=0}^{\infty} z_{jt}, \quad z_{jt} = \frac{(d)_j}{j!}\left(\frac{\theta}{1-\theta}\right)^j \left(\frac{1-\theta}{1-\theta B}\right)^d (1-B)^j y_t \tag{7}$$

where $(d)_j = d(d-1)\cdots(d-j+1)$ is the Pochhammer symbol. This is an infinite sum of generalized EWMAs applied to y_t and its successive differences.

The first component is the weighted moving average of the observations available at time t:

$$z_{0t} = \left(\frac{1-\theta}{1-\theta B}\right)^d y_t = \sum_{j=0}^{\infty} w_j^* y_{t-j}, \tag{8}$$

with weights given by

$$w_j^* = (1-\theta)^d \varphi_j, \varphi_j = -\theta \frac{1}{j}(1-d-j)\varphi_{j-1}, j > 0, \ \ \varphi_0 = 1,$$

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

$$w_j^* = (1-\theta)^d \frac{\theta^j}{j!} d^{(j)}, \tag{9}$$

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

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

$$w_j^* = w_j c_j, \ \ w_j = \lambda(1-\lambda)^j, \ \ c_j = \frac{\lambda^{d-1} d^{(j)}}{j!}, \tag{10}$$

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:

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

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

We label the linear filter $[(1-\theta)/(1-\theta B)]^d$ in (8) a fractional exponential smoothing (FES) filter. Section 6 will provide a discussion of its properties. The filter performs a long memory - EWMA of the available observations. It depends on two parameters, the memory parameter, d, and the parameter θ, which regulates the speed of mean reversion.

According to (12) the process y_t is decomposed as the sum of fractionally integrated processes of order $d-j, j = 0, 1, \ldots$.

$$z_{jt} = \tfrac{(d)_j}{j!}\theta^j(1-\theta)^{j-d}(1-B)^j \textstyle\sum_{k=0}^{\infty} \tfrac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}\xi_{t-k} \tag{11}$$

The components, for d in certain ranges, can be interpreted as the underlying level (z_{0t}), the underlying slope (z_{1t}), acceleration (z_{2t}), etc., as it will be discussed in the next section. For a stationary long memory process, z_{0t} is the fractional noise process, $(1-B)^d z_{0t} = (1-\theta)^d \xi_t$, whereas z_{1t} is the antipersistent process $(1-B)^{d-1} z_{1t} = d\theta(1-\theta)^{d-1}\xi_t$, etc. To extract the j-th component, the filter $(1-\theta B)^{-d} = \sum_j \varphi_j B^j$, $\varphi_j = \frac{\theta^j}{j!} d^{(j)}$, is applied to the series $(1-B)^j y_t$, and the outcome is rescaled by $(d)_j \theta^j (1-\theta)^{d-j}$. The filters can be approximated by truncating the weights at lag m, and rescaling them so that their sum is one (see Percival and Walden, 1993). Alternatively, a suitable number of initial values, $y_j, j \leq 0$ can be backcasted using the FerIMA model.

If d is integer, then the number of components is finite. When $d = 1$,

$$y_t = z_{0t} + z_{1t}, \quad (1 - B)z_{0t} = (1 - \theta)\xi_t, \;\; z_{1t} = \theta\xi_t,$$

and the two components can be interpreted as the permanent and the transitory components in the series (see the next section).

When $d = 2$,

$$y_t = z_{0t} + z_{1t} + z_{2t}, \quad (1 - B)^2(z_{0t} + z_{1t}) = (1 - \theta B)(1 - \theta)\xi_t, \quad z_{3t} = \theta^2\xi_t$$

so that z_{0t} is a stochastic level (an integrated RW), which is coincident with double exponential smoothing (Brown, 1963). The latter is the optimal predictor for the IMA(2,2) process with a MA root θ^{-1} with multiplicity 2; z_{1t} is a stochastic slope (a RW), and their sum is IMA(2,1), and the process z_{2t} is white noise.

The next section discusses the interpretation of the components and their relation with the multistep predictor.

5 Permanent-Transitory (Beveridge and Nelson) Decomposition

The decomposition of the FerIMA process proposed above arises as a special case of the following result, which can be viewed as a generalization of the Beveridge and Nelson (1981, BN henceforth).

Assume $d > 0$ and let y_t be the (possibly fractionally) integrated process

$$\Delta^d y_t = \psi(B)\xi_t, \quad \psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots, \quad \sum_j \psi_j^2 < \infty,$$

and $\xi_t \sim$ WN$(0, \sigma^2)$. Consider the expansion of the Wold polynomial

$$\psi(B) = \sum_{j=0}^{r-1} \psi_j(1)(1 - B)^j + \psi_r(B)(1 - B)^r,$$

where

$$\psi_{j-1}(B) = \psi_{j-1}(1) + \Delta\psi_j(B), j = 1, 2, \ldots, \quad \psi_0(B) = \psi(B).$$

Two interesting particular cases arise:

- 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 a given integer $r > 0$, the following decomposition is valid:

$$y_t = z_{0t} + z_{1t} + \cdots z_{r-1,t} + \tilde{z}_{rt}, \tag{12}$$

where

$$z_{jt} = \frac{\psi_j(1)}{(1-B)^{d-j}}\xi_t, \quad j = 0, \ldots, r-1,$$

is an FN($d-j$) process, and

$$\tilde{z}_{rt} = \frac{\psi_r(B)}{(1-B)^{d-r}}\xi_t.$$

For integer $d = r > 0$, $\tilde{z}_{r,t}$ 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 fastly to zero as the forecast horizion increases. The sum $\sum_{j=0}^{d-1} z_{jt}$ 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.

More generally, for both fractional and integer values of d, we set $r = [d]$, where $[d]$ is the nearest integer to d, in the expression (12) to obtain a generalized BN decomposition:

$$y_t = m_t + e_t, \quad m_t = z_{0t} + \cdots + z_{[d]-1,t}, \quad e_t = \frac{\psi_r(B)}{(1-B)^{d-[d]}}\xi_t.$$

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, its integration order being $d - [d] \in (-0.5, 0.5)$.

It is perhaps useful to highlight some particular cases:

- When $d \in [0, 0.5)$ the component m_t is identically equal to zero and all the series is transitory. In fact, $[d] = 0$ and the multistep forecast takes the form

$$\tilde{y}_{t+l|t} = \left[\frac{\psi(B)}{(1-B)^d} B^{-l}\right]_+ \xi_t. \tag{13}$$

- When $d \in (0.5, 1.5)$, y_t admits the following nonstationary-stationary decomposition:

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

  If $d = 1$ and $\psi(B) = (1 - \theta B)$, then m_t is an EWMA of y_t and e_t is white noise. 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.

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

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

  If $d = 2$ and $\psi(B) = (1 - \theta_1 B - \theta_2 B^2)$, then m_t can be computed according to the Holt-Winters recursive formulae (see Harvey, 1989) and e_t is white noise.

- In general, the component e_t is a stationary process featuring long memory ($d > [d]$) or antipersistence ($d < [d]$).

It should be noticed that the above decomposition differs from the one proposed by Ariño and Marmol (2004) for nonstationary fractional processes with $d \in (0.5, 1.5)$. The latter is based on a different interpolation argument and decomposes $y_t = m_t^* + e_t^*$, where, in terms of our notation, $e_t^* = \frac{\psi_1(B)}{\Gamma(d)} \xi_t$. 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 and in the case $d \in (0.5, 1)$ differs from the long run prediction of the series, which is equal to zero.

6 Fractional Exponential Smoothing Filters

Both the Riskmetrics volatility estimates and the generalized FES component in (8), result from the application of linear filters (with infinite impulse response) whose properties can be investigated in the frequency domain.

Letting $\mathrm{w}(B) = \sum \mathrm{w}_j B^j$ denote a generic linear filter, we denote its transfer function by $G(\omega) = \mathrm{w}(e^{-\imath\omega})$, $e^{-i\omega} = \cos\omega - \imath \sin\omega$, where $\imath$ is the imaginary unit and $\omega \in [0, \pi]$ is the angular frequency in radians. As it is well known, see for instance Percival and Walden (1993), the gain function, $|G(\omega)|$, provides a useful characterisation about how the linear filter modifies the amplitude of the cyclical components in the series. A low-pass filter is a filter that passes low frequency fluctuations and reduces the amplitude of fluctuations with frequencies higher than a cutoff frequency ω_c (see e.g. Percival and Walden, 1993). The latter is defined as the angular frequency at which a monotonically decreasing gain is equal to 1/2. Correspondingly, the filter is said to pass the fluctuations with period greater than $2\pi/\omega_c$ and suppress to a large extent (e.g. compress by a factor small than 0.5) those with smaller period ($\omega > \omega_c$). The cutoff frequency or period is a useful summary measure for defining the characteristic properties of a low-pass filter, although it is not the unique, as we shall see immediately.

Figure 3 compares the gains of the two Riskmetrics filters, RM1994 and RM2006. The cutoff frequency is very close, being equal to $\omega_c = 0.0889$ for RM1994 and $\omega_c = 0.0765$, which correspond to a period of 70.69 and 82.13 observations, respectively. However, the RM1994 filter is more concentrated at the cutoff. The concentration can be measured by

$$\beta^2(\omega_c) = \frac{\int_0^{\omega_c} |G(\omega)|^2 d\omega}{\int_0^{\pi} |G(\omega)|^2 d\omega}$$

The above concentration measure was defined and analysed according to different perspectives by Tufts and Francis (1970), Papoulis and Bertran (1970), Eberhard (1973) and Slepian (1978). As a result, the RM2006 volatility estimate will be slightly smoother than RM1994. However, the output of the filters will be very similar.

The FES filter $\mathrm{w}(B) = (1-\theta)^d(1-\theta B)^{-d}$ has gain

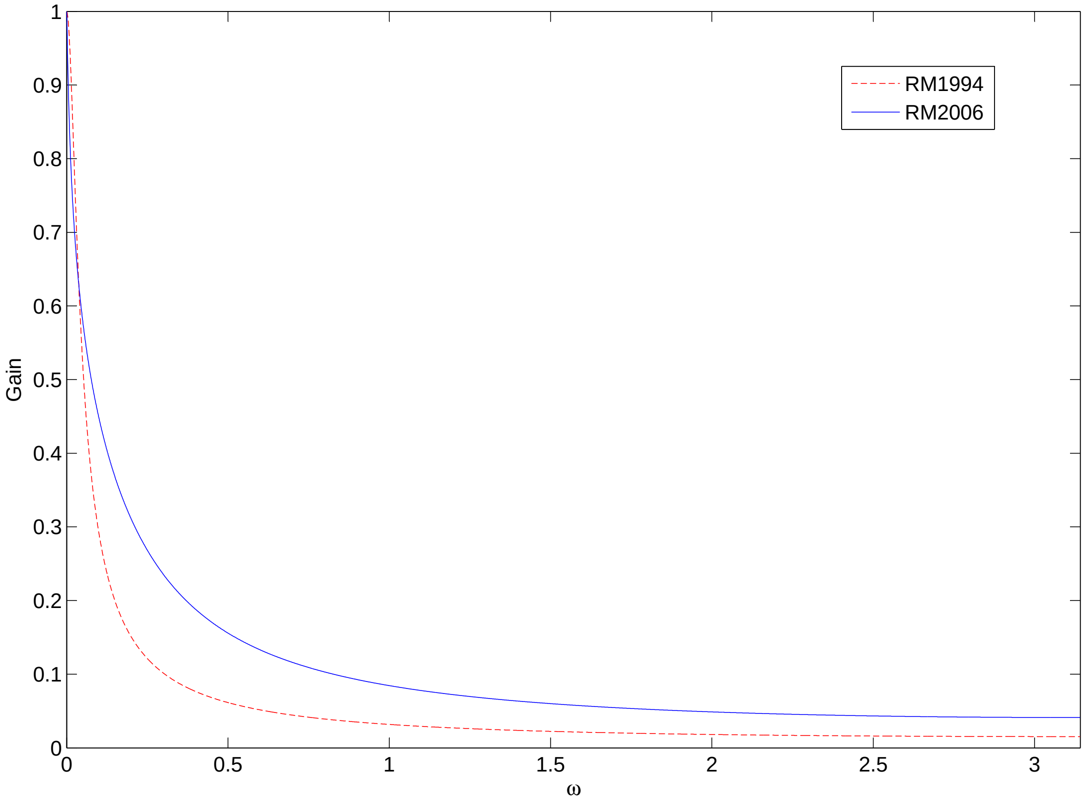
$$|G(\omega)| = \left[\frac{(1-\theta)^2}{1+\theta^2 - 2\theta\cos\omega}\right]^{d/2} = \left[\frac{1}{1+2\lambda^{-2}(1-\lambda)\cos\omega}\right]^{d/2}$$

and cutoff frequency

$$\omega_c = \arccos\left(1 - \frac{2^{2/d}-1}{2g}\right), g = \frac{1-\lambda}{\lambda^2}.$$

Figure 2 displays the combinations of d and θ, respectively in the interval (0,2] and [0,1], giving the same cutoff frequency ω_c, for some values of ω_c. The curve for $\omega_c = 0.0765$ provides the combinations that will deliver filtered estimates with comparable smoothness with respect to RM2006; for instance, we need $d = 0.7$, $\theta = 0.98$ to obtain a similar filter. Other cutoff frequencies taken into consideration are $\omega_c = 0.5$, corresponding to a period of 13 observations, $\omega_c = 1$, corresponding to a period of 6 observations, and $\omega_c = \pi$, corresponding to a period of 2 observations. For

Figure 1: Plot of the gains of the RM1994 and RM2006 filters versus the angular frequency ω.



stationary values of d and θ less than 0.5 the FES filter is an all-pass filter (i.e. $|G(\pi)| < 0.5$), and in order to obtain a substantial amount of smoothing we need to have θ very close to 1. Small variations of the θ parameter in the neighbourhood of 1 cause big changes of the cutoff frequency. Obviously, $d = 1$, $\theta = 0.94$ yields the RM1994 ES filter.

7 Estimation and Forecasting with the FerIMA model

Estimation of the parameters d and θ characterising the FerIMA process can be carried out by frequency domain maximum likelihood. Given a time series realisation $\{y_t, t = 1, 2, \ldots, n\}$, and letting $\omega_j = \frac{2\pi j}{n}$, $j = 1, \dots, \lfloor \frac{n-1}{2} \rfloor$, denote the Fourier frequencies, where $\lfloor \cdot \rfloor$ is the largest integer not greater than the argument, the periodogram, or sample spectrum, is defined as

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

where $\bar{y} = n^{-1} \sum_{t=1}^{n} y_t$.

Letting

$$f(\omega) = \frac{1}{2\pi} \left(\frac{1 + \theta^2 - 2\theta \cos\omega}{2(1 - \cos\omega)} \right)^p \sigma^2$$

denote the (pseudo) spectral density of y_t, the Whittle likelihood is:

$$\ell(d, \theta, \sigma^2) = - \sum_{j=1}^{\lfloor (n-1)/2 \rfloor} \left[\ln f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)} \right]. \tag{14}$$

The maximiser of (14) is the Whittle pseudo maximum likelihood estimator of (d, θ, σ^2). We refer to Dahlhaus (1989), Giraitis, Koul and Surgailis (2012), and Beran et al. (2013) and for the properties of the estimator in the long memory case. In the nonstationary case, the consistency and the asymptotic normality of the Whittle estimator has been proven by Velasco and Robinson (2000). Tapering may be needed to eliminate polynomial trends, see Velasco and Robinson (2000), but we do not contemplate this possibility here. Notice that we have excluded the frequencies $\omega = 0$, and π from the analysis; the latter may be included with little effort, and their effect on the inferences is negligible in large samples.

The l-step-ahead forecast of the FerIMA model is obtained from each components forecasts, i.e. $\tilde{y}_{t+l|t} = \sum_j \tilde{z}_{j,t+l|t}$. The components are characterised by decreasing levels of predictability: in fact, the prediction error variance increases with the order of the component, j.

Alternatively, the predictor can be obtained directly from the model specification. Denoting by $\bar{y}_t = t^{-1} \sum_{j=0}^{t-1} y_{t-j}$, the mean of the available sample observations at time t, the l-steps-ahead predictor of y_t is

$$\begin{aligned} \tilde{y}_{t+l|t} &= \bar{y}_t + \textstyle\sum_{j=0}^{t-1} \pi_{jl} (y_{t-j} - \bar{y}_t), \\ &= \textstyle\sum_{j=0}^{t-1} \pi^*_{jl} y_{t-j}, \quad \pi^*_{jl} = \pi_{jl} + \frac{1}{t}(1 - \pi_{jl}). \end{aligned}$$

The weights π_{jl} are computed by the Durbin-Levinson algorithm (see e.g. Palma, 2007) and in large sample they are obtained as follows:

$$\pi_{jl} = \sum_{i=0}^{l-1} \psi_i \pi_{j+l-i}$$

where ψ_i is the coefficient of the Wold polynomial associated to B^i, $y_t = \psi(B)\xi_t$ and $\pi(B)$ is the AR polynomial in the infinite AR representation $\pi(B) y_t = \xi_t$.

8 The Empirical Performance of the FerIMA Volatility Predictor

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 background information for realized measures can be found in the survey articles by by McAleer and Medeiros (2008) and Andersen, Bollerslev and Diebold (2010). The series range from 03/01/2000 to the 22/01/2014 for a total of 3,674 daily observations.

Denoting by RV_t a generic realized volatility series, we focus on its logarithmic transformation, that is we take $y_t = \ln RV_t$. We are interested in assessing the properties of the FerIMA predictor discussed in the previous section, in comparison to three well established alternative: RM1994, which is the standard exponential smoothing predictor with $\lambda = 0.06$, the RM2006 methodology, and the heterogeneous autoregressive (HAR) model proposed by Corsi (2009), which is specified as follows:

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

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. This specification captures the long memory feature of RV via a long autoregression, yet preserving the parsimony, and has proven to be very effective for forecasting volatility, rapidly becoming one of the discipline’s standards.

We perform a recursive forecasting experiment such that starting from time $n_0 = 500$ we compute the one step ahead volatility predictions according to the four methods and we proceed adding one observation at time. For the HAR And FerIMA specifications we re-estimate the parameters each time a new observation is added. The experiment yields 3,174 one step ahead prediction errors for each forecasting methodology to be used for the comparative assessment.

Denoting by $\tilde{y}_{k,t|t-1}$ the prediction arising from method k, where k is an element of the set $\{$RM1994, RM2006, HAR, F$\}$, F standing for FerIMA, and by $v_{k,t} = y_t - \tilde{y}_{k,t|t-1}$, the corresponding prediction error, $t = n_0 + 1, \dots, n$, we compare the mean square forecast error,

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

and compute the Diebold-Mariano-West test of equal forecasting accuracy. Denoting $d_{k,t} = v_{k,t}^2 - v_{F,t}^2$ the quadratic loss differential, the Diebold-Mariano-West test of the null hypothesis of equal forecast accuracy, $H_0 : \mathrm{E}(d_{k,t}) = 0$, versus the one sided alternative $H_1 : \mathrm{E}(d_{kt}) > 0$, is the test statistic

$$DM_k = \frac{\bar{d}_k}{\sqrt{\sigma_k^2}}, \quad \bar{d}_k = \frac{1}{n - n_0} \sum_{t=n_0+1}^{n} d_{k,t}, \quad \sigma_k^2 = \frac{1}{n - n_0} \left[c_0 + 2 \sum_{j=1}^{J-1} \frac{J-j}{J} c_j \right],$$

where c_j is the sample autocovariance of $d_{k,t}$ at lag j and σ_k^2 is a consistent estimate of the long run variance of the loss differential. We set the truncation lag equal to $J = 22$. See Diebold and Mariano (1995) and West (1996). The null distribution of the test is Student's t with $n - n_0 - 1$ degrees of freedom.

The results for the 21 realized volatility series ($y_t = \ln RV_t$) are reported in table 2. The FerIMA predictor is characterised by a lower MSFE and systematically outperforms the RM1994 and RM2006 predictors, with a single exception (*All Ord.* series). Only in 3 out of 21 cases the HAR predictor has a lower MSFE. In terms of the Diebold-Mariano-West test, we reject the null that the RM1994 and RM2006 predictors have the same forecast accuracy as the FerIMA predictor at the 5% significance level in all but two cases (*All Ord.* and *Hang Seng*). When the HAR predictor is compared to the FerIMA predictor, we do not reject in four cases (*All Ord.*, *DIJA*, *Hang Seng*, and IPC Mexico).

Hence, the evidence is strongly in favour of the FerIMA predictor. We also observe that the performance of RM2006 does not differ substantially from RM1994, since, as it was anticipated in section 6, the two predictors are very similar. Also, HAR systematically outperforms both RiskMetrics predictors.

The last two columns of the table report the values of the estimated FerIMA parameters θ and d. The memory parameter is always in the nonstationary region (the average of the estimated values being 0.60, with standard deviation 0.04), whereas the moving average parameter ranges from 0.21 to 0.65 with an average value 0.36 and standard deviation 0.12.

Figure 3 displays the logarithm of the realized volatility series for the S&P 500 index, along with the RM2006 filtered series and the component $m_t = z_{0t}$ extracted from the FerIMA model, computed according to 8, replacing the unknown parameters by their maximum likelihood estimates $\tilde{d} = 0.56$ and $\tilde{\theta} = 0.35$. The bottom left plot is the stationary component $y_t - m_t = e_t$ and the bottom right plot is the deviation from the RM2006 filtered series. It is noticeable that a substantial part of the variation of y_t is absorbed by the component z_{0t}, whereas RM2006 seems to oversmooth the series. The overall message is that volatility is a strongly persistent process with a stationary component contributing little. Notice that if the θ estimate were close to zero, then the series would be a FN(d) process and all the variability would be absorbed by z_{0t}.

Table 2: Log-Realized volatility series. Recursive forecasting exercise: comparison of one-step-ahead predictive performance. The first three columns present the relative MSFE ratios $\frac{\text{MSFE}_F}{\text{MSFE}_{RM2006}}$, $\frac{\text{MSFE}_F}{\text{MSE}_{RM1994}}$, $\frac{\text{MSFE}_F}{\text{MSFE}_{HAR}}$, respectively. The next three columns report the p-values of the Diebold-Mariano test of equal forecast accuracy versus the alternative that the FerIMA (abbreviated to F) is more accurate.

	Relative Mean Square Forecast Error			p-values of DM test				
Series	F vs RM2006	F vs RM1994	F vs HAR	F vs RM2006	F vs RM1994	F vs HAR	$\tilde{d}$	$\tilde{\theta}$
S&P 500	0.84	0.85	0.96	0.000	0.000	0.070	0.63	0.44
FTSE 100	0.74	0.76	0.88	0.000	0.000	0.000	0.65	0.35
Nikkei 225	0.76	0.75	0.84	0.000	0.000	0.000	0.55	0.23
DAX	0.75	0.77	0.87	0.000	0.000	0.000	0.61	0.29
Russel	0.86	0.86	0.95	0.000	0.000	0.021	0.58	0.41
All Ord.	1.03	1.04	1.09	0.833	0.887	1.000	0.64	0.65
DJIA	0.91	0.91	1.01	0.016	0.017	0.694	0.65	0.50
Nasdaq 100	0.75	0.75	0.83	0.000	0.000	0.000	0.59	0.21
CAC 40	0.75	0.76	0.88	0.000	0.000	0.000	0.62	0.31
Hang Seng	0.97	0.98	1.04	0.255	0.341	0.935	0.64	0.55
KOSPI Composite Index	0.76	0.76	0.86	0.000	0.000	0.000	0.61	0.27
AEX Index	0.73	0.74	0.87	0.000	0.000	0.000	0.63	0.31
Swiss Market Index	0.73	0.76	0.91	0.000	0.000	0.001	0.66	0.36
IBEX 35	0.75	0.75	0.85	0.000	0.000	0.000	0.58	0.26
S&P CNX Nifty	0.81	0.79	0.85	0.000	0.000	0.000	0.53	0.26
IPC Mexico	0.94	0.94	0.98	0.030	0.020	0.141	0.55	0.50
Bovespa Index	0.80	0.79	0.87	0.000	0.000	0.000	0.51	0.23
S&P/TSX Composite Index	0.87	0.88	0.92	0.003	0.002	0.002	0.57	0.40
Euro STOXX 50	0.82	0.82	0.91	0.000	0.000	0.001	0.60	0.37
FT Straits Times Index	0.82	0.83	0.88	0.000	0.000	0.000	0.58	0.36
FTSE MIB	0.78	0.79	0.88	0.000	0.000	0.000	0.59	0.29

9 Conclusions

We have dealt with the problem of forecasting volatility and decomposing it into meaningful components in the presence of long memory and possible nonstationarity.

We have reviewed the solutions available in the literature and have concluded that the recent RM2006 yields results that are not substantially different from the exponential smoothing predictor with fixed smoothing constant known as RM1994. From the point of view of signal extraction, when applied to realized volatility series (rather than squared or absolute returns or similar noisy proxies), both methodologies yield estimates of underlying volatility that are very stable and imply a very smooth estimate.

After reviewing some plausible extensions of ES to the fractionally integrated framework, we have looked at the properties of a signal extraction filter and the predictor arising from the FerIMA model, a specification originally formulated by Hosking (1981), proposing a decomposition into fractional noise components of decreasing order and offering an interpretation in the light of the well known Beveridge and Nelson decomposition.

A forecasting experiment has illustrated the potential of the FerIMA model for forecasting daily realized volatility, showing that it outperforms both RiskMetrics predictors and the Heterogeneous Autoregressive Model.

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

Acknowledgements

The author thanks Federico Carlini at CREATES for discussion on the fractional lag operator. The author gratefully acknowledge 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.

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] Amado, C., and Teräsvirta, T. (2013). Modelling volatility by variance decomposition. *Journal of Econometrics*, 175, 142-153.

[3] Amado, C., and Teräsvirta, T. (2014). Modelling changes in the unconditional variance of long stock return series. *Journal of Empirical Finance*, 25(C), 15-35.

[4] Andersen, T.G., Bollerslev, T., and Diebold, F.X. (2010). Parametric and Nonparametric Volatility Measurement. In Y. Aït-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, Chapter 2, pp. 67-128. Amsterdam: Elsevier Science B.V.

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

[6] Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (2001). The Distribution of Exchange Rate Volatility. *Journal of the American Statistical Association*, 96, 4255.

[7] Arteche, J. (2004). Gaussian Semiparametric Estimation in Long Memory in Stochastic Volatility Models and Signal plus Noise Models. *Journal of Econometrics*, 119, 131-154.

[8] Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013), *Long-Memory Processes Probabilistic Properties and Statistical Methods*, Springer-Verlag Berlin Heidelberg.

[9] Bauwens, L., Hafner, C.M., and Pierret, D. (2013). Multivariate Volatility Modeling of Electricity Futures. *Journal of Applied Econometrics*, 28, 743 761.

[10] 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.

[11] Bollerslev, T. and Wright, J.H. (2000). Semiparametric Estimation of Long-Memory Volatility Dependencies: The Role of High Frequency Data. *Journal of Econometrics*, 98, 81106.

[12] Brown, R.G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series*. NJ: Prentice-Hall, Englewood Cliffs.

[13] Chan, N. H. and Palma, W. (1998). State Space Modeling of Long-Memory Processes. *The Annals of Statistics*, 26, 719–740.

[14] Colacito, R., Engle R.F., Ghysels, E. (2011). A component model for dynamic correlations. *Journal of Econometrics*, 164, 45–59.

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

[16] Cox, D. R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society, Series B*, 23, 414-422.

[17] Dahlhaus, R. (1989). Efficient Parameter Estimation for Self Similar Processes. *The Annals of Statistics*, 17, 4, 1749-1766.

[18] Deo, R.S. and Hurvich, C.M. (2001). On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory*, 17, 686-710.

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

[20] Diebold, F. X. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13, 253 - 263.

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, 83106.

[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 funda- mentals. *Review of Economic and Statistics*, 95, 776797.

[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. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*, 4, 1-28.

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

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

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

[30] Giraitis, L., Koul, H. and Surgailis, D. (2012). *Large Sample Inference for. Long memory Processes*, Imperial College Press, London.

[31] Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227238.

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

[33] 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.

[34] 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.

[35] Gray, H.L., Woodward, W. A. and Zhang, N. (1989). On Generalized Fractional Processes. *Journal of Time Series Analysis*, 10, 233–257.

[36] Hafner, C., and Linton, O.B. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics*, 159, 5573.

[37] Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

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

[39] 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.

[40] Hodrick, J. R., Prescott E. C. (1997). Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit and Banking*, 29, 116.

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

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

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

[44] McAleer, M. and Medeiros, M. C. (2008). Realized Volatility: A Review. *Econometric Reviews*, 27, 10–45.

[45] Palma, W. (2007). *Long-memory Time Series: Theory and Methods*. Wiley. Hoboken, New Jersey.

[46] Percival D., Walden A. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.

[47] Pillai, T.R., Shitan. M. and Peiris, M.S. (2012) Some Properties of the Generalized Autoregressive Moving Average (GARMA (1,1;δ_1, δ_2)) Model. *Communications in Statistics - Theory and Methods*, 41, 699–716.

[48] Proietti, T., and Luati A. (2012). Generalised Linear Spectral Models. Forthcoming in Shephard, N. and Koopman, S.J. (2014), *Unobserved Components and Time Series Econometrics*, Oxford University Press, Oxford, UK.

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

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

[51] Slepian D., Pollak H.O. (1961), Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty, I, *The Bell System Technical Journal*, 40, 43-64.

[52] Taylor, S.J. (2005). *Asset Price Dynamics, Volatility and Prediction*. Princeton University Press.

[53] Tanaka, K. (2004). Frequency domain and wavelet-based estimation for long-memory signal plus noise models. In A. Harvey. S.J. Koopman and N. Shephard (eds.), *State Space and Unobserved Component Model. Theory and Application*, ch. 4, p. 75–91, Cambridge University Press. 13, 109–131.

[54] Tiao, G. C., and Tsay, R. S. (1994). Some advances in non-linear and adaptive modeling in time-series. Journal of Forecasting, 13, 109–131.

[55] Tiao, G.C., and Xu, D. (1993). Robustness of maximum likelihood estimates for multi-step predictions: the exponential smoothing case. *Biometrika*, 80, 623 641.

[56] 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.

[57] West, K. (1996). Asymptotic inference about predictive ability, *Econometrica*, 64, 1067 - 1084.

[58] Whittle, P. (1953). Estimation and Information in Stationary Time Series, *Arkiv för Matematik*, 2, 423–434.

[59] Whittle, P. (1983). *Prediction and Regulation by Linear Least-square Methods* (2nd edition). Basil Blackwell, Oxford, UK.

[60] 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 2: Combinations of d (horizontal axis) and θ (vertical axis) giving the same cutoff frequency for FES filter.

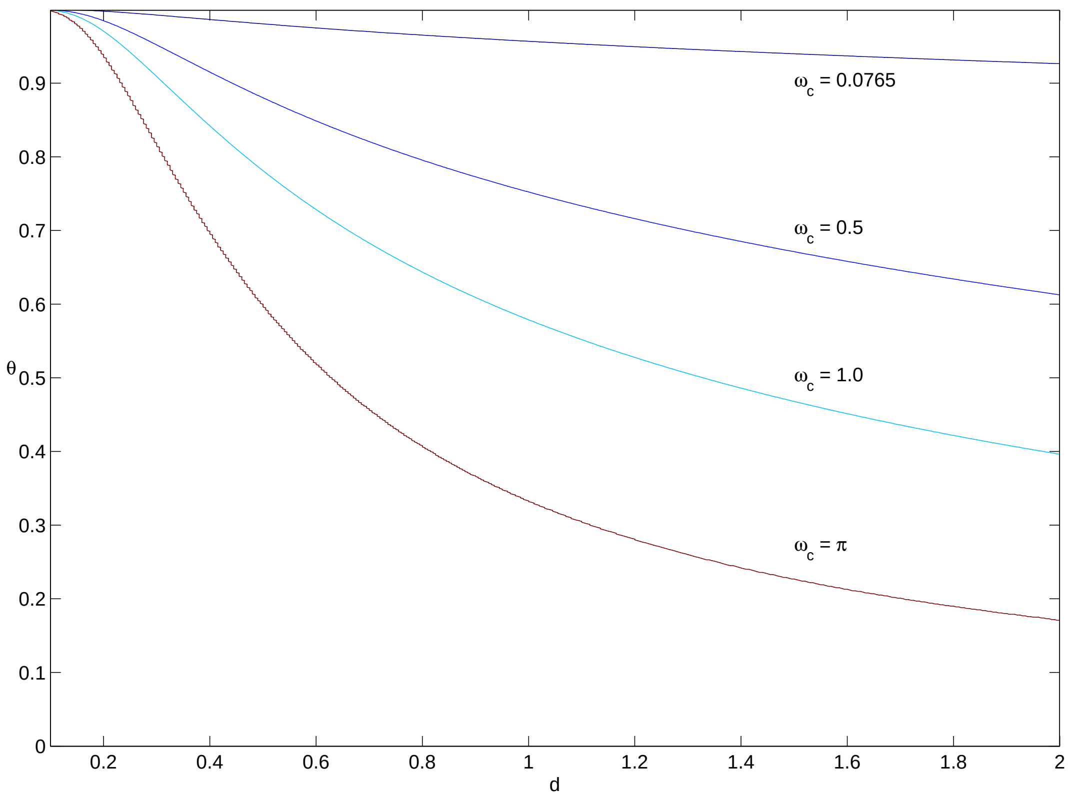


Figure 3: Logarithms of 5-minutes daily realized volatility for the SP500 index (red points) and FES and RM2006 components.