AGGREGATION OF EXPONENTIAL SMOOTHING PROCESSES WITH
AN APPLICATION TO PORTFOLIO RISK EVALUATION

Giacomo SBRANA¹ and Andrea SILVESTRINI²

March 2010

Abstract

In this paper we present a unified framework to analyse contemporaneous and temporal
aggregation of exponential smoothing models (EWMA), making use of a representation based
on macro processes. Focusing on a widely employed class of integrated moving average (IMA)
models, we derive algebraic solutions to recover the parameters of the model for the contempo-
ranously and temporally aggregated time series. In addition, an empirical application dealing
with Value-at-Risk (VaR) prediction at different sampling frequencies for an equally weighted
portfolio composed of multiple indices is presented. In the framework of EWMA estimates
of volatility, we suggest an alternative approach based on the moving average parameters of
the aggregated IMA model. Empirical results show that VaR predictions delivered by this
suggested approach are at least as accurate as those obtained applying the standard univariate
RiskMetrics™ methodology.

Keywords: Contemporaneous and temporal aggregation; EWMA; Volatility forecasting; Value-at-Risk
JEL Code: C10, C32, C43

¹Université de Strasbourg, Bureau d’Économie Théorique et Appliquée (BETA), 67085 Strasbourg Cedex, France.
E-mail: <sbrana@unistra.fr>

²Bank of Italy, Economics and International Relations Area, 00184 Roma, Italy. E-mail:
<andrea.silvestrini@bancaditalia.it>

The authors would like to thank Luc Bauwens, Daniele Coin, Riccardo De Bonis, Helmut Lütkepohl, Claudia Miani,
Matteo Piazza and David Veredas for useful comments and suggestions. The views expressed in this paper are those
of the authors and do not necessarily reflect the positions or policies of the Institutions in which they work.
1 Introduction

It has been widely recognized in the econometric literature that most macroeconomic time series are aggregates over the temporal or the cross-section dimension (Granger and Lee, 1999). Whenever we sum disaggregate data along the time dimension, we are performing a temporal aggregation of the original time series. For instance, annual GDP is the sum of quarterly GDP over four subsequent periods regularly spaced. A different kind of aggregation, through individuals rather than through time, is termed contemporaneous aggregation. For example, HICP inflation (overall index) is a contemporaneous aggregation along the inflation sub-indices (e.g. processed food, unprocessed food, energy, non-energy, etc.).

On the one hand, the interest in temporal aggregation is motivated by the fact that the empirical verification of economic theories is often based on the analysis of data sampled at different frequencies. Indeed, time series encountered in practice may be observed at mixed frequencies. In such cases, in order to make the database homogeneous, higher frequency data are temporally aggregated to lower frequencies. The econometric analysis is then conducted on transformed data, and the estimation results may depend upon the choice of the time interval and on the aggregation scheme. The literature on temporal aggregation is vast and comprises, among others, Amemiya and Wu (1972), Tiao (1972), Brewer (1973), Wei (1978), Weiss (1984), Stram and Wei (1986), Lütkepohl (1987, 1989), Wei and Stram (1988), Nijman and Palm (1990), Wei (1990), Drost and Nijman (1993), Marcellino (1999), Breitung and Swanson (2002), Hafner (2008, 2009) and Amendola, Niglio and Vitale (2009). Gómez and Aparicio-Pérez (2009), very recently, propose a general approach to deal with temporal aggregation, interpolation and prediction of disaggregate data. We refer to Silvestrini and Veredas (2008) for an updated survey of temporal aggregation in univariate and multivariate ARIMA-GARCH time series models.

On the other hand, contemporaneous aggregation is closely connected to the relationship between the micromodels, as derived from the microeconomic theory, and the corresponding models at the macro level, after aggregation. These relevant issues have been deeply analysed in the rigorous treatment provided by Forni and Lippi (1997), where the focus is on the consequences of contemporaneous aggregation in the context of general dynamic economic models. According to Forni and Lippi (1997), the properties and the parameters of the models at the macro level bear a relationship to the corresponding parameters of the micromodel. This is relevant to reconcile the apparent conflict between the properties of the aggregated data and the micromodel: in this sense, aggregation theory constitutes a link between micro theory and empirical research at the macro level based on aggregated data.

In the time series context, contemporaneous aggregation of ARMA models has been extensively discussed by Granger and Morris (1976), Robinson (1978), Granger (1980), Lewbel (1994), Garderen, Lee and Pesaran (2000), Giacomini and Granger (2004), Zaffaroni (2004), among others. Granger (1990) provides a review of “large-scale”, “small-scale”, and temporal aggregation of time series models. Very recently, in the context of panel data, Petkovic and Veredas (2009) study the impact of cross-section and temporal aggregation in linear static and dynamic models. A related strand of theoretical literature focuses on the effect of contemporaneous aggregation on forecasting efficiency and on the choice of the optimal level of aggregation in forecasting. We refer to Rose (1977), Tiao and Guttman (1980), Wei and Abraham (1981), Kohn (1982), Lütkepohl (1984a, 1984b, 1984c, 1987, 2004), Pino, Morettin and Mentz (1987). Consider, for instance, the problem of predicting Euro area variables: in principle, this can be done by pooling country specific forecasts or, alternatively, directly forecasting the aggregate variable. These issues have been empirically
investigated by an increasing number of applied papers, especially in the latest years. See Sbrana (2008), among many others, for a recent contribution.

A joint framework to deal with contemporaneous and temporal aggregation of vector ARMA (VARMA) type of models has been suggested and formalized by Lütkepohl (1984a, 1987). In particular, this author shows that contemporaneous and temporal aggregation of a VARMA model corresponds to a linear transformation of a “macro process”, which constitutes a different representation of the original VARMA. Therefore, the theory of linear transformations of VARMA processes can be applied to derive results on the consequences of simultaneous temporal and contemporaneous aggregation, as argued by Lütkepohl (1984b). Furthermore, sticking to the framework of linear transformations, it is possible to prove the closeness of VARMA models after aggregation, namely, the fact that the aggregated model is still in the VARMA class and keeps the same structure.

However, although VARMA processes are closed, the parameter values and the orders of the aggregated models vary as a function of the frequency and level of aggregation. According to a number of existing results in the econometric literature, the parameter values and the orders of the aggregated models can be derived by establishing the relationship between the autocovariance structures of the disaggregated and aggregated models. In particular, to recover the aggregated parameter values, in most of the cases it is necessary to solve nonlinear systems of equations, by means of numerical algorithms, as detailed by Granger and Morris (1976), Hamilton (1994, pp. 102-108), Wei (1990) or Marcellino (1999).

The purpose of this work is to integrate the existing results in the aggregation literature, by focusing on the effect of temporal and contemporaneous aggregation in the context of exponential smoothing methods. As well known, in general, linear Exponentially Weighted Moving Average (EWMA) methods have equivalent representations based on ARIMA models. By “equivalent”, as pointed out by Gardner and McKenzie (1985), we mean that the forecasts produced by the smoothing techniques are minimum mean squared error (MSE) forecasts for the corresponding ARIMA models. These authors explain in detail how to establish a link between several exponentially weighted moving average methods and the corresponding ARIMA models. In particular, working with a standard formulation of an exponential smoothing system and setting the damping parameter to zero, they obtain an ARIMA(0,1,1), or IMA (1,1), with no constant term as equivalent ARIMA model. This smoothing method is also known as simple exponential smoothing or N-N, meaning no trend with no seasonality model (see again Gardner and McKenzie, 1985). Likewise, Harvey (1989) shows that the IMA(1,1) model leads to the same forecasts produced by an EWMA and that the IMA(1,1) is the reduced form of a random walk plus noise structural model.

In this broad framework, our main contribution is to derive algebraic solutions for the parameters of the contemporaneously and temporally aggregated model, assuming that the original disaggregate model is an integrated moving average of order one. The solutions represent exact functions linking the unknown aggregate parameters with the data generating process. This is, in our view, an important contribution since moving average parameters are nonlinear function of the autocorrelation function. In this way, we show that no complicated numerical algorithm is needed for the calculation of the aggregated moving average parameters. Although the focus is on a specific class of moving average processes, results are valid for any aggregation frequency and for every level of contemporaneous aggregation. To our knowledge, the results

---

1 We refer to Gardner (1985, 2006) for a comprehensive review and thorough discussion.

2 A similar discussion may be found in Wei (1990, pp. 73-76).
presented in this paper can be considered as a novelty and constitute an improvement in the research field of aggregation.

Finally, a few more words about the motivation of the paper. It is worth noting that the simple exponential smoothing method/IMA model is quite popular in macroeconomics and widely employed due to its simplicity and ability to capture nonstationarity. In addition, this type of process has a long tradition and it is one of the most recommended method in forecasting time series. The optimal properties of the simple exponential smoothing were first discussed in Muth (1960). The EWMA became the cornerstone process at the origin of the rational expectations theory as in Muth (1961). The author stressed the importance of the EWMA process in modeling inflation. Since then, several works decided to employ this model. For example, Nelson and Schwert (1977) and subsequently Barsky (1987) claim that the EWMA is the best forecasting process, outperforming other univariate processes, in modeling inflation rate. Also Pearce (1979) supports the use of EWMA in forecasting price levels. Similar results have been achieved more recently by Rossana and Seater (1995), who argue that many economic time series sampled at relatively low frequencies can often be approximated by an IMA (1,1) model and by Stock and Watson (2007).

The EWMA is not only employed in macroeconomics, but also in finance. In particular, this model is widely used to produce forecasts of volatilities of financial data by the RiskMetrics™ methodology. This well known technique was developed by J.P. Morgan to perform volatility prediction and is extremely popular nowadays. See Zaffaroni (2008) for a recent analysis of the theoretical properties and performance of this model in the multivariate context.

The rest of this paper is structured as follows. In Section 2 we set up the econometric framework, which is based on the definition of “macro processes” introduced by Lütkepohl (1984a, 1987). In Section 3 we derive algebraic solutions for the parameters of the contemporaneously aggregated process. In Section 4 we derive algebraic solutions for the parameters of the temporally aggregated process. In Section 5 we illustrate how the results on temporal and contemporaneous aggregation can be applied in practice analysing the problem of volatility prediction and Value-at-Risk (VaR) calculation at different sampling frequencies. Section 6 contains some final remarks. The proofs are relegated to an appendix.

2 The econometric framework: the IMA(1,1) model for the original data

It is assumed that the data generating process is a vector integrated moving average process of order one, that is, a vector IMA(1,1). The econometric framework is based on the commonly named “macro processes”, introduced by Lütkepohl (1987) to cover the aggregation issue in full generality. In particular, we use a representation similar to the one proposed by Lütkepohl (2007, p. 441):

---

3 For example, Makridakis, Wheelright, and McGee (1983) in their classic text on forecasting say “When forecasts are needed for thousands of items, as is the case in many inventory systems, smoothing methods are often the only acceptable methods.” Moreover, Willis (1987): “Exponential Smoothing is one of the most used and most written about forms of forecasting. Material on this can be found in almost any forecasting text, and in many books on operations management and management science.” In addition, “Their [Exponential smoothing models] central advantage is not so much forecast accuracy, but rather system efficiency (pp. 196).” See also Chapter 4 of Box, Jenkins and Reinsel (2008).

---
A few words of clarification are at order. The model in (1) is an IMA(1,1) for the disaggregate data. In (1), each \( y_{k(t-1)+i}, i = 1, 2, j = 1, 2, \ldots, k, \) is an \((N \times 1)\) vector and each \( \varepsilon_{k(t-1)+j}, i = 1, 2, j = 1, 2, \ldots, k, \) is an \((N \times 1)\) vector. Furthermore, \( \theta \) is an \((N \times N)\) matrix. Note that \( k \) represents the temporal aggregation frequency. For instance, when \( k = 3 \), the original vector process is temporally aggregated over three subsequent periods. Similarly, \( N \) is the order of contemporaneous aggregation. For example, if \( N = 4 \), we are summing four processes along the cross-section dimension (think of gross domestic product, i.e. GDP, in one period, which is the sum of consumption, gross investment, government spending and net exports in that period). If \( N = k = 4 \), we are summing four processes across the cross-section and the data are also summed over four subsequent periods (e.g., sum of quarterly consumption, gross investment, government spending and net exports, and sum of the result over four periods. What we get is annual GDP).

As it will become clear in the sequel, the macro process in (1) allows to consider contemporaneous and temporal aggregation jointly. In what follows we explain how to derive algebraic solutions to recover the parameters of the model for the contemporaneously and temporally aggregated series. For the sake of clarity, we discuss contemporaneous aggregation in Section 3 and temporal aggregation in Section 4, separately. However, we remark that our focus is on simultaneous temporal and contemporaneous aggregation: in other words, these two forms of aggregation should be thought as being acting at the same time on the original data.
3 Parameters of the contemporaneously aggregated process

Without loss of generality, let us focus on the first row of (1):

\[ y_{k(t-1)+1} = y_{k(t-1)} + \varepsilon_{k(t-1)+1} + \theta \varepsilon_{k(t-1)} \]  

In (2), the size of the \( y \) and \( \varepsilon \) vectors and of the \( \theta \) matrix has been explicitly written out.

To deal with contemporaneous aggregation of the process in (2), it is useful to consider the following \( N \)-variate system representation:

\[
\begin{pmatrix}
(1 - L)y_{1,k(t-1)+1} \\
(1 - L)y_{2,k(t-1)+1} \\
\vdots \\
(1 - L)y_{N,k(t-1)+1}
\end{pmatrix} =
\begin{pmatrix}
(1 + \theta_{11}L) & \theta_{12}L & \cdots & \theta_{1N}L \\
\theta_{21}L & (1 + \theta_{22}L) & \cdots & \theta_{2N}L \\
\vdots & \vdots & \ddots & \vdots \\
\theta_{N1}L & \theta_{N2}L & \cdots & (1 + \theta_{NN}L)
\end{pmatrix}
\begin{pmatrix}
\varepsilon_{1,k(t-1)+1} \\
\varepsilon_{2,k(t-1)+1} \\
\vdots \\
\varepsilon_{N,k(t-1)+1}
\end{pmatrix}
\]

where \( L \) is the usual lag operator, such that \( Lx_t = x_{t-1} \). In addition, by assumption, \( \varepsilon'_{k(t-1)+1} = (\varepsilon_{1,k(t-1)+1}, \varepsilon_{2,k(t-1)+1}, \ldots, \varepsilon_{N,k(t-1)+1}) \) is a vector of white noise innovations such that \( E(\varepsilon_s) = 0 \) and \( E(\varepsilon_s \varepsilon'_s) = \Sigma \). More specifically:

\[
E(\varepsilon_{k(t-1)+1}\varepsilon'_s) =
\begin{pmatrix}
\sigma^2_1 & \sigma_{12} & \cdots & \sigma_{1N} \\
\sigma_{21} & \sigma^2_2 & \cdots & \sigma_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{N1} & \sigma_{N2} & \cdots & \sigma^2_N
\end{pmatrix}
\quad if \quad s = k(t-1) + 1
\]
\[
E(\varepsilon_{k(t-1)+1}\varepsilon'_s) = O_N \quad otherwise.
\]

We present the following preparatory lemma which establishes the order of the contemporaneously aggregated process, which is an IMA(1,1) for any level of aggregation.

**Lemma 1** Focusing on system (3), after taking the first differences, it is possible to show that the process generated after contemporaneous aggregation is an MA(1), that is:

\[ y_{k(t-1)+1} = F((1 - L)y_{k(t-1)+1}) = (1 + \theta L)a_{k(t-1)+1}, \]

where \( F \) is an \((1 \times N)\) generic aggregation vector. In addition, from the well-known properties of the MA(1) process, we have:

\[ E(y^2_{k(t-1)+1}) = (1 + \theta^2)\sigma_a^2 \quad E(y_{k(t-1)+1} y_{k(t-1)}) = \theta \sigma_a^2, \]

where \( \sigma_a^2 \) is the variance of \( a_{k(t-1)+1} \) and \( \theta \) is the MA parameter of the contemporaneously aggregated process, by definition.

\[^{4}\text{We can focus on any of the rows of (1). Results in Section 3 are clearly unaffected.}\]
Proof. The result stems from the fact that, in general, summing up across \( i \) moving average processes of order \( q_i \) leads to an MA(\( q^* \)) where \( q^* \leq \max(q_i) \). A detailed proof can be found in Lütkepohl (2007, page 436).

To complement the analysis, we allow for two different contemporaneous aggregation schemes:

- Simple sum of variables, that is, \( F = (1, 1, ..., 1) \);
- Weighted sum of variables, that is, \( F = (\omega_1, \omega_2, ..., \omega_N) \), with \( \sum_{i=1}^{N} \omega_i = 1 \).

**Proposition 1** Assuming the simple sum of variables aggregation scheme, the contemporaneously aggregated MA parameter \( \theta \) is:

\[
\theta = \delta \pm \sqrt{\delta^2 - 1},
\]

with
\[
|\delta| > 1,
\]

where
\[
\delta = \frac{\sum_{i=1}^{N} \sigma_i^2 + \sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sigma_{ij} + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \alpha_i \alpha_j \sigma_{ij}}{2 \left( \sum_{i=1}^{N} \alpha_i \sigma_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \alpha_j \sigma_{ij} \right)}.
\]

\( \alpha_N = \sum_{i=1}^{N} \theta_{iN} \). Note that when \( \delta < -1 \) the invertible solution (i.e., \( |\theta| < 1 \)) is \( \theta = \delta + \sqrt{\delta^2 - 1} \), when \( \delta > 1 \) the invertible solution is \( \theta = \delta - \sqrt{\delta^2 - 1} \). Concerning the innovation variance of the contemporaneously aggregated model,\(^5\) it holds:

\[
\sigma_a^2 = \frac{\sum_{i=1}^{N} \alpha_i \sigma_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \alpha_j \sigma_{ij}}{\theta}.
\]

Finally, it can also be shown that the innovations of the contemporaneously aggregated model are a white noise process.

Proof. Let us define:
\[
\delta = \frac{E(y_{k(t-1)+1}^2)}{2E(y_{k(t-1)+1}y_{k(t-1)})}.
\]

Note that \( \delta \) in (8) represents the inverse of the autocorrelation function of the aggregate IMA(1,1) model, that is, the variance divided by (two times) the first-order autocovariance. Equation (6) follows from (8), after direct calculation of the variance and first-order autocovariance of the aggregate IMA(1,1) model. To recover the unknown parameters, the right-hand side (RHS) of (8) has to be equated to the inverse of the autocorrelation function of the contemporaneously aggregated IMA(1,1) model. Therefore:

\[
\frac{E(y_{k(t-1)+1}^2)}{2E(y_{k(t-1)+1}y_{k(t-1)})} = \frac{(1 + \theta^2)}{2\theta}.
\]

Equation (7), which gives the innovation variance of the contemporaneously aggregated model, stems directly from \( E(y_{k(t-1)} y_{k(t-1)}) = \theta \sigma_a^2 \), after simple calculation.

\(^5\)The expression in (5) was previously used by Ku and Seneta (1998) to show the relation between the parameters of two independent autoregressive processes of order one and those of the sum process.
We complete the proof by showing that the aggregate error $a_{k(t-1)+1} = \sum_{i=1}^{N} \varepsilon_{i,k(t-1)+1} + \sum_{i=1}^{N} \alpha_i \varepsilon_{i,k(t-1)} - \theta a_{k(t-1)}$, with $\alpha_N = \sum_{i=1}^{N} \theta_i N$, is a white noise process. It can be easily seen that the first-order autocovariance is:

$$E(a_{k(t-1)+1} a_{k(t-1)}) = \sum_{i=1}^{N} \alpha_i \sigma_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \alpha_j \sigma_{ij} - \theta E(a_{k(t-1)}^2) = 0.$$  

Higher-order autocovariances are equal to zero, as well.

To illustrate how the previous results can be applied in practice, in the sequel we propose a simple numerical example.

**Example 3.1** Consider the following trivariate system of equations:

\[
\begin{pmatrix}
y_{1,k(t-1)+1} \\
y_{2,k(t-1)+1} \\
y_{3,k(t-1)+1}
\end{pmatrix} =
\begin{pmatrix}
y_{1,k(t-1)} \\
y_{2,k(t-1)} \\
y_{3,k(t-1)}
\end{pmatrix} + \begin{pmatrix}
1 + 0.587L & -0.087L & 0.3392L \\
-0.216L & (1 + 0.6201L) & 0.3452L \\
-0.277L & 0.1241L & (1 + 0.5642L)
\end{pmatrix} \begin{pmatrix}
\varepsilon_{1,k(t-1)+1} \\
\varepsilon_{2,k(t-1)+1} \\
\varepsilon_{3,k(t-1)+1}
\end{pmatrix}.
\]

If the variance-covariance matrix is

\[
\Sigma = \begin{pmatrix}
0.182 & 0.5 & 0.7 \\
0.5 & 0.14 & 0.1725 \\
0.7 & 0.1725 & 0.16
\end{pmatrix},
\]

then the simple sum of variables yields $y_{k(t-1)+1} = (1 + 0.7646L)a_{k(t-1)+1}$, with $\sigma_a^2 = 2.5545$.

In the following corollary we extend the previous results to cover the aggregation scheme of a weighted sum of variables.

**Corollary 1** Assuming a weighted aggregation scheme, the contemporaneously aggregated MA parameter $\theta$ can be recovered as:

$$\theta = \delta \pm \sqrt{\delta^2 - 1}, \quad \text{with } |\delta| > 1,$$

To guarantee invertibility (i.e., $|\theta| < 1$), when $\delta < -1$ the invertible solution is $\theta = \delta + \sqrt{\delta^2 - 1}$, on the contrary, when $\delta > 1$ the invertible solution is $\theta = \delta - \sqrt{\delta^2 - 1}$. Moreover

$$\delta = \frac{\sum_{i=1}^{N} \sigma_i^2 \omega_i^2 + \sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 \omega_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sigma_{ij} \omega_i \omega_j + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \alpha_i \alpha_j \sigma_{ij} \omega_i \omega_j}{2 \left( \sum_{i=1}^{N} \sigma_i^2 \omega_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \alpha_i \sigma_{ij} \omega_i \omega_j \right)}.$$

The innovation variance of the contemporaneously aggregated model is:

$$\sigma_a^2 = \frac{\sum_{i=1}^{N} \alpha_i \sigma_i^2 \omega_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \alpha_j \sigma_{ij} \omega_i \omega_j}{\theta}.$$
4 Parameters of the temporally aggregated process

In the previous section we have given the parameters of the contemporaneously aggregated process. After contemporaneous aggregation, the macro process in (1) simplifies to:

\[
\begin{pmatrix}
1 & 0 & \ldots & 0 \\
-1 & 1 & \ldots & 0 \\
0 & -1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
y_{k(t-1)+1} \\
y_{k(t-1)+2} \\
y_{k(t-1)+3} \\
\vdots \\
y_{k(t-1)+k}
\end{pmatrix}
= \begin{pmatrix}
0 & 0 & \ldots & 1 \\
0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0
\end{pmatrix}
\begin{pmatrix}
y_{k(t-2)+1} \\
y_{k(t-2)+2} \\
y_{k(t-2)+3} \\
\vdots \\
y_{k(t-2)+k}
\end{pmatrix}
+ \begin{pmatrix}
1 & 0 & \ldots & 0 \\
\theta & 1 & \ldots & 0 \\
0 & \theta & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
a_{k(t-1)+1} \\
a_{k(t-1)+2} \\
a_{k(t-1)+3} \\
\vdots \\
a_{k(t-1)+k}
\end{pmatrix}
+ \begin{pmatrix}
0 & 0 & \ldots & \theta \\
0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0
\end{pmatrix}
\begin{pmatrix}
a_{k(t-2)+1} \\
a_{k(t-2)+2} \\
a_{k(t-2)+3} \\
\vdots \\
a_{k(t-2)+k}
\end{pmatrix}
\quad t \in \mathbb{N}.
\]

(10)

In (10), each row corresponds to an IMA(1,1) process. Without loss of generality, let us focus on the last row:

\[
(1 - L)y_{kt} = \theta(L)a_{kt},
\]

(11)

where \(a_{kt}\) is a white noise innovation term such that \(a_{kt} \sim (0, \sigma^2_a)\). \(\theta(L) = 1 + \theta L\) is a moving average polynomial of length 1 and \(|\theta| < 1\). The random variable \(y_{kt}\) is observed at the disaggregate frequency \(t\). It is a realization of the stochastic process \(\{y_t\}_{t=0}^{\infty}\).

We define the temporally aggregated process as:

\[
y_{\tau} = \sum_{j=0}^{k-1} y_{kt-j},
\]

(12)

Consequently, \(\{y_{\tau}\}_{\tau=0}^{\infty}\) evolves in aggregate time units \(\tau\) according to the aggregation scheme in (12). After temporal aggregation, sample information for the random variable at the aggregate frequency is available only every \(k^{th}\) period \((k, 2k, 3k, \ldots)\), where \(k\), an integer value larger that one, is the aggregation frequency or the order of temporal aggregation. It is interesting to note that the aggregation scheme in (12) corresponds to a linear transformation applied to the macro process in (10), that is, multiplying (10) by an aggregation vector of ones of size \((1 \times k)\). Besides (12), other important aggregation schemes are possible: namely, systematic sampling, weighted averaging and phase averaging sampling. We refer to Silvestrini and Veredas (2008), among others, for a discussion.

\footnote{We can focus on any of the rows of (10). Results in Section 4 are unaffected.}
Following Weiss (1984), Stram and Wei (1986), Wei (1990) and Marcellino (1999), it can be shown that the temporal aggregation of (10) is an IMA(1, \( r \)) vector process with

\[
\gamma = \left\lfloor \frac{2(k - 1) + 1}{k} \right\rfloor,
\]

where the lag operator is in aggregate time units, say \( B = L^k \), and \( \lfloor b \rfloor \) indicates the integer part of a real number \( b \).

The following Lemma is a straightforward consequence of (13):

**Lemma 2** The temporal aggregation of the IMA(1,1) process in (10) is an IMA(1,1) in aggregate time units, whatever the aggregation frequency \( k \).

As a consequence, the temporal aggregation in (12) corresponds to an IMA(1,1), which may be represented as:

\[
(1 - L^k)y_{\tau} = (1 + \Theta L^k)a_{\tau},
\]

where \( y_{\tau} \) and \( a_{\tau} \) are the temporally aggregated time series and innovation sequence, respectively. Moreover, \( \Theta \) is the MA parameter and \( \sigma_a^2 \) is the innovation variance of the temporally aggregated process in (14).

Stram and Wei (1986) show that it is possible to express the autocovariances of the temporally aggregated model as a function of the autocovariances of the original model. Actually, their discussion is valid for a generic ARIMA\((p,d,q)\) model, hence also in the special case of an IMA(1,1). To calculate the parameters of the temporally aggregated model, \( \Theta \) and \( \sigma_a^2 \), they suggest to equalize the autocovariance structures of the models in (11) and in (14). Indicating with \( \Gamma(j) \) and \( \gamma(j) \) the j-th order autocovariances for the models in (14) and in (11), respectively, they prove that this relationship between the autocovariances of the aggregate and the disaggregate IMA(1,1) model holds:

\[
\Gamma(j) = (1 + L + \ldots + L^{k-1})^4 \gamma(k(j + 2) - 2), \quad j = 0, 1, \ldots.
\]

Note that, in (15), \( (1 + L + \ldots + L^{k-1}) \) is the temporal aggregation polynomial.

Let us denote with \( \Gamma \) and \( \gamma \) the column vectors where the j-th element is equal to \( \Gamma(j) \) and \( \gamma(j) \), respectively. We can express equation (15) in matrix form, that is, \( \Gamma = \tilde{A}_k(3k - 2)\gamma \), or equivalently:

\[
\begin{pmatrix}
\Gamma(0) \\
\Gamma(1) \\
\vdots \\
\Gamma(3k - 2)
\end{pmatrix} = \tilde{A}_k(3k - 2)
\begin{pmatrix}
\gamma(0) \\
\gamma(1) \\
\vdots \\
\gamma(3k - 2)
\end{pmatrix}.
\]

On the left-hand side (LHS) of (16), \( \Gamma \) is a \((2 \times 1)\) vector. On the RHS of (16), \( \tilde{A}_k(3k - 2) \) is a \((2 \times (3k - 1))\) aggregation matrix while \( \gamma \) is a \(((3k - 1) \times 1)\) vector. Note that \( \Gamma \) is a vector containing the autocorrelations of the temporally aggregated model; \( \gamma \), conversely, is a vector with the autocorrelations of the model at the disaggregate frequency. Stram and Wei (1986) suggest to solve the system in (16) to derive the parameters of the temporally aggregated model.
A second alternative method to calculate the parameters of the temporally aggregated IMA(1,1) model is proposed by Weiss (1984). It requires to apply the summing operator \((1-L)^k\) to the model in (11). If \(k = 12\), equation (11) becomes:

\[
(1 - L^{12})^2 (1 - L) y_{2t} = (1 - L^{12})^2 (1 + \theta L) a_{2t}.
\]  

(17)

Since \(y_\tau = (1-L^{12}) y_{2t}\), we get:

\[
(1 - L^{12}) y_\tau = (1 - L^{12})^2 (1 + \theta L) a_{2t}.
\]

Once all the calculations on the RHS side of (17) have been developed, we equate the autocovariance structures of the resulting model and of the temporally aggregated one which, as we know, is an IMA(1,1). In this way we can infer the temporally aggregated parameters. Note that the calculations may become very cumbersome as the aggregation frequency increases.

The two methods outlined here above are already known in the literature. A third alternative approach is proposed in what follows. In particular, for an IMA(1,1) model at the disaggregate frequency, we can prove the following result, which enables us to write down the system that has to be solved to calculate the unknown parameters of the temporally aggregated model. This system represents the relationship between the autocorrelation function of the original IMA(1,1) model and the temporally aggregated counterpart, which is again an IMA(1,1).

**Proposition 2** The parameters of the temporally aggregated model in (14) may be obtained by solving the system:

\[
\begin{align*}
\Gamma(0) &= (1 + \Theta^2) \sigma_a^2 = \left( \sum_{j=0}^{k-1} ((j + 1) + j \theta)^2 + \sum_{j=k}^{2k-1} ((2k - j - 1) + (2k - j) \theta)^2 \right) \sigma_a^2 \\
\Gamma(1) &= \Theta \sigma_a^2 = \left( \sum_{j=k}^{2k-1} ((2k - j - 1) + (2k - j) \theta)((j + 1 - k) + (j - k) \theta) \right) \sigma_a^2
\end{align*}
\]

(18)

To calculate the MA parameter of the temporally aggregated IMA(1,1) process, the system in (18) can be expressed as

\[
\frac{1 + \Theta^2}{2\Theta} = \Delta,
\]

where

\[
\Delta = \frac{\sum_{j=0}^{k-1} ((j + 1) + j \theta)^2 + \sum_{j=k}^{2k-1} ((2k - j - 1) + (2k - j) \theta)^2}{2 \left( \sum_{j=k}^{2k-1} ((2k - j - 1) + (2k - j) \theta)((j + 1 - k) + (j - k) \theta) \right)}.
\]

which delivers

\[
\Theta^2 - 2\Delta \Theta + 1 = 0 \Rightarrow \Theta = \Delta \pm \sqrt{\Delta^2 - 1}, \quad \text{with} \quad |\Delta| > 1.
\]

(19)

As mentioned in Section 3, the condition \(|\Delta| > 1\) guarantees invertibility of the temporally aggregated process.
Finally, once $\Theta$ is known, $\sigma_a^2$ can be obtained as

$$\sigma_a^2 = \frac{\left(\sum_{j=k}^{2k-1}((2k-j-1) + (2k-j)\theta)\right)(j+1+k) + (j-k)\theta)}{\Theta} \sigma_a^2. \tag{20}$$

**Proof.** See the Appendix. \qed

Once the analytic form of the system is known, it is immediate to calculate the unknown parameters of the temporally aggregated model. In this way, as illustrated in the following example, we show that no complicated numerical algorithms are needed to recover the temporally aggregated parameters and that the calculation procedure has achieved some simplification.

**Example 4.1** Let us assume, as in (17), a temporal aggregation frequency $k = 12$. In practice, this situation arises when, for instance, we are temporally aggregating a monthly time series to annual frequency. Reminding that $\Gamma(j)$ and $\gamma(j)$ are the $j$-th order autocovariances for the models in (14) and in (11), respectively, and applying (18), we obtain the variance and the first-order autocovariance:

$$
\gamma(0) = (1 + (2 + \theta)^2 + (3 + 2\theta)^2 + (4 + 3\theta)^2 + (5 + 4\theta)^2 + (6 + 5\theta)^2 + (7 + 6\theta)^2 + (8 + 7\theta)^2 + (9 + 8\theta)^2 + (10 + 9\theta)^2 + (11 + 10\theta)^2 + (12 + 11\theta)^2 + (10 + 11\theta)^2 + (9 + 10\theta)^2 + (8 + 9\theta)^2 + (7 + 8\theta)^2 + (6 + 7\theta)^2 + (5 + 6\theta)^2 + (4 + 5\theta)^2 + (3 + 4\theta)^2 + (2 + 3\theta)^2 + (1 + 2\theta)^2 + \theta^2)\sigma_a^2
$$

$$= (1156 + 2288\theta + 1156\theta^2)\sigma_a^2;$$

$$
\gamma(1) = ((11 + 12\theta)(1 + 2\theta) + (10 + 11\theta)(9 + 10\theta)(3 + 2\theta) + (8 + 9\theta)(4 + 3\theta) + (7 + 8\theta)(5 + 4\theta) + (6 + 7\theta)(6 + 5\theta) + (5 + 6\theta)(7 + 6\theta) + (4 + 5\theta)(8 + 7\theta) + (3 + 4\theta)(9 + 8\theta) + (2 + 3\theta)(10 + 9\theta) + (1 + 2\theta)(12 + 11\theta))\sigma_a^2
$$

$$= (286 + 584\theta + 286\theta^2)\sigma_a^2.$$

For the temporally aggregated IMA(1,1) model expressed in aggregate time units: $\Gamma(0) = (1 + \Theta^2)\sigma_a^2$, and $\Gamma(1) = \Theta\sigma_a^2$. Thus, the system that has to be solved to calculate the parameters of the temporally aggregated IMA(1,1) model is

$$
\Gamma(0) = (1 + \Theta^2)\sigma_a^2 = (1156 + 2288\theta + 1156\theta^2)\sigma_a^2
$$

$$\Gamma(1) = \Theta\sigma_a^2 = (286 + 584\theta + 286\theta^2)\sigma_a^2.$$

This corresponds to the second degree equation:

$$\frac{1 + \Theta^2}{2\Theta} = \Delta,$$

where

$$\Delta = \frac{(1156 + 2288\theta + 1156\theta^2)}{2(286 + 584\theta + 286\theta^2)}$$

and

$$\Theta = \Delta \pm \sqrt{\Delta^2 - 1}.$$

Last, as in the previous example, $\sigma_a^2$ is calculated on the basis of (20).
5 A financial application

In this section, we present an empirical application to the problem of Value-at-Risk (VaR) calculation, prediction and backtesting. In financial risk management, one of the central issues is to track certain features of the conditional distribution of returns $r_t$, such as portfolio’s VaR, which represents the $\alpha$-th quantile of the portfolio’s return conditional distribution, namely

$$\Pr[r_t < -VaR_t^{(\alpha)}|\mathcal{F}_{t-1}] = \alpha,$$

where $\mathcal{F}_{t-1}$ is the information set at time $t-1$ (see Tsay, 2005).

The exercise is applied to an equally weighted portfolio composed by NASDAQ and Standard & Poor’s 500 (S&P500) stock indices. Daily adjusted closing prices data, provided by Datastream, have been employed from July 09, 1999 to February 17, 2010 (2769 observations). A similar data set was analysed by Parra Palaro and Hotta (2006) and, more recently, by Rombouts and Verbeek (2009). A plot of the daily continuously compounded returns (i.e., $r_t = \log P_t - \log P_{t-1}$, $t = 2, \ldots, 2769$, where $P_t$ is the original price at time $t$) of NASDAQ and S&P500 is presented in Figure 1. The two time series display a few jumps, which are clearly visible along the whole sample and are particularly large for NASDAQ. Moreover, the NASDAQ and S&P500 financial log-returns exhibit the well known property of volatility clustering.

[FIGURE 1 ABOUT HERE]

Descriptive statistics of log-returns are displayed in Table 1. The two indices share similar features: both average returns are close to zero, although NASDAQ is more dispersed than S&P500. Both unconditional distributions have ticker tails than the Gaussian benchmark, while NASDAQ has lower kurtosis than S&P500. Furthermore, they are not symmetric: for NASDAQ, skewness is positive but close to zero; for S&P500, on the other hand, skewness is negative (i.e., the data are spread out more to the left of the mean) and the asymmetry is more pronounced. Consistently, the Jarque-Bera test, which is a test of distributional Gaussianity, rejects the null of normality for both indices at 5 percent level.

[TABLE 1 ABOUT HERE]

Figure 2 shows the sample autocorrelation functions (ACF) of daily log-returns of S&P500 and NASDAQ (panels a, b) and the ACF of daily squared log-returns (panels c, d). In each plot, the maximum time lag is 50. The starred lines represent the confidence bands of the sample autocorrelation. As expected, the ACF of daily log-returns is similar to that of a white noise: except in a very few cases, it is within the two standard error limits. Conversely, the plot of the ACF of the squared log-returns displays a strong form of serial correlation for both indices. Moreover, it exhibits a very slow decay.

[FIGURE 2 ABOUT HERE]

---

7VaR is important because it is the basis of risk measurement and has a variety of empirical applications, essentially in risk management and for regulatory requirements. For instance, Basel II Capital Accord imposes to financial institutions to meet capital requirements based on VaR estimates at a confidence level of 5 or 1 percent.
In the empirical application, log-returns series of NASDAQ (\(N_{\text{AS}} t\)) and S&P500 (\(SP_t\)) are considered such that an equally weighted portfolio of both indices is constructed. The log-returns of the equally weighted portfolio (\(PTF_t\)) may be approximated as

\[ PTF_t \approx 0.5N_{\text{AS}} t + 0.5SP_t. \]  

(21)

Equal weighting is a common index weighting scheme. It is used in this application to provide an illustration of contemporaneous aggregation.

The aim of the exercise is to estimate 1-period horizon VaR for the portfolio in (21), by employing the EWMA model for conditional volatility introduced by RiskMetrics\textsuperscript{TM} (1996). Of course, other approaches can be followed for volatility forecasting; e.g., univariate and multivariate ARCH/GARCH models have proven particularly useful in financial econometrics for modeling return volatilities.\textsuperscript{8} In this exercise we focus on the RiskMetrics\textsuperscript{TM} approach because it incorporates the EWMA model for estimating the conditional volatility. Moreover, it is easy to estimate and widely used for risk management purposes.

In the original RiskMetrics\textsuperscript{TM} (1996) methodology, the log prices are assumed to follow the process\textsuperscript{9}

\[ \log P_t = \log P_{t-1} + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim i.i.d. N(0,1). \]  

(22)

where \(\varepsilon_t\) is an independent and identically distributed (i.i.d.) Gaussian sequence. As a consequence of (22), the compounded log-returns \(r_t\) are conditionally Gaussian and uncorrelated:

\[ r_t | \mathcal{F}_{t-1} \sim i.i.d. N(0, \sigma_t^2). \]  

(23)

In RiskMetrics\textsuperscript{TM} (1996), the conditional volatility \(\sigma_t^2\) is assumed to evolve according to the following recursion

\[ \sigma_{t+1|t}^2 = \lambda \sigma_{t|t-1}^2 + (1-\lambda)r_t^2, \]  

(24)

where \(\lambda (0 < \lambda < 1)\), defined as the “decay factor”, is the smoothing parameter of the EWMA scheme. Note that equation (24) represents the 1-step-ahead volatility forecast. In addition, the recursion in (24) is exactly equivalent to an integrated GARCH(1,1) model without constant term for the conditional volatility of the log-returns. Finally, it can be easily shown that (24) can be re-parameterized as an IMA(1,1) model for the squared returns,

\[ (1-L)r_t^2 = \eta_t - \lambda \eta_{t-1}, \]  

(25)

where by definition \(\eta_t := r_t^2 - \sigma_{t|t-1}^2\).

Besides volatility, VaR calculation depends on a tail quantile of the portfolio’s log-return conditional distribution. Figure 3 shows the left tail of the histogram of the portfolio’s log-returns in (21). The daily returns have been normalized by dividing for the square root of the volatility in (24).\textsuperscript{10} For illustrative purposes, a Student-t distribution with twelve degrees of freedom (dotted line), rescaled to have a unit variance,

\textsuperscript{8} More recently, Ghysels, Rubio and Valkanov (2009) propose a mixed-data sampling (MIDAS) approach to calculate multi-period predictions of stock market return volatilities. We refer to Manganelli and Engle (2001) for a survey of the most popular univariate VaR methodologies and to Ghysels, Rubio and Valkanov (2009) for additional explanations on MIDAS.

\textsuperscript{9} See RiskMetrics\textsuperscript{TM} (Technical Document, 1996, p. 85).

\textsuperscript{10} Note that, to draw the normalized log-returns in Figure 3, the \(\lambda\) parameter has been fixed to 0.94 over the whole sample.
and a standard Normal (solid line) are displayed. The Student-t density seems to fit the data reasonably well. It is often well suited to deal with the fat-tailed and leptokurtic features. Indeed, consistently with residual properties and as suggested by Zumbach (2006), the graph suggests that the Student-t distribution provides a better description of the tails than the Gaussian distribution.

[FIGURE 3 ABOUT HERE]

In the empirical application, we calculate 1-period horizon VaR predictions for the portfolio in (21), working with stock price data sampled at six different time frequencies (1/2/3/5/10/20 business days). In this way we provide an illustration of how the temporal aggregation scheme works at relevant sampling frequencies. The EWMA recursion in (24) is used to deliver 1-step-ahead volatility forecasts.

Concerning the estimation of the EWMA model, RiskMetrics™ (1996) suggests to select the decay factor in (24) by searching for the smallest root mean squared prediction error over different values of \( \lambda \) (the resulting \( \lambda \) is termed “optimal decay factor”). Formally:

\[
\lambda_{opt} = \arg\min_{\lambda \in [\lambda_{\text{min}}, \lambda_{\text{max}}]} \frac{1}{T} \sum_{t=2}^{T} (r_t^2 - \sigma_{t-1}^2(\lambda))^2,
\]

where \([\lambda_{\text{min}}, \lambda_{\text{max}}]\) is a compact set over which the optimization takes place (e.g., [0,1]).

Zaffaroni (2008) points out that the estimation method in (26) is non-consistent for \( \lambda \). In addition, the resulting estimator lacks of the usual asymptotic statistical properties. As an alternative, he suggests the pseudo maximum likelihood estimator (PMLE), which requires to maximize a Gaussian pseudo log-likelihood function with respect to the unknown decay factor.

In the application, different procedures are examined to estimate \( \lambda \): we refer to “METHOD 1” if the decay factor is chosen as in (26), as in RiskMetrics™(1996), while “METHOD 2” is a maximum likelihood (ML) estimator. Since we make an explicit Student-t assumption for the normalized log-returns, we can estimate the decay factor by maximizing a Student-t log-likelihood. The log-likelihood function has two unknown parameters: the decay factor \( \lambda \) and the Student’s degrees of freedom parameter, which is estimated jointly with \( \lambda \). Note, therefore, that “METHOD 2” refers to the ML estimator, not to the PMLE estimator suggested by Zaffaroni (2008).

There is a third way to recover the decay factor: it can be directly inferred from the estimated parameters of a multivariate IMA(1,1) model for the vector of squared log-returns, applying the theoretical results on contemporaneous and temporal aggregation presented in the previous sections. We shall call “METHOD 3” our proposed approach for the determination of \( \lambda \). As a matter of fact, equation (25) allows the theoretical framework presented in this paper to be employed to calculate the decay factor. Given that the portfolio

\[11\]We are aware that different distributional assumptions can be made: yet, for simplicity, we focus only on the Student-t and on the Gaussian.

\[12\]We refer to See RiskMetrics™ (Technical Document, 1996, Section 5.3) for further details.

\[13\]The MATLAB function fmincon.m is used for ML estimation. The \( \lambda \) parameter has a lower bound of 0.005 and an upper bound equal to 0.995.
in (21) is the weighted sum of two indices at daily frequency, we propose to estimate a trivariate IMA(1,1)
model for the system composed by the two conditional variances and the autocovariance, namely,
\[
\text{vech}(r_t r_t' = \text{vech}(r_{t-1} r_{t-1}') + \text{vech}(\eta_t) - A \text{vech}(\eta_{t-1}),
\]
which is the multivariate version of (25) and where \(A\) is a 3x3 moving average coefficient matrix. Once
obtained the estimates for the MA matrix of coefficients, the decay factor is analytically computed for 1-day-
ahead portfolio’s volatility forecasts (i.e., \(\theta\) in (5)), applying the results on contemporaneous aggregation
of the IMA(1,1) model. Similarly, results on the temporal aggregation of the IMA(1,1) can be applied if
interested in a decay factor for portfolio’s volatility predictions at a different time frequency (i.e., \(\Theta\) in (19)).

Regarding vector IMA(1,1) estimation, we employ the method proposed by Galbraith, Ullah and Zinde-Walsh (2002),
which relies on a vector autoregressive (VAR) approximation to the vector MA process. This
procedure consists of basically two steps. In the first step, a VAR model is estimated by the Yule-Walker
estimator (Lütkepohl, 2007, Section 3.3.4) with a lag length (\(p > 1\)) that allows to recover the vector of
innovations whose properties are close to the vector MA process. We use the Yule-Walker estimator since it
delivers estimates of the VAR matrix coefficients which are always in the stability region. In the second
step, the vector MA coefficient matrices are obtained by using some known relations between the coefficient
matrices of an infinite-order VAR model and their vector MA counterparts (see, for instance, Lütkepohl,
2007, Section 2.1.2).

Somehow, it can be seen that the estimation of the decay factor based on the contemporaneously and
temporally aggregated vector IMA(1,1) model can be considered as a variant of the RiskMetrics\textsuperscript{TM}
approach and of the standard ML, bearing in mind that all these procedures aim to calculate in different ways the
decay factor, which is subsequently used as an input for the EWMA estimator of the portfolio’s conditional
volatility in (24).

In order to rank these three estimation methods and to determine the accuracy of the resulting VaR
estimates, we use standard backtests of VaR models. Backtesting is a statistical procedure consisting of
calculating the percentage of times that the actual portfolio returns fall outside the VaR estimate, and
comparing it to the confidence level used.

To implement the backtesting exercise, we carry out an out-of-sample analysis. We consider 1-step ahead
forecasts using a fixed rolling window scheme. The size of the window is set to 1000 observations, i.e.,
to generate VaR estimates, a training period of 1000 observations of past returns is used. The use of a
fixed rolling window scheme is very common in the financial econometrics literature, see, e.g., Fama and
MacBeth (1973), and is standard practice in RiskMetrics\textsuperscript{TM}.

We split the sample in 2 parts: from \(t=1\) to \(t=1000\); from \(t=1001\) to \(t=2768\). The second part of the
sample is used as out-of-sample validation period. For all estimation methods under consideration, the
decay factor is chosen over a range of values comprised between \(\lambda_{\text{min}} = 0.005\) and \(\lambda_{\text{max}} = 0.995\). The
three examined methods are implemented in order to calculate 1-step-ahead volatility forecasts, based on

---

\(14\) We are grateful to Helmut Lütkepohl who pointed out this property.
\(15\) Due to the type of exercise, in selecting the VAR(p) lag length, we adopt the criterion in (26), as for the RiskMetrics\textsuperscript{TM} approach. That is, the choice of the lag length of the VAR is made in order to analytically obtain a decay factor that minimizes the RMSE over different values.
\(16\) The developed MATLAB code is available from the authors upon request.
(24), and 1-period horizon VaR predictions. Using daily data with a rolling estimation window of 1000 observations and starting the out-of-sample forecasting at \( t=1001 \), we have a total of 1767 estimates of the aggregated MA parameter ("METHOD 3") and of the decay factor ("METHOD 1" and "METHOD 2").

When a different time frequency is adopted for VaR prediction, for "METHOD 1" and "METHOD 2", the decay factors are estimated starting directly from the aggregated time series. For "METHOD 3", on the other hand, the decay factor is systematically sampled at the corresponding desired frequency, such as the sequence,

\[
\{\lambda_{tk}\}_{t=1}^{1767/k}, \quad k = 1, 2, 3, 5, 10, 20,
\]

where \([b]\) indicates the integer part of a real number \( b \).

Working with daily data, we have a total of 1767 tests for VaR at each confidence level \((\alpha=0.10, 0.05, 0.025, 0.01)\). For backtesting, we count the proportion of observations where the actual portfolio loss exceeded the estimated VaR for \( \alpha=0.10, 0.05, 0.025, 0.01 \). This can be done by means of the exception indicator (at a confidence level \( \alpha \)), a binary variable which takes the value one if the log-returns exceed in absolute value the predicted VaR and zero otherwise:

\[
I_t(\alpha) = \begin{cases} 1 & \text{if } |r_t| > VaR_t(\alpha) \\ 0 & \text{otherwise} \end{cases}, \quad t = 1001, 1002, \ldots, 2768.
\]

Similarly at the other time frequencies. If the VaR model is correctly specified and the confidence level \( \alpha \) is, say, 95%, we expect portfolio log-returns to exceed the VaR estimates on about 5% of the cases. Formally:

\[
Pr(I_t(\alpha) = 1) = \alpha.
\]

Equation (28) represents the "unconditional coverage" hypothesis, which can be statistically tested. If the number of exceptions is significantly lower or higher than \( \alpha \), the VaR model contributes to an overestimation or underestimation of risk.

To test whether the hypothesis of correct unconditional coverage holds, we use the likelihood ratio test statistic proposed by Kupiec (1995), which under the null is given by

\[
-2 \log \left( \frac{1 - \hat{\alpha}}{\hat{\alpha}} \right)^{T-N_{ex}(\alpha)} \left( \frac{\hat{\alpha}}{\alpha} \right)^{N_{ex}(\alpha)} \sim \chi^2_{1}, \quad T \to \infty,
\]

where \( T \) is the number of data used for backtesting, \( N_{ex}(\alpha) = \sum_{t=1}^{T} I_t(\alpha) \) is the number of exceptions, \( \hat{\alpha} = \frac{N_{ex}(\alpha)}{T} \) is the proportion of exceptions (i.e., the percentage of log-returns exceeding the VaR predictions).

Tables 2, 3 and 4 report the empirical backtesting results. Table 2 refers to the standard RiskMetrics™ approach, Table 3 to ML estimation, Table 4 to vector IMA(1,1) estimation and subsequent aggregation of the estimated model. At all frequencies and for all the estimation methods, we adopt quantiles of a standardized Student’s t-distribution (with twelve degrees of freedom) to provide VaR predictions. The degrees of freedom have been fixed to twelve since this value was close to what we got, on average, when estimating the decay factor (and the Student’s degrees of freedom) with maximum likelihood, i.e., by applying “METHOD 2”. For a given significance level \( \alpha \), each cell in the table displays the out-of-sample
empirical coverage (i.e., proportion of exceptions) and, for VaR prediction, the p-value of the corresponding Kupiec test (in brackets). As already mentioned, $T$ is the length of the VaR sample. A p-value smaller than $\alpha$ implies a rejection of the null hypothesis (bold numbers).

Unsurprisingly, the three methods perform similarly and reasonably well for $T = 1767, 883, 589, 353$ (i.e., 1/2/3/5-days sampling frequencies). Evaluation results indicate that they pass the Kupiec test most of the times. At 3-days frequency ($T = 589$) and $\alpha = 0.01$, the observed failure rate for “METHOD 1” is significantly higher than the theoretical value at the specified confidence level. At 5-days frequency ($T = 353$) and $\alpha = 0.01$, both for “METHOD 1” and for “METHOD 3”, the observed failure rates are significantly higher than the theoretical values.

A notable exception is represented by the 10-days frequency and the 20-days frequency: in this case, “METHOD 2” is largely outperformed by “METHOD 3” and “METHOD 1”. This is due to the fact that smaller sample sizes are employed in ML estimation. Indeed, moving from high to low frequencies reduces the available information set. For instance, when $k = 20$, ML estimation is conducted on only 50 data points. As a consequence, this has a detrimental impact on the performance of the ML estimator such that, for 10-days and 20-days sampling frequencies, the p-values of the Kupiec test reject the null hypothesis at all the confidence levels (except 10 percent). Of course, “METHOD 3” is unaffected by this problem since the decay factor is always inferred from the vector MA coefficient matrix, which is estimated at daily frequency. In this way, “METHOD 3” provides a rule to infer estimates of the decay factor and volatility predictions virtually for every forecast horizon (although, in the context of VaR prediction, it seems unlikely to work with a sampling frequency greater than 20, corresponding to twenty business days).

In general, even when the null hypothesis is not rejected, it is worth noting that “METHOD 3” reports higher (or equal) p-values than “METHOD 1” most of the times. Moreover, comparing “METHOD 2” with “METHOD 3”, the ML procedure performs similarly to our multivariate approach, except at the lower frequencies (10/20-days) where “METHOD 3” outperforms “METHOD 2”. This seems to suggest the use of “METHOD 3” when moving from high to low frequencies, i.e., when risk evaluation has to be carried out at medium-term horizon.

To conclude, from the backtesting analysis, the vector IMA(1,1) approach for the decay factor calculation seems to perform at least as good as the standard RiskMetricsTM criterion. This result has to be evaluated with caution since it is based on a simple exercise and on a single data set. Further empirical research is needed to compare the different EWMA estimation methods in VaR analysis.

6 Concluding remarks

In this paper we deal with temporal and contemporaneous aggregation of IMA(1,1) processes, which constitute the equivalent ARIMA representation of the simple exponential smoothing. Relying on the closeness of VARMA processes with respect to linear transformations, we derive algebraic solutions for the unknown parameters of the aggregated model.

Contrary to the common knowledge, we provide evidence that no complicated numerical algorithms need to be implemented to infer the parameters of the aggregated model. Indeed, we derive algebraic solutions to recover them. In this way, we show that it is possible to establish a direct mapping between the parameters
of the original model and those of the model for the corresponding contemporaneously and temporally aggregated data. The mapping fully relies on autocovariances and can be easily obtained by applying the moment-based procedure discussed in the paper.

To illustrate the practical relevance of aggregation results for the vector IMA(1,1) model, we have presented an application of the techniques discussed to the calculation and prediction of VaR for a portfolio composed by NASDAQ and S&P500 stock indices, at different sampling frequencies. An empirical comparison shows that VaR predictions based on the vector IMA(1,1) model are at least as accurate as those based on the standard approach suggested by RiskMetrics™ or on maximum likelihood.

This paper represents a starting point. The aforementioned results leave many open empirical issues. Further research should investigate on comparing aggregate and disaggregate approaches in forecasting portfolio’s conditional volatility, in the framework of Exponentially Weighted Moving Average estimates of volatility. In other words, using the vector IMA(1,1) model, it would be interesting to empirically assess whether aggregating across volatility predictions for single assets is more accurate than predicting portfolio’s volatility directly.

7 APPENDIX

For the proof of Proposition 2, we need this preparatory Lemma.

**Lemma 3** It is possible to express the square of the polynomial $1 + L + L^2 + \ldots + L^{k-1}$, whatever $k$, as:

$$
\left( \sum_{j=0}^{k-1} L^j \right)^2 = \sum_{j=0}^{k-1} (j+1)L^j + \sum_{j=k}^{2k-1} (2k-j-1)L^j. 
$$

(30)

**Proof.** We use induction to prove Lemma 3.

- **Step 1.** Check. For $k = 1$:

$$
\left( \sum_{j=0}^{0} L^j \right)^2 = 1^2 = \sum_{j=0}^{0} (j+1)L^j + \sum_{j=1}^{1} (2-j-1)L^j = 1.
$$

- **Step 2.** Induction. We assume that the result holds true for $k = n$:

$$
\left( \sum_{j=0}^{n-1} L^j \right)^2 = \sum_{j=0}^{n-1} (j+1)L^j + \sum_{j=n}^{2n-1} (2n-j-1)L^j.
$$

We need to show that the result follows for $k = n + 1$, under the hypothesis it works for $k = n$, i.e.,

$$
\left( \sum_{j=0}^{n} L^j \right)^2 = \sum_{j=0}^{n} (j+1)L^j + \sum_{j=n+1}^{2n+1} (2n+1-j)L^j.
$$

(31)
Note that we can express a sum over \( n + 1 \) terms as a sum over the first \( n \) terms plus the final term. For \( k = n + 1 \), therefore, the square of the sum over \( n + 1 \) terms may be developed as:

\[
\left( \sum_{j=0}^{n} L^j \right)^2 = \left( \sum_{j=0}^{n-1} L^j + L^n \right)^2
\]

\[
= \left( \sum_{j=0}^{n-1} L^j \right)^2 + L^{2n} + 2 \left( \sum_{j=0}^{n-1} L^j \right) L^n
\]

\[
= \sum_{j=0}^{n-1} (j+1)L^j + \sum_{j=n}^{2n-1} (2n-j-1)L^j + L^{2n} + 2 \left( \frac{1-L^n}{1-L} \right) L^n \quad \text{(from the assumption)}
\]

\[
= \sum_{j=0}^{n-1} (j+1)L^j + \sum_{j=n}^{2n-1} (2n-j-1)L^j + \frac{L^n}{1-L} (2 - L^{n+1} - L^n).
\]

As a consequence:

\[
\left( \sum_{j=0}^{n} L^j \right)^2 = \sum_{j=0}^{n-1} (j+1)L^j + \sum_{j=n}^{2n-1} (2n-j-1)L^j + \frac{L^n}{1-L} (2 - L^{n+1} - L^n). \tag{32}
\]

To prove that this is the desired result for \( k = n + 1 \), from (31) and (32), it is enough to show that:

\[
\left( \sum_{j=0}^{n} (j+1)L^j + \sum_{j=n+1}^{2n+1} (2n+1-j)L^j \right) - \left( \sum_{j=0}^{n-1} (j+1)L^j + \sum_{j=n}^{2n-1} (2n-j-1)L^j \right) = \frac{L^n}{1-L} (2 - L^{n+1} - L^n). \tag{33}
\]

To show that the equation above is true, we develop (33) as:

\[
\left( \sum_{j=0}^{n} (j+1)L^j - \sum_{j=0}^{n-1} (j+1)L^j \right) + \left( \sum_{j=n+1}^{2n+1} (2n+1-j)L^j - \sum_{j=n}^{2n-1} (2n-j-1)L^j \right)
\]

\[
= (n+1)L^n + \left( \sum_{j=n+1}^{2n+1} (2n+1-j)L^j - \sum_{j=n}^{2n-1} (2n-j-1)L^j \right)
\]

\[
= (n+1)L^n + \left( \sum_{j=n+1}^{2n+1} (2n-j)L^j + \sum_{j=n+1}^{2n} L^j - \sum_{j=n}^{2n-1} (2n-j)L^j + \sum_{j=n}^{2n-1} L^j \right)
\]

\[
= 2L^n + L^{2n} + 2 \sum_{j=n+1}^{2n-1} L^j = 2L^n + L^{2n} + 2L^n \frac{L^n}{1-L} = \left( \frac{L^n}{1-L} + 1 \right) 2L^n + L^{2n}
\]

\[
= \left( \frac{1}{1-L} \right) 2L^n + L^{2n} = \frac{L^n}{1-L} (2 - L^{n+1} - L^n). \tag{34}
\]

Therefore (34) confirms (33). This completes Step 2. We have proven Lemma 3.

\[ \square \]

**Proof of Proposition 2.** The original model at the disaggregate frequency is an IMA(1,1), say,

\[
(1-L)y_{kt} = (1 + \theta L) a_{kt}. \tag{35}
\]
Bearing in mind the temporal aggregation scheme in (12), that we list here below for the sake of the exposition,

\[ y_\tau = \sum_{j=0}^{k-1} y_{kt-j} = \left( \frac{1 - L^k}{1 - L} \right) y_{kt}, \]

and applying it to both sides of (35), we get:

\[ (1 - L)y_\tau = (1 + \theta L) \left( \frac{1 - L^k}{1 - L} \right) a_{kt}. \]

Since the temporally aggregated model is an IMA(1,1) and sample information for \( y_\tau \) is available only every \( k^{th} \) period, the equation here above needs to be transformed into:

\[ (1 - L^k)y_\tau = (1 + \theta L) \left( \frac{1 - L^k}{1 - L} \right)^2 a_{kt}. \]

Let us focus on the first row of (18). The autocovariance of order zero of the temporally aggregated model defined in (36) is:

\[ \Gamma(0) = E \left[ (1 - L^k)y_\tau \times (1 - L^k)y_\tau \right] \]

\[ = E[(1 + \theta L)(1 + L + \ldots + L^{k-1})^2a_{kt} \times (1 + \theta L)(1 + L + \ldots + L^{k-1})^2a_{kt}]. \]

From (30) we know that:

\[ \left( \sum_{j=0}^{k-1} jL^j \right)^2 = \sum_{j=0}^{k-1} (j + 1)L^j + \sum_{j=k}^{2k-1} (2k - j - 1)L^j. \]

Therefore, based on this, we can express \( (1 - L^k)y_\tau = (1 + \theta L)(1 + L + \ldots + L^{k-1})^2a_{kt} \) in (36) as:

\[ (1 - L^k)y_\tau = (1 + \theta L) \left( \sum_{j=0}^{k-1} (j + 1)L^j + \sum_{j=k}^{2k-1} (2k - j - 1)L^j \right) a_{kt}. \]
The expression here above can be written out as:

\[
(1 - L^k)y_r = \left( \sum_{j=0}^{k-1} (j + 1)L^j + \sum_{j=k}^{2k-1} (2k - j - 1)L^j + \sum_{j=0}^{k-1} (j + 1)\theta L^{j+1} + \sum_{j=k}^{2k-1} (2k - j - 1)\theta L^{j+1} \right) a_{kt}
\]

\[
= \left( \sum_{j=0}^{k-1} (j + 1)L^j + \sum_{j=k}^{2k-1} (2k - j - 1)L^j + \sum_{j=0}^{k-1} (j + 1)\theta L^{j+1} + \sum_{j=k}^{2k-1} (2k - j - 1)\theta L^{j+1} \right) a_{kt}
\]

\[
= \left( \sum_{j=0}^{k-1} (j + 1)L^j + \sum_{j=1}^{k-1} j\theta L^j + \sum_{j=k}^{2k-1} (2k - j - 1)L^j + \sum_{j=k+1}^{2k-1} (2k - j)\theta L^j \right) a_{kt}
\]

\[
= \left( 1 + \sum_{j=1}^{k-1} j\theta L^j + (k - 1) + k\theta) L^k + \sum_{j=k+1}^{2k-1} (2k - j - 1)L^j + \sum_{j=k+1}^{2k-1} (2k - j)\theta L^j \right) a_{kt}
\]

\[
= \left( 1 + \sum_{j=1}^{k-1} j\theta L^j + ((k - 1) + k\theta)L^k + \sum_{j=k+1}^{2k-1} (2k - j - 1)L^j + \sum_{j=k+1}^{2k-1} (2k - j)\theta L^j \right) a_{kt}
\]

Thus, the autocovariance of order zero is:

\[
\Gamma(0) = \left( 1 + \sum_{j=1}^{k-1} (j + 1) + j\theta)^2 + ((k - 1) + k\theta)^2 + \sum_{j=k+1}^{2k-1} ((2k - j - 1) + (2k - j)\theta)^2 \right) \sigma_a^2
\]

\[
= \left( \sum_{j=0}^{k-1} ((j + 1) + j\theta)^2 + \sum_{j=k}^{2k-1} ((2k - j - 1) + (2k - j)\theta)^2 \right) \sigma_a^2,
\]

\[
(38)
\]

as we aimed to prove.

Let us switch to the second row of (18). We need to calculate the first-order autocovariance of the temporally aggregated model in (36), that is:

\[
\Gamma(1) = E \left[ (1 - L^k)y_r \times (1 - L^k)y_{r-1} \right]
\]

\[
= E[(1 + \theta L)(1 + L + \ldots + L^{k-1})^2 a_{kt} \times (1 + \theta L)(1 + L + \ldots + L^{k-1})^2 a_{k(t-1)}].
\]

From (38), we already know that:

\[
(1 - L^k)y_r = \left( 1 + \sum_{j=1}^{k-1} (j + 1) + j\theta)L^j + ((k - 1) + k\theta)L^k + \sum_{j=k+1}^{2k-1} ((2k - j - 1) + (2k - j)\theta)L^j \right) a_{kt}.
\]

Similarly:

\[
(1 - L^k)y_{r-1} = \left( L^k + \sum_{j=k+1}^{2k-1} ((j + k - 1) + (j - k)\theta)L^j + ((k - 1) + k\theta)L^{2k} + \sum_{j=2k+1}^{3k-1} ((3k - j - 1) + (3k - j)\theta)L^j \right) a_{kt}.
\]

22
Consequently:

\[
\Gamma(1) = E \left[ (1 - L^k)^{y_r} \times (1 - L^k)^{y_{r-1}} \right]
\]

\[
= \left( (k - 1) + k\theta + \sum_{j=k+1}^{2k-1} ((2k - j - 1) + (2k - j)\theta)((j + 1 - k) + (j - k)\theta) \right) \sigma_a^2
\]

\[
= \left( \sum_{j=k}^{2k-1} ((2k - j - 1) + (2k - j)\theta)((j + 1 - k) + (j - k)\theta) \right) \sigma_a^2.
\]

The second part of Proposition 2 follows from the properties of second degree equations. The proof is now complete. \qed
References


March 11, 2010
Table 1: NASDAQ and S&P500 log-returns summary statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>NASDAQ</th>
<th>S&amp;P500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>2768</td>
<td>2768</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.132546</td>
<td>0.109572</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.101684</td>
<td>-0.094695</td>
</tr>
<tr>
<td>Mean</td>
<td>-8.193816e-005</td>
<td>-8.813140e-005</td>
</tr>
<tr>
<td>Median</td>
<td>2.046986e-004</td>
<td>2.636947e-005</td>
</tr>
<tr>
<td>Range</td>
<td>0.234231</td>
<td>0.204267</td>
</tr>
<tr>
<td>St.Dev.</td>
<td>0.018709</td>
<td>0.013596</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.091602</td>
<td>-0.104871</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.344184</td>
<td>10.965030</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>2.180435e+003</td>
<td>7.322017e+003</td>
</tr>
</tbody>
</table>


Table 2: Backtesting VaR Models: Proportion of exceptions ($\hat{\alpha}$)

<table>
<thead>
<tr>
<th>METHOD</th>
<th>CONFIDENCE LEVEL ($\alpha$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha=0.10$</td>
</tr>
<tr>
<td></td>
<td>$T$</td>
</tr>
<tr>
<td>ESTIMATION: MEAN SQUARED PREDICTION ERROR MINIMIZATION</td>
<td></td>
</tr>
<tr>
<td>METHOD 1</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>883</td>
</tr>
<tr>
<td></td>
<td>589</td>
</tr>
<tr>
<td></td>
<td>353</td>
</tr>
<tr>
<td></td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>88</td>
</tr>
</tbody>
</table>

At all frequencies, a Student-t distribution with 12 degrees of freedom is used to provide a VaR estimate. A p-value smaller than $\alpha$ implies a rejection of the null hypothesis (bold numbers). $T$ denotes the length of the VaR sample.
Table 3: Backtesting VaR Models: Proportion of exceptions (\(\hat{\alpha}\))

<table>
<thead>
<tr>
<th>METHOD</th>
<th>CONFIDENCE LEVEL ((\alpha))</th>
<th>T</th>
<th>(\alpha=0.10)</th>
<th>(\alpha=0.05)</th>
<th>(\alpha=0.025)</th>
<th>(\alpha=0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ESTIMATION: MAXIMUM LIKELIHOOD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>METHOD 2</td>
<td>1767</td>
<td>0.1076</td>
<td>(0.2931)</td>
<td>0.0532</td>
<td>(0.5378)</td>
<td>0.0255</td>
</tr>
<tr>
<td>METHOD 2</td>
<td>883</td>
<td>0.0964</td>
<td>(0.7180)</td>
<td>0.0601</td>
<td>(0.1820)</td>
<td>0.0295</td>
</tr>
<tr>
<td>METHOD 2</td>
<td>589</td>
<td>0.0986</td>
<td>(0.9123)</td>
<td>0.0527</td>
<td>(0.7640)</td>
<td>0.0340</td>
</tr>
<tr>
<td>METHOD 2</td>
<td>353</td>
<td>0.0824</td>
<td>(0.2573)</td>
<td>0.0625</td>
<td>(0.2995)</td>
<td>0.0369</td>
</tr>
<tr>
<td>METHOD 2</td>
<td>176</td>
<td>0.1314</td>
<td>(0.1839)</td>
<td>0.0914</td>
<td>(0.0235)</td>
<td>0.0686</td>
</tr>
<tr>
<td>METHOD 2</td>
<td>88</td>
<td>0.1264</td>
<td>(0.4278)</td>
<td>0.1034</td>
<td>(0.0441)</td>
<td>0.0805</td>
</tr>
</tbody>
</table>

At all frequencies, a Student-t distribution with 12 degrees of freedom is used to provide a VaR estimate. A p-value smaller than \(\alpha\) implies a rejection of the null hypothesis (bold numbers). \(T\) denotes the length of the VaR sample.
Table 4: Backtesting VaR Models: Proportion of exceptions ($\hat{\alpha}$)

<table>
<thead>
<tr>
<th>METHOD</th>
<th>T. AGGR.</th>
<th>CONFIDENCE LEVEL ($\alpha$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FREQ. (k)</td>
<td>$\alpha=0.10$</td>
</tr>
<tr>
<td>1</td>
<td>1767</td>
<td>0.1059 (0.4134)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>883 0.0964 (0.7180)</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>589 0.0986 (0.9123)</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>353 0.0852 (0.3444)</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>176 0.1086 (0.7089)</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>88 0.1149 (0.6492)</td>
</tr>
</tbody>
</table>

$T$ is the number of data used in the backtesting. ‘T. AGGR. FREQ.’ stands for Temporal Aggregation Frequency of the IMA(1,1) model for squared returns in (25). The estimation of the multivariate IMA(1,1) model is based on VAR approximation (see Galbraith, Ullah, Zinde-Walsh, 2002). The maximum VAR lag length is 100. At all frequencies, a Student-t distribution with 12 degrees of freedom is used to provide a VaR estimate. A p-value smaller than $\alpha$ implies a rejection of the null hypothesis (bold numbers). $T$ denotes the length of the VaR sample.
Figure 1: Daily log-returns series of the S&P500 and NASDAQ indices from July 09, 1999 to February 17, 2010.
Figure 2: Autocorrelation function of S&P500 and NASDAQ log-returns and squared log-returns.
Figure 3: Histogram (left tail) of the daily normalized log-returns of the portfolio in (21).