Aggregation of exponential smoothing processes with an application to portfolio risk evaluation

This version: July 5, 2011

Abstract

In this paper we propose a unified framework to analyse contemporaneous and temporal aggregation of a widely employed class of integrated moving average (IMA) models. We obtain a closed-form representation for the parameters of the contemporaneously and temporally aggregated process as a function of the parameters of the original one. These results are useful due to the close analogy between the integrated GARCH(1,1) model for conditional volatility and the IMA(1,1) model for squared returns, which share the same autocorrelation function. In this framework, we present an application dealing with Value-at-Risk (VaR) prediction at different sampling frequencies for an equally weighted portfolio composed of multiple indices. We apply the aggregation results by inferring the aggregate parameter in the portfolio volatility equation from the estimated vector IMA(1,1) model of squared returns. Empirical results show that VaR predictions delivered using this suggested approach are at least as accurate as those obtained by applying standard univariate methodologies, such as RiskMetrics.

Keywords: Contemporaneous and temporal aggregation; ARIMA; Volatility; Value-at-Risk

JEL Code: C10, C32, C43
1 Introduction

The aggregation of ARIMA models has been extensively studied in the econometric literature (see, e.g., Amemiya and Wu, 1972, Palm and Nijman, 1984, Nijman and Palm, 1990, Drost and Nijman, 1993). According to a number of existing results, the parameters and the orders of the aggregated models can be derived by establishing the relationship between the autocovariance structures of the disaggregated and aggregated models. The aggregated parameters are not always easy to determine, especially when dealing with multivariate models, since in most of the cases it is necessary to solve nonlinear systems of equations to recover them, as detailed by Granger and Morris (1976), Stram and Wei (1986), Wei (1990), Marcellino (1999) and Hafner (2008).

This paper focuses on multivariate integrated moving average (IMA) models and obtains closed-form representations for the parameters of the contemporaneously and temporally aggregated model as a function of the parameters of the original one. In particular, assuming that the data generating process is a vector IMA(1,1), our first contribution is to provide closed-form expressions for the autocovariances of the aggregated model. Those allow to derive exact functions linking the unknown aggregate parameters with the data generating process. 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.

The IMA(1,1) model is well known in time series analysis since its predictions take the form of the Exponentially Weighted Moving Average (EWMA) recursion, whose forecast function depends on a single smoothing parameter which expresses the weight by which past observations are discounted. In addition, as discussed by Harvey, Ruiz and Shephard (1994), it is possible to establish a close analogy between the integrated GARCH(1,1) - hereafter IGARCH(1,1) - model (Engle and Bollerslev, 1986) for the conditional volatility and the IMA(1,1) model for squared returns, which share the same autocorrelation function: our simulation results seem confirming that the IMA(1,1) model can be considered as a valid candidate in estimating the IGARCH(1,1). The ARMA representation of GARCH models has been widely employed in the context of temporal and contemporaneous aggregation (see Drost and Nijman, 1993; Nijman and Sentana, 1996; Hafner, 2008) and by Francq and Zakoian (2000), who exploit it to propose a least-squares estimator of weak GARCH models. In addition, the EWMA model is widely used by practitioners to produce forecasts of volatilities of financial data. It has been popularized by RiskMetrics™, a risk management
methodology for measuring market risk developed by J.P. Morgan based on the Value-at-Risk (VaR) concept.

VaR is one of the standard measures used to quantify market risk in the financial industry. VaR is the basis of risk measurement and has a variety of 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 1 percent (BCBS, 1996).

VaR evaluation (and prediction) is also the focus of our empirical application, which is an additional contribution of the paper. In the application, we illustrate how the results on temporal and contemporaneous aggregation are useful, for instance, in analysing the problem of volatility prediction and VaR calculation for a portfolio of multiple assets (contemporaneous aggregation) at several sampling frequencies (temporal aggregation). Suppose to be interested in forecasting the volatility of a portfolio composed of several assets, whose individual volatilities are driven by IGARCH(1,1) processes: the portfolio volatility can be described by specifying a univariate model for the portfolio log-returns (portfolio approach) or by contemporaneously aggregating a multivariate model for the system of asset returns contained in the portfolio (asset-level approach). Modeling the joint behaviour of the assets in a multivariate fashion can lead to more efficient estimates than those obtained by estimating univariate volatility models and to forecast improvements, mainly because a larger information set is used. At the same time, one can be interested in building portfolio volatility models working with time-aggregated returns and in obtaining time-aggregated VaR measures: Basel II, for example, requires backtesting for an horizon of ten trading days, suggesting the square-root-of-time rule of thumb (see, e.g., Wang, Yeh and Cheng, 2011). Temporal aggregation allows to address this issue straightforwardly.

In this framework, a method is proposed for deriving the MA coefficient and thus the aggregate parameter in the IGARCH(1,1) volatility equation of the portfolio, for different levels of aggregation frequency of the returns. This requires estimating a multivariate IMA(1,1) model by approximate methods, and inferring the aggregate integrated GARCH parameter as a function of the MA matrix coefficients. Our results show that VaR predictions based on the aggregation of the vector IMA(1,1) model are at least as accurate as those delivered by other benchmark univariate methods such as the Student-t EWMA maximum likelihood estimator (MLE), the Student-t GARCH(1,1) MLE and the RiskMetrics™ approach.
The rest of this paper is structured as follows. In Section 2 we set up the econometric framework, which is based on 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 first discuss the analogy between the IMA(1,1) and IGARCH(1,1) models, also from an estimation point of view; we then present the empirical application and its results. Section 6 concludes. The proofs are relegated to an appendix.

2 A joint framework for contemporaneous and temporal aggregation of the vector IMA(1,1) model

A framework to deal with contemporaneous and temporal aggregation of vector ARMA (VARMA) type of models has been suggested and formalized by Lütkepohl (1984a, 1984b, 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 simultaneous temporal and contemporaneous aggregation. Furthermore, 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.

To cover the aggregation issue in full generality, the econometric framework is based on the commonly named “macro processes” above mentioned, introduced by Lütkepohl (1987). As it will become clear in the sequel, this formulation allows to consider contemporaneous and temporal aggregation jointly.

We assume that the data generating process is an $Nk$–dimensional integrated moving average process of order one, that is, a vector IMA(1,1). In particular, we use a representation similar to the one proposed by Lütkepohl (2007, p. 441):
A few words of clarification are at order. The model in (1) is a vector IMA(1,1) with constant for the disaggregate data. In (1), each \( y_{k(i-1)+j} \), \( i = 1, 2, j = 1, 2, \ldots, k \), is an \((N \times 1)\) vector and each \( \varepsilon_{k(i-1)+j} \), \( i = 1, 2, j = 1, 2, \ldots, k \), is an \((N \times 1)\) vector. Furthermore, \( \theta \) is an \((N \times N)\) matrix. Note that \( k \) represents the temporal aggregation frequency. Similarly, \( N \) is the order of contemporaneous aggregation.\(^1\) We assume that both \( N \) and \( k \) are finite.

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

\[^1\)For instance, when \( k = 3 \), the original vector process is temporally aggregated over three subsequents periods. 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).
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 model

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

\[ y_{k(t-1)+1} = c + y_{k(t-1)} + \varepsilon_{k(t-1)+1} + \theta \varepsilon_{k(t-1)} \]  \hspace{1cm} (2)

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:

\[
\begin{pmatrix}
(1 - L)y_{1,k(t-1)+1} \\
(1 - L)y_{2,k(t-1)+1} \\
\vdots \\
(1 - L)y_{N,k(t-1)+1}
\end{pmatrix}
= 
\begin{pmatrix}
c_1 \\
c_2 \\
\vdots \\
c_N
\end{pmatrix}
+ 
\begin{pmatrix}
(1 + \theta_{11}L) & \theta_{12}L & \cdots & \theta_{1N}L \\
\theta_{21}L & (1 + \theta_{22}L) & \cdots & \theta_{2N}L \\
\vdots & \vdots & \ddots & \vdots \\
\theta_{N1}L & \theta_{N2}L & \cdots & (1 + \theta_{NN}L)
\end{pmatrix}
\begin{pmatrix}
\varepsilon_{1,k(t-1)+1} \\
\varepsilon_{2,k(t-1)+1} \\
\vdots \\
\varepsilon_{N,k(t-1)+1}
\end{pmatrix}
\]  \hspace{1cm} (3)

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 weak white noise innovations such that $E(\varepsilon_s) = 0$ and $E(\varepsilon_s \varepsilon'_s) = \Sigma_{(N \times N)}$. More specifically:

\[
E(\varepsilon_{k(t-1)+1} \varepsilon'_s) = 
\begin{pmatrix}
\sigma^2_1 & \sigma_{12} & \cdots & \sigma_{1N} \\
\sigma_{21} & \sigma^2_2 & \cdots & \sigma_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{N1} & \sigma_{N2} & \cdots & \sigma^2_N
\end{pmatrix}
\]  \hspace{1cm} if \hspace{0.5cm} s = k(t-1) + 1

\[
E(\varepsilon_{k(t-1)+1} \varepsilon'_s) = O_N
\]  \hspace{1cm} otherwise.

\[\text{We could focus on any row of (1). Results in Section 3 would not be affected.}\]
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), it is possible to show that the process generated after contemporaneous aggregation is an MA(1) model in the first differences of the data, that is:

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

where \( c = \sum_{i=1}^{N} c_i \) and \( F \) is an \((1 \times N)\) generic aggregation vector.\(^3\) Furthermore, \( a_t \) is the innovation sequence of the contemporaneously aggregated process.

In what follows we assume that \( F = (\omega_1, \omega_2, ..., \omega_N) \), with \( \sum_{i=1}^{N} \omega_i = 1 \).

**Proposition 1** Assuming a weighted aggregation scheme, the innovation variance of the contemporaneously aggregated process \( \sigma^2_a \) and the MA parameter \( \theta \) can be recovered as:

\[
\begin{align*}
\theta &= 1 - \sqrt{1 - 4\rho^2} \\
\sigma^2_a &= \frac{2\rho \gamma(1)}{1 - \sqrt{1 - 4\rho^2}}
\end{align*}
\]

with

\[
\gamma(0) = \text{E}(y_{k(t-1)+1}^2) = \sum_{i=1}^{N} (1 + \alpha_i^2)\sigma_i^2\omega_i^2 + \sum_{i=1}^{N} \sum_{j \neq i} (1 + \alpha_i\alpha_j)\sigma_{ij}\omega_i\omega_j
\]

and

\[
\gamma(1) = \text{E}(y_{k(t-1)+1} y_{k(t-1)}) = \sum_{i=1}^{N} \alpha_i\sigma_i^2\omega_i^2 + \sum_{i=1}^{N} \sum_{j \neq i} \alpha_j\sigma_{ij}\omega_i\omega_j,
\]

where \( \alpha_N = \sum_{i=1}^{N} \theta_{iN} \) and \( \rho = \frac{\gamma(1)}{\gamma(0)} \).

\(^3\)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 \( q^* \leq \text{max}(q_i) \). A detailed proof can be found in Lütkepohl (2007, page 436).


4 Parameters of the temporally aggregated model

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

\[
\begin{pmatrix}
1 & 0 & \ldots & 0 \\
1 & 1 & \ldots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & 1 \\
\end{pmatrix}
\begin{pmatrix}
y_{k(t-1)+1} \\
y_{k(t-1)+2} \\
\vdots \\
y_{k(t-1)+k} \\
\end{pmatrix}
= \begin{pmatrix}
c \\
c \\
\vdots \\
c \\
\end{pmatrix}
\begin{pmatrix}
0 & 0 & \ldots & 1 \\
0 & 0 & \ldots & 0 \\
\vdots \\
0 & 0 & \ldots & 0 \\
\end{pmatrix}
\begin{pmatrix}
y_{k(t-2)+1} \\
y_{k(t-2)+2} \\
\vdots \\
y_{k(t-2)+k} \\
\end{pmatrix}
+ \begin{pmatrix}
1 & 0 & \ldots & 0 \\
\theta & 1 & \ldots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & 1 \\
\end{pmatrix}
\begin{pmatrix}
a_{k(t-1)+1} \\
a_{k(t-1)+2} \\
\vdots \\
a_{k(t-1)+k} \\
\end{pmatrix}
+ \begin{pmatrix}
0 & 0 & \ldots & \theta \\
0 & 0 & \ldots & 0 \\
\vdots \\
0 & 0 & \ldots & 0 \\
\end{pmatrix}
\begin{pmatrix}
a_{k(t-2)+1} \\
a_{k(t-2)+2} \\
\vdots \\
a_{k(t-2)+k} \\
\end{pmatrix}
\]  

(6)

In (6), each row corresponds to an IMA(1,1) process. Without loss of generality, let us focus on the last row:4

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

(7)

where |\(\theta\)| < 1 is the contemporaneously aggregated MA parameter in (5) and \(a_{kt}\) is the weak white noise innovation term of the contemporaneously aggregated process, such that \(a_{kt} \sim (0, \sigma_a^2)\). Note that the random variable \(y_{kt}\) is observed at the disaggregate frequency \(t\).

Following Marcellino (1996), we define the temporally aggregated process as:

\[\{y_{\tau}\}_{\tau=0}^{\infty} = \left\{ \sum_{j=0}^{k-1} y_{k\tau-j} \right\}_{\tau=0}^{\infty} \]  

(8)

Consequently, \(\{y_{\tau}\}_{\tau=0}^{\infty}\) evolves in aggregate time units \(\tau\) according to the aggregation scheme in (8). After temporal aggregation, sample information for the random variable at the

---

4We could focus on any row of (6). Results in Section 4 would not be affected.
aggregate frequency $\tau$ is available only every $k^{th}$ period ($k, 2k, 3k, \ldots$), where $k$, an integer value larger than one, is the aggregation frequency or the order of temporal aggregation. It is interesting to note that the aggregation scheme in (8) corresponds to a linear transformation applied to the macro process in (6), that is, multiplying (6) by an aggregation vector of ones of size $\left(1 \times k\right)$.

Following Weiss (1984), Stram and Wei (1986), Wei (1990) and Marcellino (1999), the Lemma below can be easily proved.

**Lemma 2** The temporal aggregation of the IMA(1,1) process in (6) is an IMA(1,1) in aggregate time units, whatever the aggregation frequency $k$.

As a consequence, the temporal aggregation in (8) corresponds to an IMA(1,1), which may be represented as:

$$
(1 - L^k)y_\tau = C + (1 + \Theta L^k)a_\tau,
$$

where $y_\tau$ and $a_\tau$ are the temporally aggregated time series and innovation sequence, respectively, and $C = kc$. Moreover, $\Theta$ is the MA parameter and $\sigma_a^2$ is the innovation variance of the temporally aggregated process in (9).

At this stage, we need the first two autocovariances of the temporally aggregated process in order to derive its unknown parameters, which can be then recovered as in (5).

**Proposition 2** The parameters of the temporally aggregated process in (9) may be obtained as in (5), considering the following autocovariances:

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

where $\Gamma(0)$ is the first autocovariance and $\Gamma(1)$ is the second autocovariance.

The autocovariances of the temporally aggregated model as in (10) are not trivial to derive, and therefore they represent another contribution of this paper.

**Proof.** The derivation of $\Gamma(0)$ and $\Gamma(1)$ is provided in the Appendix.

5Besides (8), 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.
5 An empirical application to risk management

In this section we illustrate how the results on temporal and contemporaneous aggregation are useful in analysing the problem of volatility prediction and Value-at-Risk (VaR) calculation for a portfolio (contemporaneous aggregation) sampled at several time frequencies (temporal aggregation).

In particular, the aim of the empirical application is to estimate 1-period horizon VaR for a portfolio composed of multiple indices. Portfolio’s VaR represents the $\alpha$-th quantile of the portfolio’s return conditional distribution, namely,

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

where $r_t = \ln(P_t) - \ln(P_{t-1})$ is the continuously compounded return on the portfolio (whose price at time $t$ is $P_t$) over the period $t-1$ to $t$ and $\mathcal{F}_{t-1}$ is the information set at $t-1$.

One key ingredient to calculate the Value-at-Risk (VaR) of a portfolio is the conditional volatility model. In this application we focus on the IGARCH(1,1) specification, which is a restricted version of the standard GARCH(1,1) where the coefficients in the conditional volatility equation sum to one. Estimates in GARCH(1,1) models close to the unit root boundary are very often encountered in empirical analysis: hence, the existence of the IGARCH effect may be considered a reasonable assumption.

As we shall explain in Subsection 5.1, a close analogy exists between the IGARCH(1,1) model for conditional volatility and the IMA(1,1) for squared returns, since these two models share the same autocorrelation function. By giving the IGARCH(1,1) an IMA(1,1) structure, it is possible to apply all the aggregation results so far presented which, being the IMA(1,1) model closed under aggregation, are valid for any temporal aggregation frequency and level of contemporaneous aggregation.

5.1 On the analogy between IGARCH(1,1) and IMA(1,1) models

Let us consider the following IGARCH(1,1) model:

\begin{align*}
  r_t &= \sigma_t z_t \quad z_t \sim i.i.d.(0, 1) \\
  \sigma_t^2 &= c + (1 - \lambda)r_{t-1}^2 + \lambda \sigma_{t-1}^2, \quad (11)
\end{align*}
where $0 < \lambda < 1$ and $c$ is a constant term different from zero. An interesting feature of the IGARCH model is that it is not covariance stationary, although it is strictly stationary (see Nelson, 1990).

The link between the IMA(1,1) model and the IGARCH(1,1) in (11) derives from the “ARMA–in–squares” representation of GARCH models: indeed, equation (11) can be re-parameterized as an IMA(1,1) for the squared returns,

$$(1 - L)r_t^2 = c + (1 + \theta L)a_t,$$

where the moving average coefficient $\theta = -\lambda$ and $a_t = r_t^2 - \sigma_t^2$ is a disturbance term not necessarily with finite variance.\(^6\)

In (12), as already remarked in Harvey, Ruiz and Shephard (1994): “...$(1 - L)r_t^2$ is stationary and has an autocorrelation function like that of an MA(1) process, indicating an analogy with the ARIMA(0,1,1) process.”

Notice that this analogy holds even if the error term $a_t$ in (12) does not possess a finite second moment. This can be easily seen by observing that the autocovariance functions of (12) are

$$
\begin{align*}
\gamma(0) &= (1 + \theta^2)\sigma_a^2 \\
\gamma(1) &= \theta\sigma_a^2
\end{align*}
$$

where $\sigma_a^2$ is the variance of $a_t$. Such that neither the autocorrelation function $\rho = \theta(1+\theta^2)^{-1}$ is affected by the (possibly) explosive process $a_t$, nor the parameter $\theta = \frac{1 - (1-4\rho^2)^{1/2}}{2\rho}$. As a consequence, a candidate estimator of $\theta$ can be based on the IMA(1,1) model.

To see this in practice, we consider two independent strong IGARCH(1,1) processes as in (11):

$$
\begin{align*}
r_{it} &= \sigma_{it}z_{it}, \quad z_t \sim i.i.d. \, N(0,1), \quad (i = 1,2) \\
\sigma_{it}^2 &= c_i + (1 - \lambda)r_{i,t-1}^2 + \lambda\sigma_{i,t-1}^2,
\end{align*}
$$

The simulation study is based on 1000 replications, allowing the time span to be $T = 250$, 500, 1000 and 5000. For each replication, we sum the two independent strong IGARCH(1,1)

\(^6\)The multivariate version of (12) is provided by Zaffaroni (2008, p. 585).
models to obtain an aggregated weak IGARCH(1,1) model whose parameters are $c$ and $\lambda$ (see Nijman and Sentana, 1996). Only weak GARCH models are closed under temporal and contemporaneous aggregation, whereas strong and semi–strong GARCH models are not (Drost and Nijman, 1993; Nijman and Sentana, 1996). As observed by Drost and Nijman (1993), “The weak GARCH definition is quite general and captures the characterizing features of other GARCH formulations.”

The Monte Carlo is conducted for three sets of parameters in (13), in order to obtain three different aggregate parameter values for $\lambda$: 0.85; 0.90; 0.95. Furthermore, the aggregate constant term is $c = 1.00$.

For each data set, we then present the estimates of $\lambda$ derived from the IMA(1,1) representation as in (12). We employ the nonlinear least-squares estimator of the moving average coefficient based on the backcasting procedure developed by Box and Jenkins (1976), which provides unconditional least-squares estimates. The use of nonlinear least-squares estimator has been proposed by Davis (1996) to estimate the parameters of ARMA models whose innovations belong to the domain of attraction of a non-normal stable distribution (and hence have infinite variance). For illustrative purposes, we also consider the standard quasi-maximum likelihood (QML) estimator of the IGARCH(1,1) model (see Lumsdaine, 1996).

The results are presented in Table 1. Both estimation methods deliver good results both for 0.85 and 0.90, though the true $\lambda$ is slightly overestimated.

|TABLE 1 ABOUT HERE|

When $\lambda = 0.95$, representing the most “realistic” value of the IGARCH aggregate parameter when modeling empirical volatility (RiskMetrics™, for instance, suggests 0.94 for daily data and 0.97 for monthly data), the bias is almost null. In addition, it is striking to note that for $T = 250$ the least-squares estimator seems to perform better than the QML estimator (the mean value of the former is 0.98, whereas the mean value of the latter is 1.00).

As a general conclusion from the simulation study, the results confirm that the least-squares estimator based on the IMA(1,1) has a negligible bias in realistic empirical situations and can be thereby considered as a valid candidate in estimating the $\lambda$ parameter as in (11), representing the IGARCH(1,1) portfolio volatility equation.
5.2 Data and preliminary analysis

We examine an equally weighted portfolio composed by NASDAQ and Standard & Poor's 500 (S&P500) stock indices. We employ daily adjusted closing prices data, provided by Datastream, from July 09, 1999, through February 17, 2010 (2769 daily in-sample returns).\textsuperscript{7}

We consider log-returns series of NASDAQ \( r_{NAS,t} \) and S&P500 \( r_{SP,t} \) in order to construct an equally weighted portfolio of both indices.

Figure 1 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 1 ABOUT HERE]

The log-returns of the equally weighted portfolio \( r_t \) may be approximated as

\[
r_t \approx 0.5 r_{NAS,t} + 0.5 r_{SP,t}. \tag{14}
\]

Equal weighting is a common index weighting scheme. We use it in this application to provide an illustration of contemporaneous aggregation.

The aim of this empirical application is to produce 1-period horizon VaR for the portfolio in (14), by employing the IGARCH(1,1) model in (11) for the conditional volatility. We calculate 1-period horizon VaR predictions working with stock price data sampled at five different time frequencies (1/2/3/4/5 business days). In this way we provide an illustration of how the temporal aggregation scheme works at relevant sampling frequencies.

The approach taken here is to calculate the low-frequency portfolio’s volatility from the sample variance of high-frequency portfolio’s returns, as it is usually done when building realized volatility measures.\textsuperscript{8} For instance, using daily returns, it is possible to estimate the

\textsuperscript{7}A similar data set has been recently analysed by Rombouts and Verbeek (2009).

\textsuperscript{8}Realized volatility is very often used to estimate daily volatility, simply summing up intra-daily squared returns at a given sample frequency. When this frequency goes to infinity, realized volatilities converge to the “true” daily volatility, and are therefore consistent and unbiased estimates of the former.
k-day portfolio’s volatility as:
\[ r^2_\tau = \sum_{j=0}^{k-1} r^2_{k-l-j}, \]

neglecting the double sum of \( k - 1 \) autocovariances.

5.3 Estimation issues

Besides volatility, VaR calculation depends on a tail quantile of the portfolio’s log-return conditional distribution. Figure 2 shows the left tail of the histogram of the portfolio’s log-returns in (14). The daily returns have been normalized by dividing for the square root of the volatility in (11).\(^9\) For illustrative purposes, a Student-t distribution with twelve degrees of freedom (dotted line), rescaled to have a unit variance, and a standard Normal (solid line) are displayed.\(^10\) 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, consistent with residual properties, the graph suggests that the Student-t distribution provides a better description of the tails than the Gaussian distribution.

[FIGURE 2 ABOUT HERE]

The IGARCH(1,1) model in (11) can be estimated in various ways. In the application, different procedures are examined.

First, we employ the maximum likelihood estimator (MLE): “Student-t EWMA MLE”. 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 \( \lambda \).\(^11\)

Second, as a benchmark, we also consider a standard Student-t GARCH(1,1) model, whose estimation is also carried out by maximum likelihood.

\(^9\)Note that, to draw the normalized log-returns in Figure 2, the \( \lambda \) parameter has been fixed to 0.94 over the whole sample.

\(^10\)We are aware that different distributional assumptions can be made: yet, for simplicity, we focus only on the Student-t and on the Gaussian.

\(^11\)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.
Third, when the EWMA version of (11) is considered, RiskMetrics\textsuperscript{TM} (1996) suggests to select the decay factor in (11) by searching for the smallest root mean squared forecast error (MSFE) over different values of $\lambda$ (the resulting $\lambda$ is termed “optimal decay factor”).\textsuperscript{12} Formally:

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

(15)

where $[\lambda_{min}, \lambda_{max}]$ is a compact set over which the optimization takes place (i.e., $[0,1]$). Or simply to fix $\lambda$ at a certain value. We refer to “RiskMetrics EWMA” if the decay factor is chosen as in (15), by a simple grid search. We also consider the standard RiskMetrics\textsuperscript{TM} approach, in which the decay factor is fixed at 0.94 (“RiskMetrics EWMA 0.94”). Yet, it has been recently pointed out that the estimation method in (15) is non-consistent for $\lambda$ (Zaffaroni, 2008). In addition, the resulting estimator lacks of the usual asymptotic statistical properties: therefore, the whole estimation procedure is invalid, even in small samples. The bottom line is that the estimation procedure in (15), although widely employed by practitioners, has to be used very cautiously because it can have detrimental effects on the forecasting performance.

Fourth, we term “Aggregation vector IMA(1,1)” our proposed approach to infer $\lambda$. As a matter of fact, equation (12) allows the theoretical framework presented in this paper to be employed to infer the decay factor at several sampling frequencies and levels of contemporaneous aggregation. Recall, indeed, that multivariate GARCH models are characterized by equivalent VARMA representations (see, e.g., Hafner, 2008, and Zaffaroni, 2008): in particular, a multivariate IGARCH(1,1) can be re-written as a vector IMA(1,1) for the vector process that contains the squares and the cross-products of the portfolio components’ returns.

Given that the portfolio in (14) is the weighted sum of two indices, we propose to estimate a trivariate IMA(1,1) model for the system composed by the two squared returns and their cross-product. Once obtained the estimates for the MA matrix of coefficients, the decay factor is analytically inferred (i.e., $\theta$), applying the results on contemporaneous aggregation of the IMA(1,1) model discussed in Section 3. 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., $\Theta$). This procedure has a number of attractions: in particular, if the multivariate model is correctly specified, a multivariate

\textsuperscript{12}We refer to See RiskMetrics\textsuperscript{TM} (Technical Document, 1996, Section 5.3) for further details.
approach should lead to more efficient estimates than those obtained by estimating univariate volatility models. As with the other estimation methods so far discussed, we use the univariate IGARCH(1,1) model in (11) in order to produce 1-day-ahead portfolio’s volatility forecasts.

Regarding vector IMA(1,1) estimation, there are several existing vector autoregressive (VAR)-based procedures for the estimation of vector MA models. We employ the method proposed by Galbraith, Ullah and Zinde-Walsh (2002), which relies on a 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. The maximum VAR lag length is 50. We use the Yule-Walker estimator since it delivers estimates of the VAR matrix coefficients which are always in the stability region.\footnote{We are grateful to Helmut Lütkepohl who recalled us this property.}

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).\footnote{Due to the type of exercise, in selecting the VAR($p$) lag length, we adopt the criterion in (15), as for the RiskMetrics$^\text{TM}$ 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.}

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$^\text{TM}$ approach and of the standard MLE, bearing in mind that all these procedures aim to estimate in different ways the aggregate parameter ($\lambda$) of the univariate IGARCH(1,1) model in (11).

### 5.4 VaR diagnostic test results

In order to assess the quality of the estimation methods and the resulting VaR predictions, 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.\footnote{The developed MATLAB code is available from the authors upon request.} The size of the window is set to 1000 daily observations (approximately 4 years of trading), i.e., to generate VaR predictions, a training period of 1000 observations of past log-returns is used. The use of a fixed rolling window scheme is very common in the financial econometrics literature 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 first part of the sample is used for model estimation. The second part of the sample as out-of-sample validation period. The five examined estimation methods are implemented in order to calculate 1-step-ahead volatility forecasts, based on (11), and 1-period horizon VaR predictions. 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 models are estimated and VaR predictions are calculated using observations from 1 to 1000. Then the sample is rolled forward, the models are re-estimated using data from 2 to 1001, VaR predictions are updated, and so on. 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 and of the decay factor.

When a different time frequency is adopted for VaR prediction, in implementing “Student-t EWMA MLE”, “Student-t GARCH(1,1) MLE”, “RiskMetrics EWMA” and “RiskMetrics EWMA 0.94” the aggregate parameters are estimated using aggregate time series data. For “Aggregation vector IMA(1,1)”, on the other hand, the aggregate parameter is systematically sampled at the corresponding desired frequency, such as the sequence,

\[
\{\lambda_{tk}\}_{t=1}^{1767/k}, \quad k = 1, 2, 3, 4, 5,
\]

(16)

where \( \lfloor b \rfloor \) indicates the integer part of a real number \( b \). Notice that \( k = 1, 2, 3, 4, 5 \) means considering 1/2/3/4/5-days sampling frequencies.

Working with daily data, we have a total of 1767 tests for VaR at each confidence level \((\alpha=0.10, 0.05, 0.025, 0.01)\), corresponding roughly to 7 years of backtesting points. 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 (also known as “violation”, or “failure indicator”), a binary variable which takes the value 1 if the log-returns exceed in absolute value the predicted VaR and 0 otherwise:

\[
I_t(\alpha) = 1\{V_{\alpha R_{t}} | F_{t<\alpha} < r_t\}, \quad t = 1001, 1002, \ldots, 2768.
\]
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 predictions on about 5% of the cases. Formally:

$$\Pr(I_t(\alpha) = 1) = \alpha.$$ \hspace{1cm} (17)

Equation (17) represents the “correct 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); the likelihood ratio test for “independence of VaR violations” (Christoffersen, 1998), which assesses whether or not exceptions are clustered in time, is also performed. Furthermore, “correct unconditional coverage” and “independence of VaR violations” are jointly tested by means of the “correct conditional coverage” test (Christoffersen, 1998), which examines both the number and the timing of violations.

The empirical backtesting results are summarized in Tables 2-6. 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. For a given significance level $\alpha$, from top to down, each cell in the table displays the out-of-sample empirical coverage (i.e., proportion of exceptions) and, for VaR prediction, the p-values of the corresponding unconditional coverage test (Kupiec, 1995), independence of VaR violations test and conditional coverage test (Christoffersen, 1998). A p-value smaller than $\alpha$ implies a rejection of the null hypothesis (bold numbers).

The five methods perform similarly and reasonably well for $T = 1767, 883, 589, 441, 353$ (i.e., 1/2/3/4/5-days sampling frequencies). Overall, evaluation results indicate that the examined estimation methods pass the Kupiec and Christoffersen tests most of the times, with few exceptions: at 1-day frequency ($T = 1767$) and $\alpha = 0.05$, the observed failure rate for “Student-t GARCH(1,1) MLE” is significantly lower than 0.05 (risk is overestimated);
moreover, at 2-days frequency \((T = 883)\) and \(\alpha = 0.10\), the proportion of violations for “Student-t GARCH(1,1) MLE” is 0.0805, far behind 0.10; similarly, at 4-days frequency \((T = 441)\) and \(\alpha = 0.10\), the proportion of violations is 0.0773 for “Student-t GARCH(1,1) MLE”, and therefore the null hypothesis of the unconditional coverage test is not accepted. At 3-days frequency \((T = 589)\) and \(\alpha = 0.01\), the failure rate for “RiskMetrics EWMA” is significantly higher (0.0255) than the theoretical value at the specified confidence level (risk is underestimated); also at 5-days frequency \((T = 353)\) and \(\alpha = 0.01\), both for “RiskMetrics EWMA” and for “Aggregation vector IMA(1,1)”, the observed failure rates are significantly higher (0.0284) than the theoretical values.

In summary, up to 5-days frequency, it comes out that “Student-t GARCH(1,1) MLE” and “RiskMetrics EWMA” perform worse than the other competing techniques, which allow to estimate VaR more accurately. In particular, the “Aggregation vector IMA(1,1)” approach and the standard “RiskMetrics EWMA 0.94” are the most successful techniques in evaluating risk. Therefore, “Aggregation vector IMA(1,1)” provides a rule to infer estimates of the decay factor and volatility predictions up to 5-days forecast horizon which works satisfactorily well.

In general, even when the null hypothesis is not rejected, it is worth noting that “Aggregation vector IMA(1,1)” reports higher (or equal) p-values than “RiskMetrics EWMA” most of the times. This seems suggesting the use of “Aggregation vector IMA(1,1)” when moving from high to low frequencies, at least up to 5-days. The standard “RiskMetrics EWMA 0.94”, in which the decay factor is not estimated but simply fixed, delivers a very good performance (this probably explains why it is still one of the most widely used methodology for measuring market risk in financial industry). These results have to be evaluated with caution since they are based on a simple exercise and on a single data set. Further empirical research is needed to explore the performance of the aggregate vector IMA(1,1) model in VaR analysis.

6 Concluding remarks

In this paper we deal with temporal and contemporaneous aggregation of vector IMA(1,1) models. Relying on the closeness of VARMA processes with respect to linear transformations, we derive algebraic solutions for the unknown parameters of the aggregated model.

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 of the aggregated model 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 optimization procedure in (15) suggested by RiskMetrics™ or on maximum likelihood.
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:

$$
\left( \sum_{j=0}^{k-1} L^j \right)^2 = \sum_{j=0}^{k-1} (j+1) L^j + \sum_{j=k}^{2k-1} (2k-j-1)L^j. \quad (18)
$$

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.,

$$
\left( \sum_{j=0}^{n} L^j \right)^2 = \sum_{j=0}^{n} (j+1)L^j + \sum_{j=n+1}^{2n+1} (2n+1-j)L^j. \quad (19)
$$

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:

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

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

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

As a consequence:

$$
\left( \sum_{j=0}^{n} L^j \right)^2 = \sum_{j=0}^{n-1} (j+1)L^j + \sum_{j=n}^{2n-1} (2n-j-1)L^j + \frac{L^n}{1-L}(2 - L^{n+1} - L^n). \quad (20)
$$

To prove that this is the desired result for $k = n+1$, from (19) and (20), it is enough to show that:

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

(21)
To show that the equation above is true, we develop (21) as:

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

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

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

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

(22)

Therefore (22) confirms (21). This completes Step 2. We have proven Lemma 3. □

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

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

(23)

Bearing in mind the temporal aggregation scheme in (8), 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 (23), 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 above needs to be transformed into:

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

(24)

Let us focus on the first row of (10). The autocovariance of order zero of the temporally aggregated model defined in (24) is:

\[
\Gamma(0) = E \left[ (1 - L^k)y_{\tau} \times (1 - L^k)y_{\tau} \right] = E[(1 + \theta L)(1 + L + \ldots + L^{k-1})^2 a_{kt} \times (1 + \theta L)(1 + L + \ldots + L^{k-1})^2 a_{kt}].
\]

From (18) we know that:

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

(25)
Similarly:

From (26), we already know that:

Then, the autocovariance of order zero is:

Therefore, based on this, we can express \((1 - L^k) y_t = (1 + \theta L)(1 + L + \ldots + L^{k-1})^2 a_{kt}\) in (24) as:

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

The expression here above can be written out as:

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

Thus, the autocovariance of order zero is:

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

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

as we aimed to prove.

Let us switch to the second row of (10). We need to calculate the first-order autocovariance of the temporally aggregated model in (24), that is:

\[
\Gamma(1) = \mathbb{E} \left[ (1 - L^k) y_t \times (1 - L^k) y_{t-1} \right] = \mathbb{E}[(1 + \theta L)(1 + L + \ldots + L^{k-1})^2 a_{kt} \times (1 + \theta L)(1 + L + \ldots + L^{k-1})^2 a_{k(t-1)}].
\]

From (26), we already know that:

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

Similarly:

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

\[
22
\]
Consequently:

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

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

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

The proof is now complete. \( \square \)

**References**


Table 1: Univariate estimation of the IGARCH parameter

<table>
<thead>
<tr>
<th></th>
<th>IMA</th>
<th>IGARCH</th>
<th>IMA</th>
<th>IGARCH</th>
<th>IMA</th>
<th>IGARCH</th>
<th>IMA</th>
<th>IGARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Obs.</td>
<td>250</td>
<td>250</td>
<td>500</td>
<td>500</td>
<td>1000</td>
<td>1000</td>
<td>5000</td>
<td>5000</td>
</tr>
<tr>
<td>True value</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Mean</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>SD</td>
<td>0.07</td>
<td>0.07</td>
<td>0.05</td>
<td>0.03</td>
<td>0.04</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.05</td>
<td>1.05</td>
<td>1.00</td>
<td>0.98</td>
<td>0.95</td>
<td>0.92</td>
<td>0.95</td>
<td>0.88</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.16</td>
<td>0.33</td>
<td>0.61</td>
<td>0.74</td>
<td>0.58</td>
<td>0.78</td>
<td>0.63</td>
<td>0.83</td>
</tr>
<tr>
<td>True value</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>Mean</td>
<td>0.94</td>
<td>0.93</td>
<td>0.92</td>
<td>0.90</td>
<td>0.92</td>
<td>0.90</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>SD</td>
<td>0.05</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.05</td>
<td>1.06</td>
<td>1.02</td>
<td>1.02</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.92</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.56</td>
<td>0.78</td>
<td>0.66</td>
<td>0.84</td>
<td>0.71</td>
<td>0.83</td>
<td>0.71</td>
<td>0.87</td>
</tr>
<tr>
<td>True value</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Mean</td>
<td>0.98</td>
<td>1.00</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>SD</td>
<td>0.04</td>
<td>0.06</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.05</td>
<td>1.06</td>
<td>1.02</td>
<td>1.02</td>
<td>1.00</td>
<td>1.01</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.80</td>
<td>0.15</td>
<td>0.80</td>
<td>0.88</td>
<td>0.86</td>
<td>0.90</td>
<td>0.85</td>
<td>0.93</td>
</tr>
</tbody>
</table>

SD stands for standard deviation of the estimated parameter with respect to the true parameter.
Figure 1: Autocorrelation function of S&P500 and NASDAQ log-returns and squared log-returns.

Figure 2: Histogram (left tail) of the daily normalized log-returns of the portfolio in (14).
<table>
<thead>
<tr>
<th>T=1767 BACKTESTING DATA</th>
<th>RiskMetrics EWMA</th>
<th>RiskMetrics EWMA 0.94</th>
<th>Aggregation</th>
<th>Student-t</th>
<th>Student-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>(sampling frequency: 1 day)</td>
<td>(min MSFE)</td>
<td>(λ set at 0.94)</td>
<td>vector IMA(1,1)</td>
<td>EWMA MLE</td>
<td>GARCH(1,1) MLE</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.1076</td>
<td>0.1053</td>
<td>0.1059</td>
<td>0.1076</td>
<td>0.0909</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.2931</td>
<td>0.4594</td>
<td>0.4134</td>
<td>0.2931</td>
<td>0.1563</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.1602</td>
<td>0.2296</td>
<td>0.2105</td>
<td>0.0939</td>
<td>0.1885</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.2147</td>
<td>0.3096</td>
<td>0.3288</td>
<td>0.2209</td>
<td>0.1542</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0500</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0555</td>
<td>0.0555</td>
<td>0.0555</td>
<td>0.0532</td>
<td>0.0391</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.2976</td>
<td>0.2976</td>
<td>0.2976</td>
<td>0.5378</td>
<td>0.0287</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.8393</td>
<td>0.8393</td>
<td>0.8393</td>
<td>0.9977</td>
<td>0.2209</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.5695</td>
<td>0.5695</td>
<td>0.5695</td>
<td>0.8271</td>
<td>0.0432</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0250</td>
<td>0.0250</td>
<td>0.0250</td>
<td>0.0250</td>
<td>0.0250</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0255</td>
<td>0.0266</td>
<td>0.0255</td>
<td>0.0255</td>
<td>0.0187</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.8972</td>
<td>0.6672</td>
<td>0.8972</td>
<td>0.8972</td>
<td>0.0754</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.1249</td>
<td>0.1088</td>
<td>0.1249</td>
<td>0.1249</td>
<td>0.2021</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.3055</td>
<td>0.3055</td>
<td>0.3055</td>
<td>0.3055</td>
<td>0.1098</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0102</td>
<td>0.0096</td>
<td>0.0091</td>
<td>0.0102</td>
<td>0.0085</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.9354</td>
<td>0.8738</td>
<td>0.6866</td>
<td>0.9354</td>
<td>0.5137</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.5425</td>
<td>0.5425</td>
<td>0.5885</td>
<td>0.5425</td>
<td>0.6121</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.8280</td>
<td>0.8370</td>
<td>0.7963</td>
<td>0.8280</td>
<td>0.7105</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 α implies a rejection of the null hypothesis (bold numbers).
| Confidence level ($\alpha$) | 0.1000 | 0.1000 | 0.1000 | 0.1000 | 0.1000 |
| Proportion of exceptions | 0.0930 | 0.0918 | 0.0964 | 0.0964 | 0.0805 |
| P-value Unconditional coverage test | 0.4818 | 0.4132 | 0.7180 | 0.7180 | **0.0464** |
| P-value Independence test | 0.8840 | 0.3228 | 0.9379 | 0.1637 | 0.9000 |
| P-value Conditional coverage test | 0.7726 | 0.4389 | 0.9340 | 0.3552 | 0.1365 |
| Confidence level ($\alpha$) | 0.0500 | 0.0500 | 0.0500 | 0.0500 | 0.0500 |
| Proportion of exceptions | 0.0544 | 0.0533 | 0.0522 | 0.0601 | 0.0431 |
| P-value Unconditional coverage test | 0.5522 | 0.6574 | 0.7706 | 0.1820 | 0.3349 |
| P-value Independence test | 0.6760 | 0.2579 | 0.2850 | 0.4505 | 0.5700 |
| P-value Conditional coverage test | 0.7680 | 0.4779 | 0.5412 | 0.3087 | 0.5372 |
| Confidence level ($\alpha$) | 0.0250 | 0.0250 | 0.0250 | 0.0250 | 0.0250 |
| Proportion of exceptions | 0.0261 | 0.0204 | 0.0261 | 0.0295 | 0.0227 |
| P-value Unconditional coverage test | 0.8388 | 0.3672 | 0.8388 | 0.4072 | 0.6534 |
| P-value Independence test | 0.2668 | 0.3862 | 0.2668 | 0.7932 | 0.4715 |
| P-value Conditional coverage test | 0.5287 | 0.4575 | 0.5287 | 0.6853 | 0.6977 |
| Confidence level ($\alpha$) | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 |
| Proportion of exceptions | 0.0102 | 0.0113 | 0.0125 | 0.0102 | 0.0068 |
| P-value Unconditional coverage test | 0.9516 | 0.6958 | 0.4775 | 0.9516 | 0.3111 |
| P-value Independence test | 0.6665 | 0.6318 | 0.5979 | 0.6665 | 0.7742 |
| P-value Conditional coverage test | 0.9096 | 0.8259 | 0.6761 | 0.9096 | 0.5746 |

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).
## Table 4: VaR evaluation tests (T=589)

<table>
<thead>
<tr>
<th>(sampling frequency: 3 days)</th>
<th>RiskMetrics EWMA (min MSFE)</th>
<th>RiskMetrics EWMA 0.94 (λ set at 0.94)</th>
<th>Aggregation vector IMA(1,1)</th>
<th>Student-t EWMA MLE</th>
<th>Student-t GARCH(1,1) MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence level (α)</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0935</td>
<td>0.0901</td>
<td>0.1003</td>
<td>0.0986</td>
<td>0.0867</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.5978</td>
<td>0.4182</td>
<td>0.9781</td>
<td>0.9123</td>
<td>0.2737</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.5624</td>
<td>0.5542</td>
<td>0.1466</td>
<td>0.7302</td>
<td>0.4536</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.7357</td>
<td>0.6050</td>
<td>0.3485</td>
<td>0.3906</td>
<td>0.4043</td>
</tr>
</tbody>
</table>

| P-value Unconditional coverage test | 0.0500                      | 0.0500                                | 0.0500                      | 0.0500            | 0.0500                 |
| P-value Independence test       | 0.5037                      | 0.7895                                | 0.9099                      | 0.7640            | 0.3933                 |
| P-value Conditional coverage test | 0.9113                      | 0.1866                                | 0.6300                      | 0.7715            | 0.3939                 |
| Confidence level (α)           | 0.0500                      | 0.0500                                | 0.0500                      | 0.0500            | 0.0500                 |
| Proportion of exceptions       | 0.0561                      | 0.0476                                | 0.0510                      | 0.0527            | 0.0425                 |
| P-value Unconditional coverage test | 0.5037                      | 0.7895                                | 0.9099                      | 0.7640            | 0.3933                 |
| P-value Independence test       | 0.9113                      | 0.1866                                | 0.6300                      | 0.7715            | 0.3939                 |
| P-value Conditional coverage test | 0.7947                      | 0.4035                                | 0.8848                      | 0.9165            | 0.4830                 |

| P-value Unconditional coverage test | 0.0250                      | 0.0250                                | 0.0250                      | 0.0250            | 0.0250                 |
| P-value Independence test       | 0.0391                      | 0.0306                                | 0.0340                      | 0.0340            | 0.0221                 |
| P-value Conditional coverage test | 0.0426                      | 0.3995                                | 0.1841                      | 0.1841            | 0.6470                 |
| Confidence level (α)           | 0.0736                      | 0.5987                                | 0.2044                      | 0.1635            | 0.5071                 |
| Proportion of exceptions       | 0.0255                      | 0.0170                                | 0.0204                      | 0.0221            | 0.0136                 |
| P-value Unconditional coverage test | 0.0016                      | 0.1206                                | 0.0262                      | 0.0109            | 0.4048                 |
| P-value Independence test       | 0.3903                      | 0.1546                                | 0.4791                      | 0.2839            | 0.0908                 |
| P-value Conditional coverage test | 0.0047                      | 0.1088                                | 0.0657                      | 0.0221            | 0.1691                 |

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).
Table 5: VaR evaluation tests (T=441)

<table>
<thead>
<tr>
<th>RiskMetrics EWMA (min MSFE)</th>
<th>RiskMetrics EWMA 0.94 (λ set at 0.94)</th>
<th>Aggregation vector IMA(1,1)</th>
<th>Student-t EWMA MLE</th>
<th>Student-t GARCH(1,1) MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence level (α)</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0864</td>
<td>0.0864</td>
<td>0.0886</td>
<td>0.0977</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.3301</td>
<td>0.3301</td>
<td>0.4187</td>
<td>0.8733</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.1329</td>
<td>0.3308</td>
<td>0.7800</td>
<td>0.3576</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.2024</td>
<td>0.3879</td>
<td>0.6935</td>
<td>0.6407</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0500</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0568</td>
<td>0.0500</td>
<td>0.0591</td>
<td>0.0614</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.5203</td>
<td>1.0000</td>
<td>0.3944</td>
<td>0.2901</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.6270</td>
<td>0.4139</td>
<td>0.6237</td>
<td>0.7849</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.7227</td>
<td>0.7162</td>
<td>0.6169</td>
<td>0.5506</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0250</td>
<td>0.0250</td>
<td>0.0250</td>
<td>0.0250</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0295</td>
<td>0.0250</td>
<td>0.0273</td>
<td>0.0295</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.5526</td>
<td>1.0000</td>
<td>0.7634</td>
<td>0.5526</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.3893</td>
<td>0.2674</td>
<td>0.4115</td>
<td>0.3983</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.5788</td>
<td>0.5406</td>
<td>0.6821</td>
<td>0.5788</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0136</td>
<td>0.0091</td>
<td>0.0159</td>
<td>0.0136</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.4676</td>
<td>0.8457</td>
<td>0.2513</td>
<td>0.4676</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.6834</td>
<td>0.7862</td>
<td>0.6339</td>
<td>0.6834</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.7068</td>
<td>0.9458</td>
<td>0.4624</td>
<td>0.7068</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 α implies a rejection of the null hypothesis (bold numbers).
<table>
<thead>
<tr>
<th>T=353 BACKTESTING DATA</th>
<th>RiskMetrics EWMA (min MSFE)</th>
<th>RiskMetrics EWMA 0.94 (λ set at 0.94)</th>
<th>Aggregation vector IMA(1,1)</th>
<th>Student-t EWMA MLE</th>
<th>Student-t GARCH(1,1) MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence level (α)</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.1000</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0767</td>
<td>0.0824</td>
<td>0.0881</td>
<td>0.0824</td>
<td>0.0739</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.1303</td>
<td>0.2573</td>
<td>0.4472</td>
<td>0.2573</td>
<td>0.0880</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.9536</td>
<td>0.6803</td>
<td>0.8638</td>
<td>0.2947</td>
<td>0.9543</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.3478</td>
<td>0.4837</td>
<td>0.7381</td>
<td>0.3041</td>
<td>0.2329</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0500</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0568</td>
<td>0.0483</td>
<td>0.0597</td>
<td>0.0625</td>
<td>0.0483</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.5653</td>
<td>0.8827</td>
<td>0.4191</td>
<td>0.2995</td>
<td>0.8827</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.8876</td>
<td>0.8426</td>
<td>0.5116</td>
<td>0.7189</td>
<td>0.2407</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.8392</td>
<td>0.9699</td>
<td>0.5817</td>
<td>0.5472</td>
<td>0.4970</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0250</td>
<td>0.0250</td>
<td>0.0250</td>
<td>0.0250</td>
<td>0.0250</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0341</td>
<td>0.0369</td>
<td>0.0455</td>
<td>0.0369</td>
<td>0.0199</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.3001</td>
<td>0.1801</td>
<td>0.0271</td>
<td>0.1801</td>
<td>0.5242</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.4149</td>
<td>0.4934</td>
<td>0.2163</td>
<td>0.4934</td>
<td>0.1179</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.4193</td>
<td>0.3221</td>
<td>0.0405</td>
<td>0.3221</td>
<td>0.2405</td>
</tr>
<tr>
<td>Confidence level (α)</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
<td>0.0100</td>
</tr>
<tr>
<td>Proportion of exceptions</td>
<td>0.0284</td>
<td>0.0256</td>
<td>0.0284</td>
<td>0.0227</td>
<td>0.0170</td>
</tr>
<tr>
<td>P-value Unconditional coverage test</td>
<td>0.0046</td>
<td>0.0141</td>
<td>0.0046</td>
<td>0.0396</td>
<td>0.2274</td>
</tr>
<tr>
<td>P-value Independence test</td>
<td>0.2752</td>
<td>0.2154</td>
<td>0.4437</td>
<td>0.1628</td>
<td>0.0807</td>
</tr>
<tr>
<td>P-value Conditional coverage test</td>
<td>0.0099</td>
<td>0.0228</td>
<td>0.0134</td>
<td>0.0455</td>
<td>0.1050</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 α implies a rejection of the null hypothesis (bold numbers).