

School of Finance



University of St.Gallen

**FORECASTING REALIZED (CO)VARIANCES WITH A BLOC
STRUCTURE WISHART AUTOREGRESSIVE MODEL**

MATTEO BONATO

MASSIMILIANO CAPORIN

ANGELO RANALDO

WORKING PAPERS ON FINANCE NO. 2012/11

SWISS INSTITUTE OF BANKING AND FINANCE (S/BF – HSG)

NOVEMBER 2012



Forecasting realized (co)variances with a block structure Wishart autoregressive model

Matteo Bonato* Massimiliano Caporin† Angelo Rinaldo‡

Current Draft: November 2008

Abstract

The increased availability of high-frequency data provides new tools for forecasting of variances and covariances between assets. However, recent realized (co)variance models may suffer from a 'curse of dimensionality' problem similar to that of multivariate GARCH specifications. As a result, they need strong parameter restrictions, in order to avoid non-interpretability of model coefficients, as in the matrix and log exponential representations. Among the proposed models, the Wishart autoregressive model introduced by Gouriéroux et al. (2007) analyzes the realized covariance matrices without any restriction on the parameters while maintaining coefficient interpretability. Indeed, the model, under mild stationarity conditions, provides positive definite forecasts for the realized covariance matrices. Unfortunately, it is still not feasible for large asset cross-section dimensions. In this paper we propose a restricted parametrization of the Wishart Autoregressive model which is feasible even with a large cross-section of assets. In particular, we assume that the asset variances-covariances have no or limited spillover and that their dynamic is sector-specific. In addition, we propose a Wishart-based generalization of the HAR model of Corsi (2004). We present an empirical application based on variance forecasting and risk evaluation of a portfolio of two US treasury bills and two exchange rates. We compare our restricted specifications with the traditional WAR parameterizations. Our results show that the restrictions may be supported by the data and that the risk evaluations of the models are extremely close. This confirms that our model can be safely used in a large cross-sectional dimension given that it provides results similar to fully parameterized specifications.

*Matteo Bonato is a PhD Candidate in Finance at the Swiss Banking Institute of the University of Zürich and Swiss Finance Institute, Zürich, Switzerland (E-mail: bonato@isb.uzh.ch).

†Massimiliano Caporin is Assistant Professor of Economics at the Department of Economics "Marco Fanno", University of Padua, Padua, Italy (E-mail: massimiliano.caporin@unipd.it).

‡Angelo Rinaldo is senior economist at the Swiss National Bank, Zürich, Switzerland (E-mail: Angelo.Rinaldo@snb.ch). We are grateful for comments received at the International Workshop on Computational and Financial Econometrics in Neuchatel (June 2008) and the Swiss National Bank Brown Bag Seminar (July 2008). We also thank Lorian Mancini, Tim Bollerslev, Charlotte Christiansen, Valeri Voev and Fulvio Corsi for their useful feedback and suggestions.

-The views expressed in this paper are those of the author(s) and do not necessarily represent those of the Swiss National Bank-

1 Introduction

The increased availability of high-frequency data provides new tools for forecasting variances and covariances between assets. In particular, after the seminal paper by Andersen and Bollerslev (1998), the literature on realized volatility has grown enormously; see McAleer and Medeiros (2006) for a review.

While most works has focused on the study of univariate series, recently there has been growing theoretical and empirical interest in extending the results for the univariate process to a multivariate framework. In this context, two pioneering contributions have been made by Barndorff-Nielsen and Shephard (2004) and Bandi and Russel (2005). Barndorff-Nielsen and Shephard (2004) did not consider the presence of microstructure noise, whereas of the noise has been considered in Bandi and Russel (2005).

Alternative approaches to the high-frequency covariance estimator have recently been introduced by Hayashi and Yoshida (2005, 2006), Sheppard (2006) and Zhang (2006), among others. For example, instead of using calendar returns, the Hayashi and Yoshida estimator (HY) is based on overlapping tick-by-tick returns. Sheppard (2006) analyzed the conditions under which the realized covariance is an unbiased and consistent estimator of the integrated covariance. Zhang (2006) also studied the effects of microstructure noise and non-synchronous trading in the estimation of integrated covariance between assets.

Although the literature on multivariate extensions of the realized variance regarding the definition of new estimators of the realized covariances resulted in a notable amount of academic works, only a few papers provide financial applications for these new estimators.

One explanation for the scarcity of empirical contributions in multivariate realized volatility analysis is the difficulty in finding a dynamic specification of a stochastic volatility matrix which satisfies the symmetry and positivity properties of each forecasted matrix, does not suffer from the so called ‘curse of dimensionality’ and possesses a closed-form expression for the forecasts at any horizon.

In an interesting paper, de Pooter et al. (2006) investigate the benefits of high-frequency intraday data when constructing mean-variance efficient stock portfolios with daily rebalancing from the individual constituents of the S&P 100 index. The author analyzed the issue of determining the optimal sampling frequency, as judged by the performances of the estimated portfolios. As in Fleming et al. (2001, 2003), and building on the work of Foster and Nelson (1996) and Andreou and Ghysels (2002), in this paper a rolling window volatility estimator is used to forecast the conditional variance matrix $V_{t,h}$:

$$\widehat{V}_{t,h} = \exp(-\alpha_h)\widehat{V}_{t-1,h} + \alpha_h \exp(-\alpha)Y_{t-1} \quad (1)$$

where α_h can be estimated by means of maximum likelihood for the model

$$r_t = \widehat{V}_{t,h}^{1/2} z_t \quad (2)$$

with $z_t \stackrel{i.i.d.}{\sim} N(0, I)$ and Y_t as the realized covariance matrix estimated using I intraday returns of equal length $h \equiv 1/I$. r_t is the usual $n \times 1$ vector of daily returns at time t of the n assets composing the portfolio.

In a related paper, Bandi et al. (2006) evaluate the economic benefits of methods that have been suggested to optimally sample (in a MSE sense) high-frequency returns data for the purpose of realized variance and covariance estimation in the presence of market microstructure noise. However, their approach is different from that in de Pooter et al. (2006); their method is designed to select the time-varying optimal sampling frequency for each entry in the covariance matrix based on MSE criteria. Subsequently, the economic gains yielded by the MSE-based optimal sampling are evaluated by comparing the utility gains

provided by optimally sampled realized covariance with realized covariances based on fixed intervals. To forecast each entry of the covariance matrix, they adopted an ARFIMA(2, d , 2) model.

An alternative way to forecast the realized variance/covariance matrix is to adopt a matrix transformation that guarantees the positive definitiveness of the forecasts.

Bauer and Vorkink (2007) present a new matrix logarithm model of realized covariance stock returns which uses latent factors as functions of both lagged volatility and returns. The model has several advantages in that it is parsimonious, does not impose parametric restrictions, and yields positive definite covariance matrices.

In Chiriac and Voev (2008) a model based on a multivariate, fractionally integrated autoregressive moving average (ARMA) process for the elements of the Cholesky factors of the observed matrix series is proposed. Denoting with Y_t the $n \times n$ realized covariance matrix at time t , with n the number of assets considered, the Cholesky decomposition of Y_t is given by the upper triangular matrix P_t , for which $P_t P_t' = Y_t$. Then the following model is used

$$\Phi(L)D(L)(X_t - \mu) = \Theta(L)\epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_t). \quad (3)$$

$X_t = \text{vech}(P_t)$ is the vector obtained by stacking the upper triangular components of the matrix P_t in a vector, $\Phi(L)$ and $\Theta(L)$ are matrix lag polynomials and $D(L) = \text{diag}[(1 - L)^{d_1}, \dots, (1 - L)^{d_m}]$, where d_1, \dots, d_m are the degrees of fractional integration of each of the m elements of the vector X_t . μ is a vector of constants. Parameters in (3) are not directly interpretable. However, the dynamic linkages among the variances and covariances series as functions of those parameters are derived.

While both the matrix logarithmic transformation and the Cholesky decomposition have the advantage of guaranteeing the positive definiteness of the covariance matrix, they also have a major drawback: the coefficients of the model totally rule out any possible interpretation. In other words, there is no way to check the significance of the interactions between variances and covariances and thus to reduce the number of parameters in the model by imposing no or limited spillover between the variances and covariances.

A solution to this problem is represented by the Wishart autoregressive model (WAR) proposed by Gouriéroux et al. (2007). The model is based on a dynamic extension of the Wishart distribution. This specification is compatible with financial theory, satisfies the constraints on volatility matrices, has a flexible form and, most importantly, maintains the coefficients' interpretability.

The main innovation proposed in this paper is the introduction of a specific parametrization of the WAR model. In particular, we show how to achieve a great reduction of the number of parameters according to an economic criterion which is consistent with standard sectorial asset allocation approaches. The parametric structure we propose imposes a block structure on the coefficient matrices, hence we name the model *block WAR*. The use of block structures in parameter matrices is similar to that in Billio et al. (2006), Billio and Caporin (2008), Asai et al. (2008). Engle and Kelly (2008) introduce a block structure for the correlation matrix while Caporin and Paruolo (2008) present a spatial solutions to the curse of dimensionality problem in multivariate volatility models that implies a block structure on the coefficient matrices. In this paper we assume that the asset variances-covariances have no or limited spillover and that their dynamic is sector-specific. A pairwise preliminary analysis confirms this assumption and allows us to substantially reduce the number of parameters implied by the model. In addition, we propose a Wishart-based generalization of the HAR model of Corsi (2004), named HAR-WAR model. We present an empirical application based on variance forecasting and risk evaluation of a portfolio of two US treasury bills and two exchange rates. We compare our restricted specifications with the traditional WAR parameterizations.

Our results show that the restrictions may be supported by the data and that the risk evaluations of the models are extremely close. This confirms that our model can be safely used in a large cross-sectional dimension given that it provides results similar to fully parameterized specifications.

Section 2 introduces the WAR model of [Gourieroux et al. \(2007\)](#), followed by our proposed generalization. Section 3 presents the estimation procedure and show an alternative way to estimate the degrees of freedom of the model, a key element to determine if the density of the Wishart distribution exists. The dataset we used is presented in Section 4 and an empirical application based on portfolio risk evaluation is provided in Section 5. Section 6 concludes and gives directions for future research.

2 The block Wishart autoregressive model

In the following we define the basic Wishart auto regressive model of [Gourieroux et al. \(2007\)](#) and then we introduce the set alternative parametric restrictions that define the block WAR.

2.1 The Wishart Autoregressive Model

Denote by Y_t the time t (realized) covariance for a group of n assets. The sequence of stochastic positive definite Y_t matrices is said to follow a Wishart process if the following relations hold.

At first, the (realized) covariance may be represented as a sum of underlying stochastic processes

$$Y_t = \sum_{k=1}^K x_{k,t} x_{k,t}' \quad (4)$$

where $x_{k,t}$, $k = 1, 2, \dots, K$ are independent Gaussian VAR(1) processes of dimension n with a common autoregressive parameter matrix M and common innovation variance Σ :

$$x_{k,t} = M x_{k,t-1} + \epsilon_{k,t}, \quad \epsilon_{k,t} \sim N(0, \Sigma). \quad (5)$$

When Y_t is defined as in (4) and (5) we say it follows a WAR process of order 1, denoted $W[K, M, \Sigma]$. The transition density of WAR(1) depends on the following parameters: K , the scalar degree of freedom (the number of underlying VAR processes), strictly greater than $n-1$ (the number of assets minus one); M , the $n \times n$ matrix of autoregressive parameters; and Σ , the $n \times n$ symmetric and positive definite matrix of innovation covariances. We stress that the interpretation of Y_t from latent Gaussian VAR(1) processes is valid for integer valued K only.

From Proposition 2 in [Gourieroux et al. \(2007\)](#) we have:

$$E_t(Y_{t+1}) = M Y_t M' + K \Sigma. \quad (6)$$

The first conditional moment is thus an affine function of the lagged values of the volatility process. In particular, the WAR(1) process is a weak linear AR(1) process. More precisely we get:

$$Y_{t+1} = M Y_t M' + K \Sigma + \eta_{t+1}, \quad (7)$$

where η_{t+1} is a matrix of stochastic errors with a zero conditional mean. Equivalently, we may represent Y_t conditional mean in the following companion form:

$$vech(Y_{t+1}) = A(M) vech(Y_t) + vech(K \Sigma) + vech(\eta_{t+1}), \quad (8)$$

where $vech(Y)$ denotes the vector obtained by stacking the lower triangular elements of Y , and $A(M)$ is a function of M . The error term η is a weak white noise, since it features conditional heteroskedasticity and, even after conditional standardization, is not identically distributed.

In general, WAR processes with higher autoregressive order p may be considered and the Wishart process can be easily extended to include more autoregressive lags. This is accomplished by replacing the conditioning matrix MY_tM' with any symmetric positive semi-definite function of $Y_t, Y_{t-1}, \dots, Y_{t-p+1}$. However, when the autoregressive order is larger than 1, the interpretation of the Wishart process as the sum of squares of autoregressive Gaussian processes is no longer valid even for integer K . For a WAR(p) process, the equivalent of (6) reads:

$$E_t(Y_{t+1}) = \sum_{j=1}^p M_j Y_{t+1-j} M_j' + K\Sigma. \quad (9)$$

In the following, unless differently stated, we will refer only to WAR(1) specifications.

2.2 Interpretation of the coefficients

The principal drawback of many multivariate volatility models is the so-called ‘curse of dimensionality’, that is, the numbers of parameters is a power function of the cross-sectional model dimension. One of the main contributions of this paper is to provide a sensible reduction of the parameter space by imposing a set of restrictions on the standard WAR model. Our modeling approach will be presented in the following section; here we provide the intuition on parameter interpretation within the WAR model.

In the simple case of a (2×2) matrix, as done in [Gourieroux \(2007\)](#), we define the best prediction of Y_t given by a WAR(1) model. Then we present the approaches we suggest to reduce the parameter space.

Consider the (2×2) covariance matrix Y_t , the autoregressive matrix M and the innovation variance Σ :

$$Y_t = \begin{pmatrix} Y_{11,t} & Y_{12,t} \\ Y_{12,t} & Y_{22,t} \end{pmatrix}, M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

The full WAR(1) model specifies the best prediction of Y_t as:

$$E[Y_t|Y_{t-1}] = \begin{pmatrix} a_1 Y_{11,t-1} + b_1 Y_{12,t-1} + c_1 Y_{22,t-1} + d_1 & a_2 Y_{11,t-1} + b_2 Y_{12,t-1} + c_2 Y_{22,t-1} + d_2 \\ - & a_3 Y_{11,t-1} + b_3 Y_{12,t-1} + c_3 Y_{22,t-1} + d_3 \end{pmatrix} \quad (10)$$

where a_j, b_j, c_j and d_j , $j = 1, \dots, 3$ are scalar parameters. d_j corresponds to K times the entries of Σ . By construction, the prediction is a symmetric semi-definite positive matrix for any Y_{t-1} which belong to \mathcal{S}^+ , the set of symmetric positive definite matrices. To express it in terms of M we have:

$$\begin{cases} a_1 = m_{11}^2, & b_1 = 2m_{11}m_{12}, & c_1 = m_{12}^2, \\ a_2 = m_{11}m_{21}, & b_2 = m_{11}m_{22} + m_{21}m_{21}, & c_2 = m_{12}m_{22}, \\ a_3 = m_{21}^2, & b_3 = 2m_{21}m_{22}, & c_3 = m_{22}^2, \end{cases}$$

The effect of the past variances and covariances on the present volatility can be seen immediately. First, note that the full WAR model allows for spillover between variances and covariances.

Therefore, a possible strategy is to reduce the numbers of parameters by assuming no or limited spillover between the variances. For instance, setting $m_{12} = 0$ implies that the conditional variance of the first asset depends only on its past shocks and that the second asset variance does not influence the

conditional covariance. Differently, a diagonal specification of M corresponds to the absence of spillovers between variances and covariances.

This very simple example in two dimensions helps us to identify the coefficients in M that plays a role in the spillover effect between variances. Using the delta method we can, in fact, easily compute the standard errors for the a_i , b_i and c_i and thus evaluate which parameters are significant and check the appropriateness of assumption of limited spillover. We will present now four different parametrizations for the WAR process that impose no or limited spillover. We also show in the empirical analysis that the restrictions we impose on the matrix M are justified by the data.

2.3 Specifications of the block Wishart autoregressive model

To derive the block WAR model we impose a set of restrictions on the matrix M . These restrictions come from a criterion allowing assets to be grouped. Some examples are given by the economic sector of the stocks entering into an equity portfolio, the type of assets entering into a diversified equity-bond portfolio, or the geographical reference areas of a group of assets. The main intuition behind asset grouping is that the clustered variables may share common patterns or common features, and that their variance-covariance dynamic is similar. In fact, we can presume that assets belonging to the same economic sector may have a similar reaction to market shocks/news, and are similarly affected by market movements.

Clearly, groups may be defined on a data-driven basis, such as referring to the dynamic properties of the series mean and/or variances, or on mixed criteria. The comparison of alternative methods for clustering financial assets is outside the scope of this paper and will not be considered. In the following we will use *a priori* defined groups in order to present our modeling approach and to show, on an empirical basis, its advantages.

Consider the simple WAR(1) model as in Eq. 7:

$$Y_{t+1} = MY_tM' + K\Sigma + \eta_{t+1}.$$

Assume that our portfolio consists of n stocks and that we can classify them into N groups, according to some economic (or data-driven) criterion, as discussed in the previous section (such as the economic sector or the existence of common patterns in realized variances and covariances).

The N groups have dimension n_i with $\sum_i n_i = n$. In addition, the assets are ordered following a group rule, that is, assets from 1 to n_1 belong to group 1, assets from $n_1 + 1$ to $n_1 + n_2$ belongs to group 2, and so on. Given this asset classification, the autoregressive matrix M may be partitioned as follows:

$$M = \begin{pmatrix} M_{11} & \cdots & M_{1N} \\ \vdots & M_{ii} & \vdots \\ M_{N1} & \cdots & M_{NN} \end{pmatrix},$$

where M_{ij} is a matrix of dimension $n_i \times n_j$.

By imposing a particular structure on the matrices M_{ij} we be able to reduce the number of parameters of the model. We propose the following specifications:

- (i) $M_{ij} = \mathbf{0} \quad \forall i \neq j, i, j = 1, \dots, N,$
- (ii) $M_{ij} = \mathbf{0}$ and $M_{ii} = \alpha_i(\mathbf{i}_{n_i}\mathbf{i}'_{n_i}), \quad \forall i \neq j, i, j = 1, \dots, N$
- (iii) $M_{ij} = \mathbf{0}$ and $M_{ii} = (\alpha_{i,1}, \dots, \alpha^{i,n_i})(\mathbf{I}_{n_i}), \quad \forall i \neq j, i, j = 1, \dots, N$

(iv) $M_{ij} = \mathbf{0}$ and $M_{ii} = \alpha_i(\mathbf{I}_{n_i})$, $\forall i \neq j, i, j = 1, \dots, N$

where \mathbf{i}_{n_i} is a $n_i \times 1$ vector of ones and \mathbf{I}_{n_i} is the identity matrix of dimension n_i .

If assets belonging to the same group share common reactions to shocks, we can hypothesize, to some extent, that their co-volatilities also have a similar behavior. If the groups are sector-specific, model (i) implies that the variances and covariances of each asset are only influenced by the variances and covariances of assets belonging to the same class. Therefore, no volatility spillover exists between assets belonging to different sectors. We named this model *block WAR*. The number of parameters that needs to be estimated is $n(n+1)/2 + \sum_{i=1}^N n_i^2$, along with the degrees of freedom K .

A further reduction of the number of parameters is obtained by imposing a single parameter for each group, as shown in model (ii). In this case, the variance and covariance of each asset belonging to, say, group j depends on the past values of itself, on the past values of the variances of the other assets of the same group and on the covariances with those assets via a function of the unique parameter α_j . We call this model *restricted block WAR*. This models contains $n(n+1)/2 + N$ parameters in M and Σ plus K .

Model (iii) relaxes the assumption of spillover between assets belonging to the same sector. It assumes each matrix M_{ii} , $i = 1, \dots, N$ to be diagonal, i.e. the autoregressive matrix M is diagonal. In this case grouping the assets according to some criterion does not affect the parametric space. We named this model *diagonal WAR*. For this model, n parameters need to be estimated in the matrix M , plus the $n(n+1)/2$ parameters in Σ and the degrees of freedom K . One of the implications of the diagonal structure for M is that each realized variance is only a function of its past values.

If we assume again that assets belonging to the same sector have common dynamics for the variance, or if we can find a way to group assets whose volatilities obeys the same process, the number of parameters can be further reduced. This is the case for model (iv). For each group a single parameter is taken to model the dynamics of the variances for the assets in the considered group, i.e. the elements on the diagonal of each M_{ii} , $i = 1, \dots, N$ are all equal. In total only $N + n(n+1)/2 + 1$ parameters are required in this model. We refer to this model as the *restricted diagonal WAR*.

Is is worth mentioning that the specifications (i)-(iv) do not represent a complete generalization of the WAR model. In fact, we set all the off-diagonal blocks to zero. The assumption $M_{ij} = \mathbf{0} \forall i \neq j, i, j = 1, \dots, N$ can be replaced by the same structure we imposed on the matrices M_{ii} : full, scalar, diagonal and restricted diagonal. This allows us to consider not only the interactions between assets belonging to the same group, but also interactions between a limited set of groups. In this paper we stick with a structure that ignores the off-diagonal blocks and leave a full generalization of the WAR model for future works.

2.4 The block HAR-WAR model

One of the stylized facts about asset returns is the long-run temporal dependencies of return volatilities. The literature on volatility modeling has documented that such temporal dependencies are highly persistent. In particular, the low first-order autocorrelations usually found in empirical analysis (Thomakos and Wang, 2003), along with their slow decay, suggest that the logarithmic realized standard deviations do not contain a unit root but exhibit long memory.

To account for this, fractionally integrated autoregressive models (ARFIMA) have been shown to be effective in empirical modeling (see Andersen et al. (2001a) and Andersen et al. (2001b) among others). Fractional integration achieves long memory parsimoniously by imposing a set of infinite dimensional restrictions on the infinite variable lags but completely lacks a clear mathematical interpretation.

Another crucial point is that the long memory observed in the data could be only an apparent behavior generated from a process which is not really long memory. Indeed, the usual tests can indicate the presence of long memory simply because the largest aggregation level that we are able to consider is not large enough. LeBaron (2001) shows that a very simple additive model defined, as the sum of only three different linear processes (AR(1) processes) each operating on a different time frame, can display hyperbolic decaying memory, provided that the longest component has a half-life that is long relative to the test aggregation ranges. Another result from Granger (1980) shows that the sums of a high number of short memory processes can induce long memory. In Pong et al. (2004) an ARMA(2,1) is proposed to model and forecast realized volatility. The authors' choice is motivated by the research of Gallant et al. (1999), who show that the sum of two AR(1) processes is capable of capturing the persistent nature of asset price volatility. In their paper Pong et al. (2004) show that the short memory ARMA(2,1) model is as good as long memory ARFIMA models when forecasting futures volatilities. Motivated by the existence of multiple volatility components in intraday frequencies, along with the apparent long-memory characteristic, Andersen and Bollerslev (1997) formulated a version of the mixture-of-distributions hypothesis (MDH) for returns that explicitly accommodates numerous heterogeneous information arrival processes.

An alternative to ARFIMA is the heterogeneous autoregressive (HAR) model suggested by Corsi (2004) (see also Aït-Sahalia and Mancini, 2008; Corsi et al., 2007). Extending the heterogeneous ARCH model of Müller et al. (1997), the long-memory pattern is reproduced by summing of (a small number of) volatility components constructed over different horizons. The basic ideas stems from the so called 'heterogeneous market hypothesis' presented by Müller et al. (1993), which recognized the presence of heterogeneity in traders. Differently from Andersen and Bollerslev (1997), in this latter view the multi-component structure in the volatility is to be found in the heterogeneity of agents rather than in the heterogeneous nature of the information arrival.

Defining the k -period realized volatility component by the sum of the single-period realized volatilities, i.e.

$$\left(\sqrt{RV}\right)_{t-k:t-1} = \frac{1}{k} \sum_{j=1}^k \sqrt{RV_{t-j}}, \quad (11)$$

the HAR model for realized volatility of Corsi (2004), including the daily, weekly and monthly realized volatility components, is given by

$$\sqrt{RV}_t = \alpha_0 + \alpha_d + \sqrt{RV}_{t-1} + \alpha_w \left(\sqrt{RV}\right)_{t-5:t-1} + \alpha_m \left(\sqrt{RV}\right)_{t-22:t-1} + \mu_t. \quad (12)$$

In Corsi (2004) μ_t is assumed to be Gaussian white noise, whereas in Corsi et al. (2007), a standardized normal inverse Gaussian (NIG) is chosen to deal with the non-Gaussianity of the error terms.

In the spirit of the HAR model, we propose here to model the conditional realized covariance matrix Y_t with an autoregressive Wishart process which accounts for the temporal aggregation of the covariance matrix. We call this process WAR-HAR process. In the sequel, we will show that this process, can be interpreted as a particular WAR(23) process.

Define the k -period realized covariance matrix component by the sum of the single-period realized covariance matrices:

$$Y_{t-k:t-1} = \frac{1}{k} \sum_{j=1}^K Y_{t-j} \quad (13)$$

Combining a WAR(p) structure with the temporal aggregation induced by the HAR model, we write the

process Y_t as:

$$Y_t = M_1 Y_{t-1} M_1' + M_2 Y_{t-5:t-1} M_2' + M_3 Y_{t-22:t-1} M_3' + K\Sigma + \eta_t, \quad (14)$$

Now, opening the summations and aggregating according to the same lag, we get:

$$Y_t = (M_1 Y_{t-1} M_1') + (\tilde{M}_2 Y_{t-1} \tilde{M}_2' + \tilde{M}_3 Y_{t-1} \tilde{M}_3') + \dots + \quad (15)$$

$$\left(\tilde{M}_2 Y_{t-5} \tilde{M}_2' + \tilde{M}_3 Y_{t-5} \tilde{M}_3' \right) + \tilde{M}_3 Y_{t-6} \tilde{M}_3' + \dots + \tilde{M}_3 Y_{t-22} \tilde{M}_3' + \quad (16)$$

$$K\Sigma + \eta_t, \quad (17)$$

with $\tilde{M}_2 = \frac{1}{\sqrt{5}} M_2$ and $\tilde{M}_3 = \frac{1}{\sqrt{22}} M_3$.

To interpret the process as a WAR(22), we simply rewrite it as:

$$Y_t = M_1 Y_{t-1} M_1' + \sum_{i=1}^5 N_2 Y_{t-i} N_2' + \sum_{j=6}^{23} \tilde{M}_3 Y_{t-j} \tilde{M}_3' + K\Sigma + \eta_t. \quad (18)$$

where

$$N_2 : N_2 Y_t N_2' = \tilde{M}_2 Y_t \tilde{M}_2' + \tilde{M}_3 Y_t \tilde{M}_3'.$$

As for the WAR(p) process, the WAR-HAR process permits a *vech* representation, i.e.

$$vech(Y_t) = \sum_{j=1}^{22} A_j(M_1, M_2, M_3) vech(Y_{t-j}) + vech(K\Sigma) + vech(\eta_t) \quad (19)$$

where $A_j(M_1 N M_2, \tilde{M}_3)$ is a matrix function of M_1, N_2 and \tilde{M}_3 .

Since the HAR-WAR model is a WAR(22) characterized using only three autoregressive matrices, the reduction of the parametric space introduced in Section 2.3 is applied in this new context to matrices M_1, M_2 and M_3 . This originates what we called the *full HAR-WAR*, the *diagonal HAR-WAR*, the *restricted diagonal HAR-WAR*, the *block HAR-WAR* and the *restricted block HAR-WAR*.

3 Estimation

3.1 Identification

Following the exposition in Gourieroux et al. (2007), we obtain an analogous identification result for the block WAR and block WAR-HAR model. For ease of exposition we present only the estimation procedure for the WAR(1) process with diagonal autoregressive matrix M . The assumption of diagonal M , even if strict, renders the estimation extremely easy and fast. The extension to the diagonal HAR-WAR case is straightforward.

Under the assumption that $K > n - 1$ it is straightforward to show that:

- i) K and Σ are identifiable while the autoregressive coefficients in M (and thus M_1, M_2 and M_3) are identifiable up to their sign.
- ii) Σ is first-order identifiable up to a scale factor and M is first-order identifiable up to its sign. The degree of freedom K is not first-order identifiable but is second-order identifiable.

3.2 First-order identification

Following Gouriou et al. (2007), the first-order conditional moments can be used to calibrate the parameters in M and Σ , up to the sign and scale factor, respectively.

As the first-order method of moments is equivalent to non-linear least squares, the estimator is defined as:

$$\left(\hat{M}, \hat{\Sigma}^*\right) = \text{Argmin}_{M, \Sigma^*} S^2(M, \Sigma^*)$$

where

$$\begin{aligned} S^2(M, \Sigma^*) &= \sum_{t=2}^T \sum_{i < j} \left(Y_{ij,t} - \sum_{k=1}^n \sum_{l=1}^n Y_{kl,t-1} m_{ik} m_{lk} - \sigma_{ij}^* \right)^2 \\ &= \sum_{t=2}^T \left\| \text{vech}(Y_t) - \text{vech}(MY_{t-1}M' + \Sigma^*) \right\|^2 \end{aligned}$$

and $\Sigma^* = K\Sigma$.

As mentioned in Gouriou et al. (2007), any statistical software which accounts for heteroskedasticity can be used to obtain the estimates. We present here the complete procedure under the assumption that M is diagonal as we want to emphasize the quickness of the algorithm.

For each $Y_t, t = 1, \dots, T$ of dimensions $n \times n$, we consider the matrix \mathbf{Y} , of dimensions $T \times n(n+1)/2$ build with the *vech* of Y_t for each time $t = 1, \dots, T$; i.e. the i -th row of \mathbf{Y} is *vech*(Y_i).

Under the hypothesis that M is diagonal, define $a = \text{diag}(M)$ and $dg(a)$ as the diagonal matrix with the vector a as diagonal. Then

$$MY_{t-1}M' = dg(a)Y_{t-1}dg(a) = (aa') \odot Y_{t-1} \quad (20)$$

and

$$\text{vech}(MY_{t-1}M') = \text{vech}(aa') \odot \text{vech}(Y_{t-1}) \quad (21)$$

Define $[\mathbf{Y}]_2^T$ as the matrix obtained from \mathbf{Y} when dropping the last row, i.e. considering the time from T down to time 2. Define $A = \text{vech}(aa')$ and $Z = \text{vech}(\Sigma^*)$. The residual matrix W is obtained as

$$W = [\mathbf{Y}]_2^T - (A' \otimes \mathbf{i}_{T-1}) \odot [\mathbf{Y}]_1^{T-1} - Z' \otimes \mathbf{i}_{T-1} \quad (22)$$

where \mathbf{i}_{T-1} is a $T-1 \times 1$ vector of ones.

Then the minimization problem reduces to:

$$\left(\hat{M}, \hat{\Sigma}^*\right) = \text{Argmin}_{M, \Sigma^*} \left[\mathbf{i}'_{T-1} (W \odot W) \mathbf{i}_{n(n+1)/2} \right]. \quad (23)$$

With our data set of four assets and 2,174 trading days (see Section 4 for a detailed description), only 1.2 seconds for the diagonal case (0.7 seconds for the restricted diagonal case) on a Pentium 4 PC are necessary to obtain the estimates. This result, if compared with the 42 seconds required from the same data set when a DCC model (Engle, 2002) is fitted, represents a great improvement.¹ For the diagonal HAR-WAR only 5 seconds are required, and for its restricted version only 3.9 seconds. See Table 8 for all the other specifications.

¹To ensure a fair benchmark, we tested both our Matlab code and the one provided by Kevin Sheppard in his UCSD toolbox.

3.3 Second-order identification

Whereas the estimation of the entries of the autoregressive matrix M and of the innovation variance Σ (up to multiplication for a scale parameter) is relatively straightforward, the estimation of the degrees of freedom poses some challenges. We first present the estimation procedure introduced in [Gourieroux et al. \(2007\)](#) and then show how the same parameter K can be estimated relying on the fact that, given a portfolio allocation α , its volatility $\alpha'Y_t\alpha$ is gamma-distributed with a scale parameter equal to K .

Consider the simple WAR(1) model. The marginal distribution of the WAR(1) is the centered Wishart distribution, defined as $W(K, 0, \Sigma(\infty))$, where $\Sigma(\infty)$ is computed from

$$\Sigma(\infty) = M\Sigma(\infty)M' + \Sigma. \quad (24)$$

Thus, the conditional variance of a portfolio's volatility is given by:

$$V(\alpha'Y_t\alpha) = \frac{2}{K}[\alpha'\Sigma^*(\infty)\alpha]^2, \quad (25)$$

where α is a vector of dimension $(n \times 1)$ and $\Sigma^*(\infty) = K\Sigma(\infty)$. A consistent estimator of the degrees of freedom K can be computed as follows:

Step 1 Compute $\hat{\Sigma}^*(\infty)$ as solution of

$$\hat{\Sigma}^*(\infty) = \hat{M}\hat{\Sigma}^*(\infty)\hat{M}' + \hat{\Sigma}^*(\infty). \quad (26)$$

Step 2: Chose a portfolio allocation and compute its sample volatility

$$V(\alpha'Y_t\alpha) = \frac{1}{T} \sum_{t=1}^T \left[\alpha'Y_t\alpha - \frac{1}{T} \sum_{t=1}^T \alpha'Y_t\alpha \right]^2. \quad (27)$$

Step 3: A consistent estimator of K is:

$$K(\hat{\alpha}) = 2[\alpha'\hat{\Sigma}^*(\infty)\alpha]^2 / \hat{V}(\alpha'Y_t\alpha) \quad (28)$$

Step 4: A consistent estimator of Σ is $\hat{\Sigma}(\alpha) = \hat{\Sigma}^* / \hat{K}(\alpha)$.

A derivation of the above estimator for the general stationary WAR(p) process is reported in the Appendix.

This method provides consistent estimates of the degrees of freedom but is problematic in two aspects: first, it needs to estimate the matrix $\Sigma(\infty)$; second, it makes use of the estimates \hat{M} and $\hat{\Sigma}$, carrying their estimation error into the estimate of \hat{K} .

A more direct way that does not need to rely on the estimates of M and Σ comes from the distribution of the volatility of a portfolio.

Consider a portfolio allocation $\alpha \in \mathbb{R}^n$. We know that the unconditional distribution of Y_t is a $W(K, 0, \Sigma(\infty))$, a centered Wishart distribution. We can therefore easily show² that

$$\alpha'Y_t\alpha \sim \text{Ga} \left(\frac{K}{2}, 2\alpha'\Sigma(\infty)\alpha \right), \quad (29)$$

i.e. the distribution of the portfolio with allocation α is a gamma distribution with the degrees of freedom K as shape parameter. An unbiased estimator of K can be obtained simply via maximum likelihood

²See, for example, the proof given in [Meucci \(2005, Technical Appendix, p. 33-34\)](#) or the Appendix of this paper.

by fitting a gamma distribution to the process $\alpha'Y_t\alpha^3$. As shown in Bonato (2008), both estimators are unbiased but the second one is statistically more efficient. However, it is important to recall that these results are valid only if a WAR(1) is the true data generator process (DGP). This assumption, even if realistic, is far from being true, and a divergence in the values of the estimates is expected. In particular, Bonato (2008) shows that in the presence of extreme observations or when the DGP is not a Wishart process, the estimates for the degrees of freedom using the WAR model are perceptibly lower than predicted by the theory via gamma distribution.

A comparison of the two estimates should give a sort of measure of goodness of fit of the WAR model. A perfect fit should bring the two values to coincide.

The value of the degrees of freedom is the key element in determining whether the process is non-degenerate ($K \geq n$) or if it admits density ($K > n - 1$). Once the estimated degrees of freedom using the two estimators confirm the stationarity of the process, then the question of which estimator of K is to be used is no longer an issue, as the forecasted covariance matrices are independent of K . In fact, \hat{M} and $\hat{\Sigma}^*$ are first-order identifiable and are only required to compute $E_t(Y_{t+1})$, as shown in Equation (6). Recall that $\hat{\Sigma} = \hat{\Sigma}^*/\hat{K}$ and K is second-order identifiable. So we do not need \hat{K} to obtain $\hat{\Sigma}^*$.

4 The data

Our model introduces parametric restrictions by grouping the assets according to their type. For this reason we consider a portfolio composed of two currencies and two treasury bills. Bonds and currencies are in fact not likely to be correlated and thus our choice not to impose limited spillover between variances is justified *a priori*. As currencies we used USD/CHF and USD/GBP five-minute spot prices provided by Olsen and Associate Zürich. USD/CHF prices were available from 2 January 1997 to 9 August 2005 and USD/GBP series was covering the period from 2 January 1997 to 31 October 2006. The second group consists of the prices of the 10-year and 30-year U.S. treasury bills. These futures are traded at the Chicago Board of Trade (CBoT) from 7:20 to 14:00 Eastern Standard Time (EST). Our samples contain five-minute prices from 2 January 1997 to 29 June 2007. We adopted the conventional⁴ practice of using the futures contract with the largest trading volume. As the contract approached maturity (five trading days before), we moved to the next contract, ensuring no overlapping periods in the price sequence and no returns computed on prices from different contracts. Days in which at least one of the series had no match with the other three (e.g. when the CBoT was closed) were dropped. In addition, 23 October 1997, 9 September 1998, 14 April 2003 and 11 October 2004 were removed from the sample due to the presence of irregularities. This left us with 2,147 trading days.

Currencies are traded around the clock. T-bills are traded during the CBoT trading day and virtually round the clock on GLOBEX starting from 1 July 2003. As our samples start in 1997 we studied only the overlapping trading hours, i.e. the trading hours of the CBoT. To remove the overnight effect we did not consider the first 15 minutes after the opening. Table 4 reports the descriptive statistics for the five-minute and daily returns for the four asset we considered. The typical stylized fact are observed: negative skewness, excess of kurtosis in both daily and intraday returns.

Intraday returns were constructed taking the first differences of the log-prices and multiply by 100.

³When performing the ML estimation one should be careful to the parametrization of the Gamma density function. According to Meucci's notation, it would be for instance $\alpha'Y_t\alpha \sim \text{Ga}(K, \alpha'\Sigma(\infty)\alpha)$

⁴As done in Martens and van Dijk (2007) and de Pooter et al. (2006) among others.

Table 1: Summary statistics of five-minute and daily returns. Daily returns are computed as the logarithm of the difference between the closing price and opening price. Exchange rates are traded round the clock but as we are interested in a portfolio, only the trading hours coinciding with the CBoT trading hours were considered.

Return		CHF/USD	GBP/USD	T-10Y	T-30Y
5-min	Mean	0.0003	-0.0004	0.0001	0.0001
	Maximum	1.2716	0.6765	0.7856	0.7916
	Minimum	-1.3690	-0.6763	-1.0124	-0.8992
	St. dev.	0.0575	0.0433	0.0570	0.0367
	Skewness	-0.0322	-0.0145	-0.3391	-0.4123
	Kurtosis	16.1390	10.9153	11.1789	19.1486
Daily	Mean	-0.0250	-0.0277	0.0049	0.0076
	Maximum	3.1195	1.4240	1.9022	1.0802
	Minimum	-2.8374	-2.0079	-1.9112	-1.3626
	St. dev.	0.4967	0.3403	0.4970	0.3199
	Skewness	-0.1294	-0.0722	-0.3460	-0.3030
	Kurtosis	5.3625	4.8464	3.9230	4.2370

The trading day we constructed runs from 7:40 (first observation) to 14:00 (last observation), resulting in 76 five-minute returns which we used to construct the series realized covariance matrices. Figure 1 shows the realized volatility estimated from the data.

In the next step we constructed the series of realized covariance matrices using the classical estimator presented in Barndorff-Nielsen and Shephard (2004) and used, for example, in de Pooter et al. (2006):

$$Y_t = \sum_{i=1}^I r_{t-1+ih,h} r_{t-1+ih,h}' \quad (30)$$

We indicate with Y_t the realized covariance matrix at time t in order to be coherent with our previous notation and because the use of Σ would probably create confusion as Σ denotes the covariance matrix of the Gaussian vector underlying the WAR(1) model. $r_{t-1+ih,h} \equiv p_{t-1+ih} - p_{t-1+(i-1)/h}$ denotes the $(n \times 1)$ vector of returns for the i -th intraday period on day t , for $i = 1, \dots, I$, and with $n = 4$ the number of stocks. I is the number of intraday intervals, each of length $h \equiv 1/I$. In our case, with a frequency of five minutes, $I = 76$. In contrast to de Pooter et al. (2006) we did not consider overnight returns. Including overnight returns would affect only the volatility of the T-bills because currencies are traded 24 hours and their equivalent to the overnight returns would be the over-weekend return. Therefore we contend that adding overnight returns to only some components of the portfolio would induce distortion in the realized volatility of the portfolio itself.

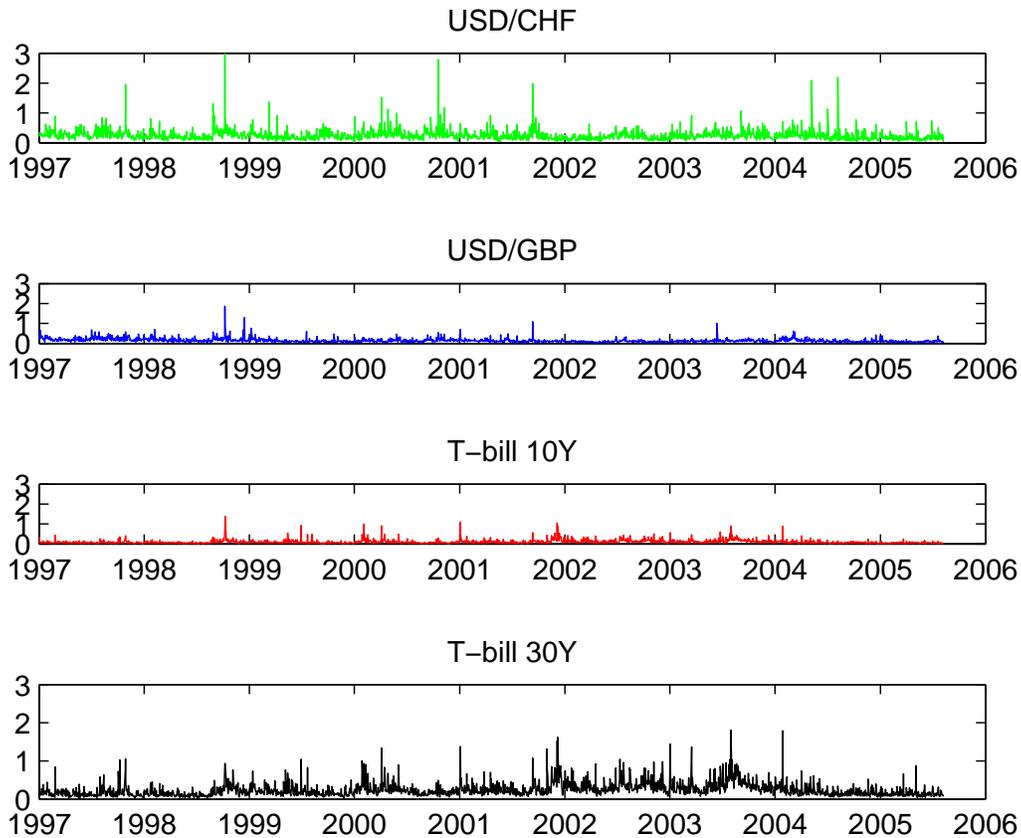


Figure 1: Daily realized volatility for the two currencies and the two treasury bonds.

Table 2: Summary statistics for the realized volatilities

Realized volatility	CHF/USD	GBP/USD	T-10Y	T-30Y
Mean	0.2511	0.1422	0.2466	0.1022
Maximum	2.9772	1.8661	1.8043	1.3761
Minimum	0.0184	0.0164	0.0276	0.0119
St. dev.	0.1856	0.1039	0.1895	0.1006
Skewness	5.5066	4.8388	2.6636	4.5772
Kurtosis	59.7536	54.3341	14.2783	37.2670

5 Empirical application

5.1 Estimation results

The first model we estimated is the full WAR(1), in which the matrix M is full. The estimates are reported in Table 3. As shown in Equation (10), the impact of the past values of realized variances and covariances on future realized variances and covariances is a function of the entries of M , so, rather than checking

the significance of the elements of M , we are interested in checking the significance of the coefficients a_i, b_i, c_i , $i = 1, \dots, 3$, i.e. the coefficients that directly affect the realized variance-covariance matrix forecasts.

0.4044	0.1033	0.0764	-0.1442
(3.3985)	(0.2273)	(0.1868)	(-0.2282)
-0.0602	0.5637	-0.0344	0.0600
(-0.2441)	(4.2327)	(-0.1067)	(0.1235)
0.0323	0.0008	0.7204	-0.1047
(0.2425)	(0.0003)	(3.3614)	(-0.3092)
-0.0128	0.0489	0.1753	0.4037
(-0.0715)	(0.2063)	(0.5773)	(0.9577)

Table 3: Estimated latent autoregressive matrix M for the full WAR(1) model. t-ratios in parenthesis.

0.0424	0.0007	-0.0011	0.0002
(7.9627)	(0.1110)	(-0.1812)	(0.0445)
	0.0197	-0.0017	-0.0023
	(3.7092)	(-0.3023)	(-0.4465)
		0.0279	0.0136
		(4.8620)	(2.7554)
			0.0123
			(2.8124)

Table 4: Estimated latent autoregressive matrix Σ for the full WAR(1) model. t-ratios in parenthesis.

Table 5 reports the estimates and the t-test values of the parameters that determine the best prediction of Y_t as given by a WAR(1) model. For simplicity we will only consider the case of two assets and report the estimates of the different pairs of combinations of the two currencies and two T-bills we used in our analysis. The parameter a_1 , which tells us the effect of the realized volatility at time $t - 1$ on the realized volatility expected at time t , is significant for all the pairs⁵. We have the same results for the coefficients b_2 and c_3 , the autoregressive parameters for the realized covariances and realized variances of the second component of the pair. The only exceptions are the couples CHF-GBP and T30-T10. In particular, for the latter pair, only the autoregressive coefficient for the 30-year U.S. treasury bill is statistically significant.

It is very important to note that the rest of the coefficients are not statistically significant for any of the different combinations of pairs. This suggests that a reduction of the parameters of the models hypothesizing a limited spillover is reasonable and to some extent necessary.

The estimates of the autoregressive matrix M for the four specifications of the WAR(1) model, the diagonal, the diagonal restricted, the block-diagonal and the restricted block-diagonal are reported in Table 6. Standard errors are in parenthesis. Starting at the top left of the table, we see that imposing the same value of the autoregressive coefficient for assets belonging to the same type is a sensible choice. Consider the diagonal WAR case. For the first two elements of the diagonal (exchange rates), we have a common

⁵Recall from (10) that $a_1 = m_{11}^2$ so that the significance test is a one-sided test with 10% level at 1.28.

	CHF-GBP	CHF-T30	CHF-T10	GBP-T30	GBP-T10	T30-T10
a_1	0.1613 (1.5543)	0.1786 (2.1789)	0.1806 (2.1754)	0.3279 (2.2469)	0.3364 (2.2960)	0.5419 (1.7310)
a_2	-0.0418 (-0.4640)	0.0130 (0.2369)	-0.0027 (-0.0340)	0.0081 (0.0857)	0.0196 (0.1500)	0.1304 (0.5772)
a_3	0.0108 (0.2190)	0.0009 (0.1184)	0.0000 (0.0170)	0.0002 (0.0429)	0.0011 (0.0750)	0.0314 (0.2874)
b_1	-0.0835 (-0.3802)	0.0260 (0.2363)	-0.0054 (-0.0338)	0.0162 (0.0857)	0.0392 (0.1500)	0.2607 (0.5638)
b_2	0.2051 (1.2183)	0.2783 (4.0238)	0.2629 (3.2827)	0.3783 (4.1564)	0.3627 (3.4078)	0.2722 (0.7471)
b_3	-0.1171 (-0.4521)	0.0406 (0.2365)	-0.0078 (-0.0340)	0.0187 (0.0856)	0.0421 (0.1491)	0.1417 (1.3579)
c_1	0.0412 (0.2635)	0.0004 (0.0799)	0.0040 (0.1065)	0.0001 (0.0417)	0.0013 (0.0830)	0.0161 (0.1876)
c_2	0.1143 (0.5561)	-0.0137 (-0.1598)	-0.0389 (-0.2135)	0.0062 (0.0833)	0.0224 (0.1662)	-0.0507 (-0.3345)
c_3	0.3173 (1.9316)	0.4356 (5.4035)	0.3815 (2.4987)	0.4361 (5.2972)	0.3883 (2.5243)	0.1602 (0.4589)

Table 5: Estimates and t-ratios for the coefficients of Equation (10). Coefficients that are significant at the 10% level are shown in bold.

parameter 0.4585 against 0.4175 and 0.5636. For the T-bills we have an autoregressive parameter for the volatilities equal to 0.6481 in front of 0.6583 and 0.6209. Including spillover between assets belonging to the same sector affects only the autoregressive parameter of the 30-years T-bill and appears unnecessary as most of the off-diagonal coefficients are not significant at the 5% level, confirming the findings reported in Table 5. The restricted block diagonal case presents estimates that are not compatible with the previous cases and this seems to suggest that this kind of specification might be too restrictive to model the covariance matrix. The estimation results for the HAR-WAR process are similar to those for the WAR process and are available upon request.

The estimated values for the degrees of freedom are reported in Table 8. To obtain the estimates the following allocation was used: $\alpha = (1 \ 1 \ 1 \ 1)'$. Different allocations led to analogous results.

All the different specifications result in a number of degrees of freedom strictly bigger than n , $n = 4$ being the number of assets making up the portfolio, and thus the Wishart process is stationary and non-degenerate. All the estimates of K are close to each other except for the restricted block WAR-HAR. The resulting degrees of freedom equal to 6.5 are slightly bigger than in the other cases and this might be due to some problem in the optimization routine. Further investigation in this direction is necessary.

In addition to the estimated degrees of freedom, Table 8 also reports the number of parameters for each model and the CPU time necessary to obtain the estimates on a Pentium IV PC. The advantage of using a diagonal model (either WAR or HAR), compared with the full counterpart, is notable. The time required to obtain the estimates ranges from 0.71 to 5 seconds, a great improvement compared, for example, with the 323 required by the diagonal BEKK of Engle and Kroner (1995), which assumes the

Block WAR				Restricted block WAR			
0.4080	0.1060			0.2740	0.2740		
(3.5332)	(0.2383)			(4.8680)			
-0.0648	0.5626			0.2740	0.2740		
(-0.2649)	(4.2528)						
		0.7216	-0.1078			0.3282	0.3282
		(3.3565)	(-0.3175)			(12.8269)	
		0.1716	0.4035			0.3282	0.3282
		(0.5640)	(0.9389)				
Diagonal WAR				Restricted diagonal WAR			
0.4175				0.4584			
(4.2792)				(5.9889)			
	0.5636				0.4584		
	(4.4107)						
		0.6583				0.6481	
		(11.1432)				(13.595)	
			0.6209				0.6481
			(6.0008)				

Table 6: Estimated latent autoregressive matrix M for the different specification of the WAR(1) model. t-ratios in parenthesis.

same autoregressive structure for the latent variance-covariance matrix⁶.

5.2 Variance Forecasting

The ability to forecast the volatility of a financial position is a key factor in many activities like risk management, portfolio optimization or option pricing, just to mention the most common. For this reason we preferred to give more emphasis to the out-of-sample forecast of the proposed model, rather than the in-sample fit and in-sample forecast. Of course, in-sample fit is important to determine the goodness of a model; however, unreported results showed that the WAR models have a very poor in-sample forecasting ability. Our suspicion is that the degrees of freedom are unlikely to be constant through time, and therefore fitting the model to the entire series is not appropriate. To check the variation of the degrees of freedom within the sample, we split the 2,147 trading days into non-overlapping periods of 30 days. We then estimated the degrees of freedom for each sub-period. Results are reported in Figure 2. We can clearly see that the degrees of freedom are far from being constant over time, with values ranging approximately between 3 and 20.

As done by banks and regulators, we use a rolling window to perform one-day-ahead out-of-sample forecasts. Our first step is to construct a portfolio with the series of two exchange rates and two treasury bills. We assume that the value of the portfolio is in dollars and that it therefore carries a long position for the treasury bills and a short position in currencies. For simplicity, we assume equal (positive) weights

⁶ Again, to estimate the parameters of the BEKK model we used the Matlab code provided by Kevin Sheppard in the UCSD Garch toolbox.

Block WAR				Restricted block WAR			
0.0424	0.0007	-0.0002	-0.0003	0.0451	-0.0049	0.0000	0.0000
(8.0529)	(0.1140)	(-0.0604)	(-0.0828)	(11.1560)	(-1.2006)	(0.0070)	(0.0069)
	0.0197	-0.0019	-0.0014		0.0228	-0.0024	-0.0016
	(3.7136)	(-0.6149)	(-0.4363)		(5.6421)	(-0.7805)	(-0.4998)
		0.0279	0.0136			0.0371	0.0127
		(4.8738)	(2.7514)			(9.9012)	(3.3877)
			0.0124				0.0076
			(2.8801)				(2.0225)
Diagonal WAR				Restricted diagonal WAR			
0.0424	0.0011	-0.0004	-0.0004	0.0406	0.0011	-0.0004	-0.0004
(8.1190)	(0.3423)	(-0.1238)	(-0.1264)	(8.5055)	(0.3536)	(-0.1201)	(-0.1201)
	0.0198	-0.0019	-0.0014		0.0230	-0.0021	-0.0015
	(3.7920)	(-0.6012)	(-0.4396)		(6.1516)	(-0.6732)	(-0.4760)
		0.0285	0.0154			0.0292	0.0151
		(5.6888)	(4.2106)			(6.6184)	(4.2865)
			0.0128				0.0121
			(3.1117)				(3.5788)

Table 7: Estimated latent autoregressive matrix Σ for the different specification of the WAR(1) model. t-ratios in parenthesis.

for the treasury bills and equal (negative) weights for the exchange rates. In particular, we assume the owner of the portfolio invests 0.75 of his wealth for each of the T-bills and short-sells 0.25 for each of the currencies to buy CHF and GBP against USD, respectively. The forecasting period runs from 2 January 2003 until 8 August 2005, resulting in 653 one-step-ahead forecasts. For each day the realized variance of the portfolio is forecast by fitting a WAR model to the series of covariance matrices and re-estimating the model at each step. As already mentioned above, the degrees of freedom are likely not to be constant and therefore at each step the model was estimated using a rolling window of 100 trading days, as done in Ait-Sahalia and Mancini (2008). Table 9 presents the results of the Mincer-Zarnowitz regression:

$$IV_t^{1/2} = b_0 + b_1 E_{t-1}[RV_t]^{1/2} + \text{error}, \quad (31)$$

where IV_t is the realized volatility of the portfolio at time t and $E_{t-1}[RV_t]$ is the forecasted realized volatility. Standard errors are reported in parenthesis. The R^2 across the models varies from 0.3209 for the full WAR(1) to the 0.3655 for the diagonal HAR-WAR. The moving windows estimation of the various WAR models delivered acceptable R^2 , that are, for instance, slightly higher than those reported in Andersen et al. (2003).

It interesting to note that the full WAR(1) model has a worse performance if compared with its restricted counterparts. This might be due to the fact that the the full model is not the most appropriate as it carries over the estimation error of the parameters into the forecasts, which means that it is not as good as a more parsimonious model. It should also be noted that, in terms of R^2 , the difference between the diagonal model and the restricted diagonal model is not relevant. Neither is the difference between the block diagonal and the restricted block diagonal. The diagonal model has the highest R^2 . This suggests

Specification	Parameters	CPU time (secs)	\hat{K}	fval	Ranking
full WAR	27	117	4.8	209.01	9
block diag. WAR	19	94	4.9	209.11	8
restr. block diag. WAR	13	21	4.8	231.96	5
diagonal WAR	15	1.22	4.8	209.39	2
restr. diag. WAR	13	0.71	4.8	209.80	1
full HAR-WAR	59	531	4.7	189.78	11
block diag. HAR-WAR	35	410	4.7	189.37	10
restr. block diag. HAR-WAR	17	92	6.5	198.52	7
diagonal HAR-WAR	23	3.5	4.6	187.45	4
restr. diag. HAR-WAR	17	2.5	4.7	187.54	3
DCC	14	42	-	-	6
diag. BEKK	18	639	-	-	12
\hat{K} via gamma dist.			7.09	s.e. (0.8)	

Table 8: Estimate of the degrees of freedom for the different specifications of the WAR and HAR-WAR models (last column). The first column reports the number of parameters for each specification. The CPU necessary to obtain the estimates are reported in the second column. *fval* is the value of the function (23) at the minimum. The last row reports the value of K when it is estimated relying on the gamma distribution for the variance of the portfolio.

that this simple parametrization is sufficient to capture the dynamics of the variances and covariances.

5.3 Distribution of the portfolio's realized volatility

As demonstrated in the Appendix, under the WAR hypothesis the realized volatility of a portfolio follows a gamma distribution with shape parameter $K/2$, where K denotes the degrees of freedom of the Wishart process and scale parameter $2\omega'\Sigma(\infty)\omega$ with $\Sigma(\infty)$ solution of

$$\Sigma(\infty)^* = M\Sigma(\infty)^*M' + \Sigma^*.$$

as in (26), where ω is the vector of portfolio weights, i.e. $\omega = [-.25 \quad -.25 \quad .75 \quad .75]'$. Figure 4 (left) displays the density of the realized volatility of the portfolio under the hypothesis that it follows a gamma distribution. The dashed red line represents the kernel density of the portfolio's realized volatility. The green dash-dot line is the density of a $Ga(K_\Gamma/2, 2\omega'\Sigma(\infty)\omega)$ where K_Γ denotes the degrees of freedom estimated via the gamma distribution. The blue line is the density of a gamma distribution but with K estimated as in Gouriou et al. (2007), Steps 1-4. Recall that to obtain both the estimates for K $\alpha = (1 \ 1 \ 1 \ 1)'$ was used.

In Figure 4 (right) we fitted a gamma distribution to the realized volatility of our portfolio. The blue line represents the kernel density of the realized variance, the blue line is the gamma fitting and the black dash dot line represents the log-normal density. Numerous studies (Andersen et al., 2003, among others) show that the logarithm of the realized volatility tends to follow a normal distribution. Is therefore no surprising that a lognormal distribution clearly better fits the distribution of the realized volatility of the portfolio. On the other hand, the fit provided by the Wishart model, i.e. the a gamma distribution, from a very rough graphical analysis, provides an acceptable alternative⁷.

⁷The assumption of a gamma distribution to model the realized volatility is also at the basis of the multiplicative model of Engle and Gallo (2006)

	b_0	b_1	R^2
full WAR(1)	0.0226 (0.0333)	0.8988 (0.0512)	0.3209
block diagonal WAR(1)	0.0004 (0.0342)	0.9349 (0.0526)	0.3262
restr. block diag. WAR(1)	0.0046 (0.0341)	0.9405 (0.0524)	0.3224
diagonal WAR(1)	0.0064 (0.0343)	0.9434 (0.0526)	0.3299
restr. diag. WAR(1)	0.0059 (0.0342)	0.9428 (0.0526)	0.3298
full HAR-WAR	0.1387 (0.0275)	0.7361 (0.0429)	0.3103
block diag. HAR-WAR	0.0685 (0.0284)	0.8439 (0.0442)	0.3584
restr. block diag. HAR-WAR	0.0647 (0.0284)	0.8440 (0.0438)	0.3623
diagonal HAR-WAR	0.0520 (0.0289)	0.8630 (0.0446)	0.3662
restr. diag. HAR-WAR	0.0550 (0.0286)	0.8594 (0.0443)	0.3655

Table 9: Out-of-sample one-day-ahead forecast of $IV^{1/2}$. The models are estimated on a rolling window of 100 days from 2 January 2003 to 8 August 2005. Standard errors in parenthesis.

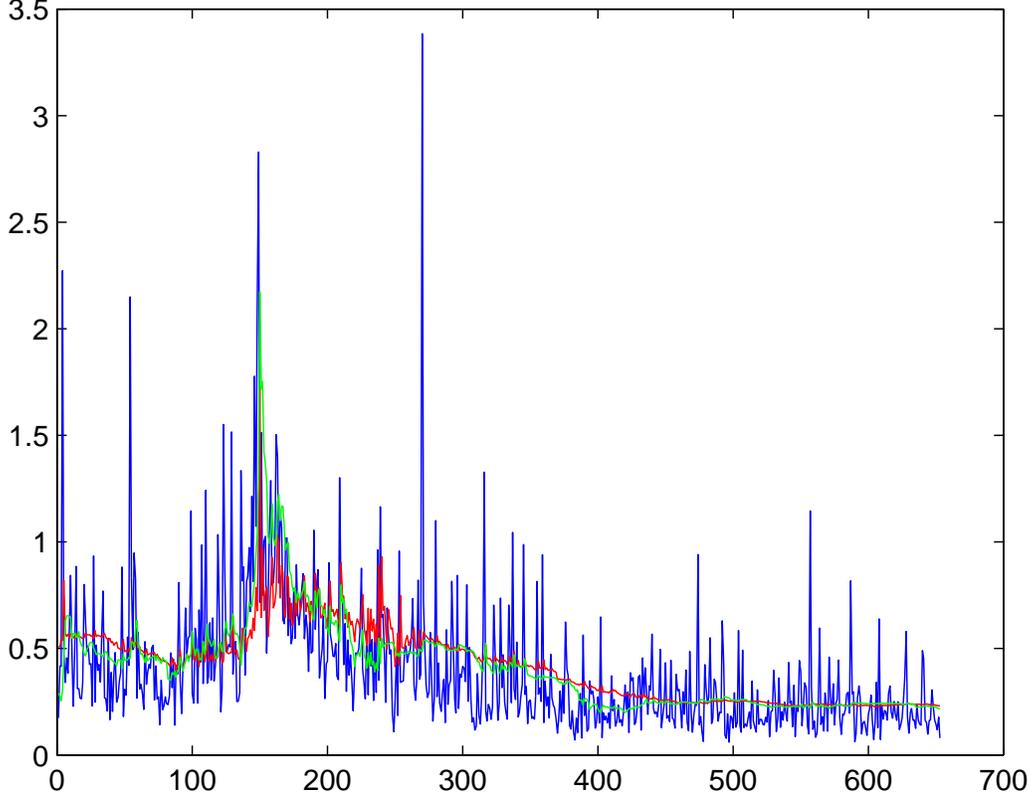


Figure 3: Out-of-sample forecast of the realized variance for the restricted diagonal WAR(1) (red line) and the restricted diagonal HAR-WAR model (green line). The blue line represents the ex-post observed realized volatility of the portfolio.

where $c(\cdot, \Omega_{t-1})$ and $h(\cdot, \Omega_{t-1})$ are functions of Ω_{t-1} (the information set at time $t - 1$), and depend on an unknown vector of parameters η ; ν_t is an independent and identically distributed (i.i.d.) process, independent of Ω_{t-1} , with $E[\nu_t] = 0$ and $E[\nu_t^2] = 1$. μ_t is the conditional mean of r_t and σ_t is its conditional variance. In our setting we assume, for simplicity, a constant mean for all the assets in our portfolio. In particular, if r_t represents the return of the portfolio, $\mu_t = \mu$ and for the (realized) variance of the portfolio we have:

$$RV_t = \omega' Y_t \omega, \quad (36)$$

where ω are the portfolio weights as previously chosen. To compute one-day-ahead forecasts for the VaR of the daily return r_t using the conditional realized volatility, we re-estimate the model in Eq. (32) with constant conditional mean while the conditional variance is proportional to $RV_{t|t-1}$, the one-step-ahead forecast of the realized volatility of the portfolio; i.e. $\sigma_t^2 = \sigma^2 RV_{t|t-1}$ (with σ^2 being an additional parameter to be estimated). σ^2 is used to ensure that the rescaled innovations have unit variance.

We used the same forecasting period as in the previous section. For each model we computed the one-day-ahead variance and then the one-day-ahead forecast of the VaR. A Gaussian distribution and a Student's t distribution were used to model the residuals z_t . Table 10 presents the performances of the different models in terms of VaR predictions. Forecasts of VaR at level $\rho = 1\%, 5\%$ and 10% were computed. For each model and distribution for ν_t , we reported the percentage of violations, i.e. the

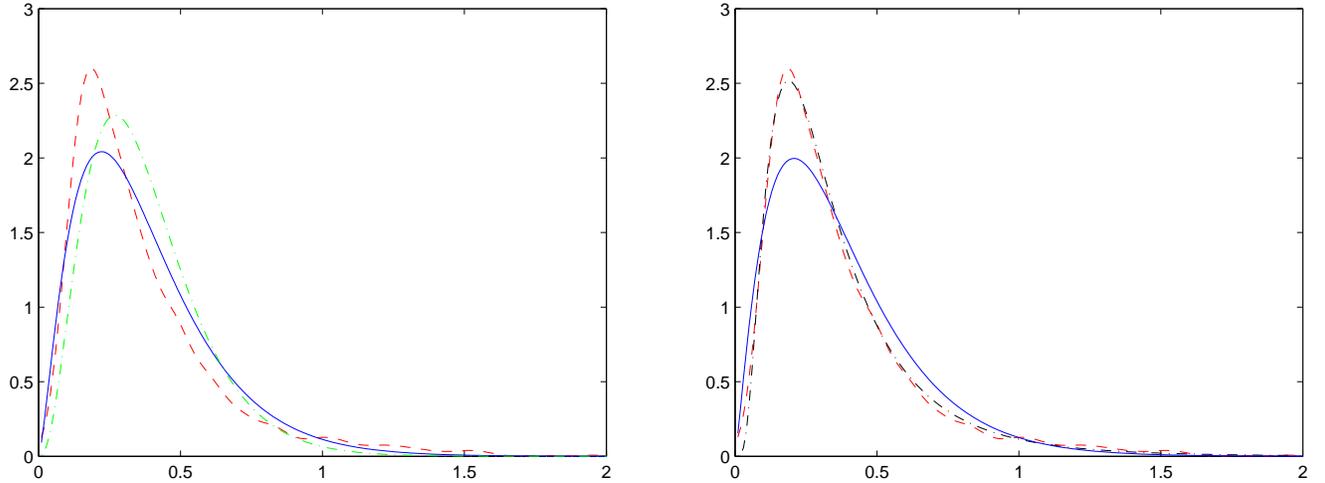


Figure 4: Kernel densities of the realized volatility of the portfolio (red dashed line), density of a $Ga(K_\Gamma/2, 2\omega'\Sigma(\infty)\omega)$ where K_Γ denotes the degrees of freedom estimated via the gamma distribution (blue line) and density of a gamma distribution (green dash-dot line) with K estimated as in Gouriéroux et al. (2007) [left-hand panel]. Kernel densities of the realized volatility of the portfolio (red dashed line), gamma (blue line) and log-normal (black dash-dot) distribution fitted to the series [right-hand panel].

percentage of times that the realized return is smaller than the forecasted VaR. A good density forecast should satisfy two criteria. First, for a given VaR level ρ , the percentage of violations should be ρ . Second, violations should be conditionally unpredictable, i.e. a violation of nominal ρ_1 VaR today should convey no information as to whether nominal ρ_2 percent VaR will be violated tomorrow.

To check the robustness of the different VaR models in this VaR forecast evaluation, we also report in Table 10 the p-values of the test proposed in Berkowitz (2001) to evaluate a density forecast. This test relies on the fact that for a given daily return r_t , if the series of one-day-ahead conditional density forecasts $\hat{f}_{t|t-1}(r_t)$ coincides with $f(r_t, \mathbb{I}_{t-1})$, it then follows under weak conditions that the sequence of probability integral transformation of r_t with respect to $\hat{f}_{t|t-1}(\cdot)$

$$u_t = \int_{-\infty}^{r_t} \hat{f}_{t|t-1}(s) ds = \hat{F}(r_t) \quad (37)$$

should be i.i.d. uniformly distributed on $(0,1)$. This transformation was first presented in Rosenblatt (1952).

If the series of u_t is distributed as an i.i.d. $U(0,1)$, then

$$z_t = \Phi^{-1} \left[\int_{-\infty}^{r_t} \hat{f}_{t|t-1}(s) ds \right] \quad \text{is an i.i.d. } N(0,1).$$

Once the series has been transformed, it is straightforward to calculate the Gaussian likelihood and construct the likelihood ratio (LR) statistics.

In particular, Berkowitz (2001) suggested a test that allows the user to intentionally ignore model failures that are limited to the interior of the distribution; the proposed LR test is based on a censored likelihood: the tail of the forecasted density is compared with the observed tail.

First, for different values of ρ the desired cutoff point $\text{VaR} = \Phi^{-1}(\rho)$ is computed. Then we define the

new variable of interest as

$$z_t^* = \begin{cases} \text{VaR} & \text{if } z_t \geq \text{VaR} \\ z_t & \text{if } z_t < \text{VaR}. \end{cases}$$

The log-likelihood function for joint estimation of μ and σ^2 is

$$L(\mu, \sigma | z^*) = \sum_{z^* < \text{VaR}} \log \frac{1}{\sigma} \phi \left(\frac{z_t^* - \mu}{\sigma} \right) + \sum_{z^* = \text{VaR}} \log \left(1 - \Phi \left(\frac{\text{VaR} - \mu}{\sigma} \right) \right) \quad (38)$$

$$= \sum_{z^* < \text{VaR}} \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (z_t^* - \mu)^2 \right) + \sum_{z^* = \text{VaR}} \log \left(1 - \Phi \left(\frac{\text{VaR} - \mu}{\sigma} \right) \right). \quad (39)$$

To construct the LR test the null hypothesis requires that $\mu = 0$, $\sigma^2 = 1$. Therefore the restricted likelihood $L(0,1)$ is compared to the unrestricted one, $L(\hat{\mu}, \hat{\sigma}^2)$. The test statistic is then

$$LR_{\text{tail}} = -2(L(0,1) - L(\hat{\mu}, \hat{\sigma}^2)) \quad (40)$$

Under the null hypothesis, the test statistic is distributed $\chi^2(2)$.

[Table 10 somewhere here]

Table 10 reports, for the different models considered and different assumptions for the residuals, the percentage of violations along with the p-value of the Berkowitz's test.

The relative number of violations is close to the theoretical one and assuming a t distribution for the residuals does not really improve the forecasting performances. For all the proposed specifications of the WAR model, the Berkowitz test does not reject the null hypothesis of appropriateness of the forecasted densities. Therefore all the models provide acceptable VaR forecasts. For the 1% VaR level, the results are somewhat surprising. The percentage of VaR violations is, for all the specifications, around 2.4% in front of a theoretical value of 1%. However, the p-values of the Berkowitz test are all higher than the rejection threshold of, say, 5%. This might be explained by the fact that the test proposed by Berkowitz is not a pointwise evaluation of the VaR violations, but rather analyzes the entire forecasted densities, or, in our case, the left tail of the distribution.

Besides the good forecasting performances of the proposed models, we want to stress the fact that there is no notable difference in the forecasting ability of the different specifications. Therefore, a very parsimonious (and thus quick to estimate) model like the restricted diagonal WAR is sufficient to model the riskiness of our portfolio.

6 Conclusions and direction for future research

In this paper we proposed a particular set of restricted specification of the WAR model for realized (co)variances. Our specifications rely on the ability to group assets according to some criterion, for example the economic sector, a common feature in the variance-covariance dynamics, and so on. This allowed us to drastically reduce the number of parameters. A comparison between the different specifications highlighted that there is no loss when a more parsimonious model is chosen. This is essentially due to the fact that the restricted model was justified by the data.

However, some aspects of the WAR process need to be clarified. In particular, the degrees of freedom seem to vary through time and it is not clear by which variables they are driven.

A straightforward extension of the present work involves applying the WAR model to solve concrete financial problems like dynamic portfolio choice, for instance.

This and other applications of the WAR model are left for future research.

References

- AÏT-SAHALIA, Y. AND L. MANCINI (2008): “Out of sample forecast of quadratic variation,” *Journal of Econometrics*, *forthcoming*.
- ANDERSEN, T. AND T. BOLLERSLEV (1997): “Heterogeneous information arrivals and return volatility dynamics: Uncovering the long run in high frequency data,” *The Journal of Finance*, 52, 975–1005.
- (1998): “Answering the skeptics: Yes, standard volatility models do provide accurate forecasts,” *International Economic Review*, 39, 885–905.
- ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND H. EBENS (2001a): “The distribution of realized stock return volatility,” *Journal of Financial Economics*, 43–76.
- ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2001b): “The distribution of realized exchange rate volatility,” *Journal of the American Statistical Association*, 96, 42–55.
- (2003): “Modeling and forecasting realized volatility,” *Econometrica*, 71, 579–625.
- ANDREOU, E. AND E. GHYSELS (2002): “Rolling sample volatility estimators: some new theoretical, simulation and empirical results,” *Journal of Business and Economic Statistics*, 20, 363–376.
- ASAI, M., M. CAPORIN, AND M. MCALEER (2008): “Block structure multivariate stochastic volatility,” *Manuscript*.
- BANDI, F. AND J. RUSSEL (2005): “Realized covariation, realized beta and microstructure noise,” *unpublished paper*.
- BANDI, F., J. RUSSEL, AND Y. ZHU (2006): “Using high-frequency data in dynamic portfolio choice,” *Econometric Reviews*, *forthcoming*.
- BARNDORFF-NIELSEN, O. AND N. SHEPHARD (2004): “Econometric analysis of realized covariation: high-frequency based covariance, regressions, and correlation in financial economics,” *Econometrica*, 72, 885–925.
- BAUER, G. AND K. VORKINK (2007): “Multivariate realized stock market volatility,” *Bank of Canada Working Papers 2007-20*.
- BERKOWITZ, J. (2001): “Testing densities forecasts, with applications to risk management,” *Journal of Business and Economic Statistics*, 19, 465–474.
- BILLIO, M. AND M. CAPORIN (2008): “A generalised dynamic conditional correlation model for portfolio risk evaluation,” *Mathematics and Computers in Simulations*, *forthcoming*.
- BILLIO, M., M. CAPORIN, AND M. GOBBO (2006): “Flexible dynamic conditional correlation multivariate GARCH for asset allocation,” *Applied Financial Economics Letters*, 2, 123–130.

- BONATO, M. (2008): "Estimating the degrees of freedom of the realized volatility Wishart autoregressive model," *Manuscript*.
- BROWNLEES, C. AND G. GALLO (2008): "Comparison of Volatility Measures: A Risk Management Perspective," .
- CAPORIN, M. AND P. PARUOLO (2008): "Spatial dependence in multivariate volatility models," *Working Paper*.
- CHIRIAC, R. (2007): "Nonstationary Wishart autoregressive model," *Working Paper*.
- CHIRIAC, R. AND V. VOEV (2008): "Modelling and forecasting multivariate realized volatility," *Working Paper*.
- CLEMENTS, M. P., A. B. GALVAO, AND J. H. KIM (2008): "Quantile forecasts of daily exchange rates returns from forecasts of realized volatility," *Journal of Empirical Finance*, 15, 729–750.
- CORSI, F. (2004): "A simple long memory model of realized volatility," *Tech. Report, University of Southern Switzerland*.
- CORSI, F., U. KRETSCHMER, S. MITTNIK, AND C. FIGORSCH (2007): "The volatility of realized volatility," *Journal of Financial Econometrics*, 27, 46–78.
- DE POOTER, M., M. MARTENS, AND D. VAN DIJK (2006): "Predicting the daily covariance matrix for S&P 100 stocks using intraday data - But which frequency to use?" *Econometric Reviews, forthcoming*.
- ENGLE, R. (2002): "Dynamic conditional correlation - a simple class of multivariate GARCH models," *Journal of Business and Economic Statistics*, 20, 339–350.
- ENGLE, R. AND G. GALLO (2006): "A Multiple Indicators Model for Volatility Using Intra-Daily Data," *Journal of Econometrics*, 131, 3–27.
- ENGLE, R. AND B. KELLY (2008): "Dynamic Equicorrelations," *Working Paper*.
- ENGLE, R. F. AND F. K. KRONER (1995): "Multivariate simultaneous generalized ARCH," *Econometric Theory*, 11, 122–150.
- FLEMING, J., C. KIRBY, AND B. OSTDIK (2001): "The economic value of volatility," *Journal of Finance*, 56, 329–352.
- (2003): "The economic value of volatility timing using 'realized' volatility," *Journal of Financial Economics*, 67, 473–509.
- FOSTER, D. AND D. NELSON (1996): "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64, 139–174.
- GALLANT, R., C. HSU, AND G. TAUCHEN (1999): "Using daily range data to calibrate volatility diffusions and extract the forward integrated variance," *The Review of Economics and Statistics*, 81, 617–631.
- GIOT, P. AND S. LAURENT (2004): "Modeling daily Value-at-Risk using realized volatility and ARCH type models," *Journal of Empirical Finance*, 11, 379–398.

- GOURIEROUX, C. (2007): "Positivity conditions for a bivariate autoregressive volatility specifications," *Journal of Financial Econometrics*, 5, 624–636.
- GOURIEROUX, C., J. JASIAK, AND R. SUFANA (2007): "The Wishart autoregressive process of multivariate stochastic volatility," *Journal of Econometrics*, forthcoming.
- GRANGER, C. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.
- HAYASHI, T. AND N. YOSHIDA (2005): "On covariance estimation of non-synchronously observed diffusion processes," *Bernoulli*, 11, 359–379.
- (2006): "Estimating correlations with nonsynchronous observations in continuous diffusion models," *Working Paper*.
- LEBARON, B. (2001): "Stochastic volatility as a simple generator of financial power-laws and long memory," *Quantitative Finance*, 1, 621 – 631.
- MARTENS, M. AND D. VAN DIJK (2007): "Measuring volatility with the realized range," *Journal of Econometrics*, 181–207.
- MCALEER, M. AND M. MEDEIROS (2006): "Realized volatility: a review," *Working Paper*.
- MEUCCI, A. (2005): *Risk and asset allocation*, Springer.
- MÜLLER, U., M. DACOROGNA, R. DAVE, R. OLSEN, R. PIQUET, AND J. VON WEIZSÄCKER (1997): "Volatilities of different time resolutions - Analyzing the dynamics of market components," *Journal of Empirical Finance*, 4, 213–239.
- MÜLLER, U., M. DACOROGNA, R. DAVE, O. PICTET, R. OLSEN, AND J. WARD (1993): "Fractals and intrinsic time - a challenge to econometricians," *XXXIX International AEA Conference on Real Time Econometrics, 14-15 Oct 1993, Luxembourg*.
- PONG, S., M. SHACKLETON, S. TAYLOR, AND X. XU (2004): "Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models," *Journal of Banking and Finance*, 28, 2514–2563.
- ROSENBLATT, M. (1952): "Remarks on a multivariate transformation," *The Annals of Mathematical Statistics*, 23, 470–472.
- SHEPPARD, K. (2006): "Realized covariance and scrambling," *Manuscript*.
- THOMAKOS, D. AND T. WANG (2003): "Realized volatility in the futures markets," *Journal of Empirical Finance*, 321–353.
- ZHANG, L. (2006): "Estimating covariation: Epps effect, microstructure noise." *Working Paper*.

A Appendix

A.1 Relation between Wishart and gamma distribution

This proof follows the one in the Technical Appendix in Meucci (2005).

If \mathbf{Y} is a Wishart distribution, then for any comfortable matrix \mathbf{A} we have

$$\mathbf{A}\mathbf{Y}\mathbf{A}' = \mathbf{A}\mathbf{X}_1\mathbf{X}_1'\mathbf{A}' + \dots + \mathbf{A}\mathbf{X}_K\mathbf{X}_K'\mathbf{A}' \quad (41)$$

$$= \mathbf{Z}_1\mathbf{Z}_1' + \dots + \mathbf{Z}_K\mathbf{Z}_K' \quad (42)$$

$$\sim W(K, \mathbf{A}\Sigma\mathbf{A}') \quad (43)$$

since

$$\mathbf{X}_t \sim N(0, \Sigma) \quad (44)$$

and

$$\mathbf{Z}_t \equiv \mathbf{A}\mathbf{X}_t \sim N(0, \mathbf{A}\Sigma\mathbf{A}'). \quad (45)$$

By taking a row vector, i.e. $\mathbf{A} \equiv a'$, each term in the sum is normally distributed as follows:

$$Z_t \equiv a'\mathbf{X}_t \sim N(0, a'\Sigma a). \quad (46)$$

Now, for any random variable

$$y_i \sim N(0, \sigma^2) \quad (47)$$

the gamma distribution with K degrees of freedom is defined as the distribution of the following variable:

$$x = y_1^2 + \dots + y_K^2 \sim \text{Ga}(K/2, 2\sigma^2). \quad (48)$$

and has p.d.f. of the form⁸

$$f(x|K/2, 2\sigma^2) = \frac{1}{(2\sigma^2)^{K/2}\Gamma(K/2)} x^{K/2-1} e^{-x/2\sigma^2}. \quad (49)$$

Therefore from (48)

$$a'\mathbf{Y}a \sim \text{Ga}(K/2, 2(a'\Sigma a)). \quad (50)$$

Note that in Meucci (2005) we have $a'\mathbf{Y}a \sim \text{Ga}(K, (a'\Sigma a))$, because a different parametrization of the gamma distribution is used.

A.2 Estimation of the degrees of freedom for a general WAR(p) process

We present here a way to derive the estimator of the degrees of freedom K in a general WAR(p) process. Differently from Chiriac (2007), we do not rely on the interpretation of a WAR process in terms of a Gaussian VAR process; in fact, for a WAR(p) process with $p > 1$ this interpretation is no longer valid (see Gourieroux et al., 2007). Instead, we use the fact that any portfolio of Wishart-distributed matrices follows a gamma distribution, as shown in the previous section.

⁸Recall that if $x \sim \text{Ga}(a, b)$, then $f(x|a, b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{-x/b}$

Let $Y_t \in \mathbb{R}^n \times \mathbb{R}^n$ be a $\text{WAR}(p)$ process:

$$E [Y_t | \mathbb{I}_{t-1}] = \sum_{j=1}^p M_j Y_{t-j} M_j' + K \Sigma. \quad (51)$$

where \mathbb{I}_{t-1} is the information set available up to time $t - 1$.

Under stationary conditions, the unconditional mean of the process, $E [Y_t]$, is obtained using the law of iterated expected values:

$$E [Y_t] = E [E [Y_t | \mathbb{I}_{t-1}]] = \sum_{j=1}^p M_j E [Y_{t-j}] M_j' + K \Sigma \quad (52)$$

As the unconditional distribution of any $\text{WAR}(p)$ process is a centered Wishart distribution, applying the definition of centered Wishart distribution, we can write:

$$Y_t = \sum_{k=1}^K z_{k,t} z_{k,t}', \quad (53)$$

where $z_{t,k} \stackrel{i.i.d}{\sim} N(0, \Sigma(\infty))$.

From (53) we have that

$$E [Y_t] = \sum_{k=1}^K E [z_{k,t} z_{k,t}'] \quad (54)$$

$$= K V [z_{k,t}] \quad (55)$$

$$= K \Sigma(\infty). \quad (56)$$

Combining this result with (53) and defining $\Sigma^*(\infty) = K \Sigma(\infty)$ and $\Sigma^* = K \Sigma$ we get

$$\Sigma^*(\infty) = \sum_{j=1}^p M_j \Sigma^*(\infty) M_j' + \Sigma^* \quad (57)$$

From (48) we know that, for any given vector $\omega \in \mathbb{R}^n$

$$\omega' Y_t \omega \sim Ga(K/2, 2\omega' \Sigma(\infty) \omega). \quad (58)$$

Knowing the variance of a gamma-distributed random variable, we have

$$V [\omega' Y_t \omega] = \frac{K}{2} (2\omega' \Sigma(\infty) \omega)^2. \quad (59)$$

$\Sigma(\infty)$ is not observable, but given the estimated matrices \hat{M}_j , $j = 1, \dots, p$ and $\hat{\Sigma}^*$ we can recover $\hat{\Sigma}^*(\infty)$ that satisfies (57). Thus:

$$V [\omega' Y_t \omega] = \frac{K}{2} \left(2\omega' \frac{\hat{\Sigma}^*(\infty)}{K} \omega \right)^2 \quad (60)$$

$$= \frac{2}{K} \left(\omega' \hat{\Sigma}^*(\infty) \omega \right)^2. \quad (61)$$

Therefore the estimated degrees of freedom are

$$\hat{K} = \frac{2(\omega' \hat{\Sigma}^*(\infty) \omega)^2}{V [\omega' Y_t \omega]} \quad (62)$$

Table 10: VaR failure rate and Berkowitz (2001) test's p -value

		10%	5%	1%
full WAR(1)	N	0.1072 (0.6608)	0.0490 (0.7038)	0.0230 (0.8174)
	t	0.1041 (0.8137)	0.0490 (0.8508)	0.0230 (0.9446)
block diagonal WAR(1)	N	0.1041 (0.6441)	0.0521 (0.6865)	0.0245 (0.7984)
	t	0.1026 (0.7836)	0.0505 (0.8209)	0.0245 (0.9157)
restr. block diag. WAR(1)	N	0.1057 (0.6677)	0.0536 (0.7093)	0.0245 (0.8184)
	t	0.1041 (0.7991)	0.0521 (0.8341)	0.0245 (0.9220)
diagonal WAR(1)	N	0.1057 (0.6705)	0.0521 (0.7121)	0.0245 (0.8208)
	t	0.1041 (0.7988)	0.0505 (0.8337)	0.0245 (0.9214)
restr. diag. WAR(1)	N	0.1057 (0.6664)	0.0521 (0.7080)	0.0245 (0.8168)
	t	0.1041 (0.7980)	0.0505 (0.8329)	0.0245 (0.9208)
full HAR-WAR	N	0.1103 (0.0697)	0.0658 (0.0800)	0.0291 (0.1112)
	t	0.1087 (0.1393)	0.0658 (0.1574)	0.0260 (0.2104)
block diag. HAR-WAR	N	0.1133 (0.2612)	0.0536 (0.2898)	0.0260 (0.3711)
	t	0.1133 (0.3929)	0.0536 (0.4292)	0.0245 (0.5297)
restr. block diag. HAR-WAR	N	0.1149 (0.3722)	0.0551 (0.4076)	0.0245 (0.5057)
	t	0.1149 (0.4991)	0.0551 (0.5392)	0.0245 (0.6474)
diagonal HAR-WAR	N	0.1118 (0.4440)	0.0475 (0.4831)	0.0245 (0.5909)
	t	0.1103 (0.5716)	0.0475 (0.6141)	0.0245 (0.7281)
restr. diag. HAR-WAR	N	0.1133 (0.3707)	0.0475 (0.4065)	0.0245 (0.5063)
	t	0.1133 (0.5333)	0.0475 (0.5751)	0.0245 (0.6881)