FLEXIBLE MULTIVARIATE GARCH MODELING WITH AN APPLICATION TO INTERNATIONAL STOCK MARKETS

Olivier Ledoit, Pedro Santa-Clara, and Michael Wolf*

Abstract—This paper offers a new approach to estimating time-varying covariance matrices in the framework of the diagonal-vech version of the multivariate GARCH(1,1) model. Our method is numerically feasible for large-scale problems, produces positive semidefinite conditional covariance matrices, and does not impose unrealistic a priori restrictions. We provide an empirical application in the context of international stock markets, comparing the new estimator with a number of existing ones.

I. Introduction

THE goal of this paper is to estimate conditional covariance matrices. Since the covariance matrix is an essential ingredient in risk management, portfolio selection, and tests of asset pricing models, this is a very important problem in practice. Estimating conditional covariance matrices is a multivariate extension of the simpler problem of estimating conditional variances. In the univariate case, many methods are available, ranging from the simple rolling-window estimation method to the sophisticated models of latent stochastic volatility. The most popular method, however, for estimating conditional variances is the GARCH(1,1) model. We do not claim that it is the best method, because a method that is uniformly better than the others does not seem to exist. On the other hand, many studies have shown that the univariate GARCH(1,1) gives reasonable results, and it can be safely assumed that it will remain in use for some time to come; for example, see Andersen, Bollerslev, and Lange (1999) and Lee and Saltoğlu (2001). For these reasons, multivariate extensions of the univariate GARCH(1,1) model have long been of interest.

The most general multivariate GARCH-style model commonly considered is defined by

$$\mathrm{E}[x_{i,t}|\Omega_{t-1}] = 0, \tag{1}$$

$$\begin{aligned}\mathrm{Cov}[x_{i,t}, x_{j,t}|\Omega_{t-1}] &= h_{ij,t} \\ &= c_{ij} + a_{ij}x_{i,t-1}x_{j,t-1} + b_{ij}h_{ij,t-1},\end{aligned} \tag{2}$$

where Ω_{t-1} denotes the conditioning information set available at time $t-1$, and $x_{i,t}$ denotes the realization of the i^{th} variable ($i = 1, \ldots, N$) at time t. The parameter values satisfy $a_{ij}, b_{ij} \geq 0\ \forall i, j = 1, \ldots, N$, and $c_{ii} > 0\ \forall i = 1, \ldots, N$. Equation (2) is known as the *diagonal-vech* model. It assumes that the conditional covariance of variables x_i and x_j depends on its lagged value and on past realizations of the product x_ix_j only (Bollerslev, Engle, and Wooldridge, 1988). Also, equation (1) assumes that the variables have zero conditional mean, which can always be justified by taking them to be residuals coming from some regression model. Although more general models can be thought of, they typically involve too many parameters to be of practical interest.

The natural way to estimate the conditional covariance matrix is to compute the (quasi) maximum likelihood estimates of the parameters c_{ij}, a_{ij}, and b_{ij} from observations of all the variables in the vector x. Unfortunately, this is not computationally feasible for matrices of dimension $N > 5$ (Ding and Engle, 1994): there are too many parameters, $3N(N+1)/2$, and they interact in a way that is too intricate for existing optimization algorithms to converge. Another problem is that the estimation of the general diagonal-vech model does not necessarily yield conditional covariance matrices that are positive semidefinite.

The existing literature avoids these difficulties by imposing additional structure on the problem. For example, Ding and Engle (1994) give a sequence of 20 nested models that are particular special cases of equation (2), by specifying, for example, that the conditional correlations should be constant, or that there is some factor structure in the conditional covariance matrix. Additional models can be found in Engle and Kroner (1995), Engle and Mezrich (1996), and Engle (2002), among others. Apart from being tractable, these models typically also ensure that the resulting conditional covariance matrices are positive semidefinite.

Although it can be useful to impose *sensible* restrictions for forecasting purposes, there is also the danger of employing restrictions that are strongly violated by the data. We therefore seek a way to estimate the unrestricted model, to later compare it against more restrictive models using data from international stock markets.

Our basic idea proceeds in two steps. The first step is to obtain each set of coefficient estimates $\hat{c}_{ij}$, $\hat{a}_{ij}$, and $\hat{b}_{ij}$ separately for every (i, j). This can be achieved simply by estimating a two-dimensional or one-dimensional GARCH(1,1) model (for $i \neq j$ or $i = j$ respectively), which is computationally feasible using a traditional method such as maximum likelihood. We bring together the outputs of these separate estimation procedures into matrices $\hat{C} = [\hat{c}_{ij}]_{i,j=1,\ldots,N}$, $\hat{A} = [\hat{a}_{ij}]_{i,j=1,\ldots,N}$, and $\hat{B} = [\hat{b}_{ij}]_{i,j=1,\ldots,N}$. However, the coefficient matrices $\hat{C}$, $\hat{A}$, and $\hat{B}$ are generally incompatible with each other in the sense that they yield conditional covariance

Received for publication June 29, 2000. Revision accepted for publication July 29, 2002.

* Credit Suisse First Boston; The Anderson School, UCLA; and Universitat Pompeu Fabra, respectively.

The third author's research was supported by DGES grant BEC2001-1270. This paper was finished while he was visiting Bocconi University, whose hospitality is gratefully acknowledged.

We thank Frank de Jong, Bruno Gérard, Hans Mikkelsen, and Robert Whitelaw for comments. We are grateful to the editor John Campbell and two anonymous referees for suggestions and constructive criticisms that have led to improvement of the paper. Ilya Sharapov was extremely helpful with the numerical computations. All errors are our own.

matrices that are not positive semidefinite. Therefore, our second step is to transform the estimated parameter matrices $\hat{C}$, $\hat{A}$, and $\hat{B}$ in such a way that they yield conditional covariance matrices that are guaranteed to be positive semidefinite, where the transformation is chosen to be the least disruptive possible (according to some metric). In addition, we obtain GARCH(1,1) parameters that correspond to covariance-stationary processes, in contrast to the implicit model behind the exponential smoothing scheme that is quite popular for large-dimensional covariance matrices and used by RiskMetrics, for example.

In summary, the main advantage of our estimation method is that it is the first to allow estimation of the full-blown diagonal-vech model for dimensions larger than $N = 5$ without imposing any a priori restrictions. Our conditional covariance matrices are only forced to be positive semidefinite, but they generally turn out to be positive definite and well conditioned, which is a characteristic that (purely on economic grounds) we would expect from the true covariance matrix (as long as we consider a menu of nonredundant assets). An additional advantage is the reduced computational cost compared to traditional multivariate models; see section IIIC.

The paper proceeds as follows. Section II develops the new estimation method. Section III gives an empirical application to international stock markets. Section IV concludes. An appendix highlights some computational issues.

II. Estimation Method

It is important to understand precisely why it is so difficult to estimate the unrestricted diagonal-vech model in equation (2) by maximum likelihood. Although there are many parameters, $3N(N + 1)/2$, this cannot be the only source of the problem. The number of parameters in the *unconditional* covariance matrix is of the same order of magnitude, $N(N + 1)/2$, and estimating the unconditional covariance matrix by the sample covariance matrix is computationally trivial. Computing the sample covariance matrix is easy because it can be done in a *decentralized* fashion: for every variable, compute its sample variance (this is a univariate problem) and insert it into the diagonal; for every pair of variables, compute their sample covariance (this is a bivariate problem) and insert it at the appropriate place off the diagonal. Thus, a large-sample covariance matrix can be constructed by solving $N(N + 1)/2$ univariate or bivariate estimation problems.

Could the same decentralized process be used to compute the diagonal-vech estimator? Not directly. The crucial problem is the compatibility of the parameters that come out of all the univariate or bivariate estimations. The compatibility constraint is that the resulting covariance matrices must be positive semidefinite. To pursue our analogy, in the case of the sample covariance matrix, the mathematical form of the estimators guarantees that the sample covariance matrix constructed by putting together the individual sample variances and sample covariances is positive semidefinite. On the other hand, for the diagonal-vech model, positive definiteness is not automatic.

The rest of this section develops an approach to deal with these problems.

A. Decentralized Estimation of Multivariate GARCH(1,1)

Consider what happens when we try to decentralize the estimation process for the diagonal-vech model. As we said, this constitutes the first step of our estimation procedure. This step itself can be divided into two substeps, corresponding to the estimation of the diagonal and the off-diagonal coefficients, respectively.

Diagonal Coefficients: We estimate a univariate GARCH(1,1) process for every one of the variables by *quasi* maximum likelihood, and we get consistent estimators $\hat{c}_{ii}$, $\hat{a}_{ii}$, and $\hat{b}_{ii}$. Separately for each $i = 1, \ldots, N$, we solve the quasi-likelihood maximization program, assuming conditional normality:

$$\max_{c_{ii},a_{ii}+b_{ii}\leq 1-\epsilon} \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi h_{ii,t}}} e^{-x_{it}^2/(2h_{ii,t})} \quad \text{s.t.} \quad h_{ii,t} = c_{ii} + a_{ii}x_{i,t-1}^2 + b_{ii}h_{ii,t-1}. \tag{3}$$

Here, ϵ is a small number[1] to ensure $a_{ii} + b_{ii} < 1$; see the discussion following equations (9)–(11). For each i, we have a simple univariate GARCH(1,1) estimation problem, which many commercial packages solve quickly. The estimator is in general not efficient, as the conditional distribution may be different from normal, but it is consistent (see, for example, Campbell, Lo, and MacKinlay, 1997, section 12.2).

Off-Diagonal Coefficients: From the above, we get parameter estimates $\hat{c}_{ii}$, $\hat{a}_{ii}$, and $\hat{b}_{ii}$. We can use them to construct conditional variance estimates $\hat{h}_{iit}$. In the second stage, we use these estimates to specify quasi-likelihood functions for the off-diagonal elements. Separately for each $i = 1, \ldots, N$ and $j = i + 1, \ldots, N$, we solve

$$\max_{c_{ij},a_{ij},b_{ij}} \prod_{t=1}^{T} \frac{1}{2\pi\sqrt{\det(H_{ij,t})}} e^{-X_{ij,t}' H_{ij,t}^{-1} X_{ij,t}/2} \tag{4}$$

$$\text{s.t.} \quad X_{ij,t} = \begin{bmatrix} x_{i,t} \\ x_{j,t} \end{bmatrix}, \quad H_{ij,t} = \begin{bmatrix} \hat{h}_{ii,t} & h_{ij,t} \\ h_{ij,t} & \hat{h}_{jj,t} \end{bmatrix}, \quad \text{and} \quad h_{ij,t} = c_{ij} + a_{ij}x_{i,t-1}x_{j,t-1} + b_{ij}h_{ij,t-1}. \tag{5}$$

This quasi likelihood is obtained by restricting attention to the 2×2 submatrix of variables x_i and x_j, fixing the

[1] In this work, we have used $\epsilon = 10^{-3}$; this value could be changed depending on the context.

conditional variances at their first-stage values $\hat{h}_{ii}$ and $\hat{h}_{jj}$, and assuming normality.[2] As on the diagonal, quasi-likelihood theorems ensure consistency. The problem (4)–(5) is easy to solve using standard optimization algorithms, since there are only three free parameters. The positive definiteness of the conditional covariance submatrix $H_{ij,t}$ is ensured by imposing the following bounds in the estimation process: $|c_{ij}| \leq (\hat{c}_{ii}\hat{c}_{jj})^{1/2}$, $0 \leq a_{ij} \leq (\hat{a}_{ii}\hat{a}_{jj})^{1/2}$, and $0 \leq b_{ij} \leq (\hat{b}_{ii}\hat{b}_{jj})^{1/2}$, as Ding and Engle (1994) show.

B. *Compatibility Constraints*

As noted before, the estimators of the coefficients c_{ij}, a_{ij}, and b_{ij} obtained separately for every (i, j) in section IIA are not compatible with one another, in the sense that the forecasted covariance matrix may not be positive semidefinite. This subsection analyzes the mathematical relations that they must satisfy in order to become compatible.

Positive Semidefinite Conditional Covariance Matrix: Following the notation of Ding and Engle (1994), let $C = [c_{ij}]_{i,j=1,\ldots,N}$, $A = [a_{ij}]_{i,j=1,\ldots,N}$, and $B = [b_{ij}]_{i,j=1,\ldots,N}$ denote matrices containing the parameters of the model. Let $H_t = [h_{ijt}]_{i,j=1,\ldots,N}$ denote the conditional covariance matrix at time t. Denote the matrix of cross-products of variables observed at time t by $\Sigma_t = [x_{i,t}x_{j,t}]_{i,j=1,\ldots,N}$. Then equation (2) can be rewritten as

$$H_t = C + A * \Sigma_{t-1} + B * H_{t-1}, \tag{6}$$

where the symbol $*$ denotes the Hadamard product. The Hadamard product of two matrices $U = [u_{ij}]_{i,j=1,\ldots,N}$ and $V = [v_{ij}]_{i,j=1,\ldots,N}$ is defined as the elementwise product $U * V = [u_{ij}v_{ij}]_{i,j=1,\ldots,N}$. Similarly, let $\div$ denote elementwise division: $U \div V = [u_{ij}/v_{ij}]_{i,j=1,\ldots,N}$ and let $\wedge$ denote elementwise exponentiation: $U^{\wedge p} = [u_{ij}^p]_{i,j=1,\ldots,N}$.

Ding and Engle (1994) show that a sufficient condition to guarantee that the conditional covariance matrix H_t is positive semidefinite almost surely (a.s.) is that C, A, and B are positive semidefinite. We derive a somewhat weaker sufficient condition.

Proposition 1. If $C \div (1 - B)$, A, and B are positive semidefinite, then the conditional covariance matrix is positive semidefinite.

Proof of Proposition 1. Substituting equation (6) into itself recursively yields

$$\begin{aligned} H_t &= \sum_{k=0}^{\infty} B^{\wedge k} * C + \sum_{k=0}^{\infty} B^{\wedge k} * A * \Sigma_{t-k-1} \\ &= C \div (1 - B) + \sum_{k=0}^{\infty} B^{\wedge k} * A * \Sigma_{t-k-1}. \end{aligned} \tag{7}$$

The Hadamard product of two positive semidefinite matrices is positive semidefinite; for example, see Styan (1973). In addition, the sum of two positive semidefinite matrices is positive semidefinite. Finally, the matrix of cross products of realizations Σ_{t-k-1} is positive semidefinite a.s. by construction. Therefore, inspection of equation (7) shows that, under the conditions stated in proposition 1, the conditional covariance matrix H_t is guaranteed to be positive semidefinite a.s. □

A simple example for which our condition holds but not the one in Ding and Engle (1994) is given by

$$B = \begin{bmatrix} 0.9 & 0.84 \\ 0.84 & 0.8 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1.0 & 1.1 \\ 1.1 & 1.0 \end{bmatrix}.$$

It is easy to check that here $C \div (1 - B)$ is positive semidefinite but C is not. While this example may or may not be economically relevant, it illustrates that the sufficient condition of Ding and Engle (1994) can indeed be weakened.

Proposition 2. If:

- the conditional multivariate distribution of the vector x_t is continuous with unbounded support for all t;
- $\forall i, \forall j, b_{ij} < 1$;
- the conditional covariance matrix H_t is positive semidefinite a.s. for all t,

then it is necessary that the parameter matrix $C \div (1 - B)$ be positive semidefinite.

Proof of Proposition 2. We make a proof by contradiction. Suppose that $C \div (1 - B)$ has at least one negative eigenvalue $\lambda < 0$. Expand the conditional covariance matrix as:

$$\begin{aligned} H_t = C \div (1 - B) &+ \sum_{k=0}^{t-1} B^{\wedge k} * A * \Sigma_{t-k-1} \\ &+ B^{\wedge t} * [H_0 - C \div (1 - B)]. \end{aligned} \tag{8}$$

Let maxeig $(\cdot)$ denote the maximum eigenvalue of a matrix. Since all the elements of B have absolute value strictly below 1, we have

[2] We do not have to impose the constraint $a_{ij} + b_{ij} < 1$ at this stage; see the discussion following equations (9)–(11).

$$B^{\wedge t} \to 0 \quad \text{as} \quad t \to \infty,$$

$$B^{\wedge t} * [H_0 - C \div (1 - B)] \to 0 \quad \text{as} \quad t \to \infty,$$

$$\text{maxeig}\,(B^{\wedge t} * [H_0 - C \div (1 - B)]) \to 0 \quad \text{as} \quad t \to \infty.$$

Therefore, there exists a T large enough so that maxeig $(B^{\wedge T} * [H_0 - C \div (1 - B)]) < -\lambda/2$.

The matrix $A * (xx')$ goes to the null matrix as the vector x goes to the null vector. Hence, there exists a neighborhood N_0 of the null vector such that $\forall x \in N_0$, maxeig $(A * (xx')) < -\lambda/(2T)$. Since all the elements of B have absolute value strictly below 1, that implies: $\forall x \in N_0$, $\forall k = 0, \ldots, T - 1$, maxeig $(B^{\wedge k} * A * (xx')) < -\lambda/(2T)$. In the event that $\forall k = 0, \ldots, T - 1$, $x_k \in N_0$, we have

$$\text{maxeig}\left(\sum_{k=0}^{T-1} B^{\wedge k} * A * \Sigma_{t-k-1} + B^{\wedge T} * [H_0 - C \div (1 - B)]\right) < -\lambda.$$

Therefore, by equation (8) the conditional covariance matrix H_T is not positive semidefinite if this event occurs. Since the conditional multivariate distribution of the vector x_t is continuous with unbounded support for all t, the event has a positive probability of happening, which leads to a contradiction. This proves that $C \div (1 - B)$ cannot have any strictly negative eigenvalue. □

The assumption that the elements of B have absolute value strictly below 1 is innocuous, because it comes from the variance and covariance stationarity of the multivariate GARCH(1,1) process. Similarly, we can prove that the positive semidefiniteness of the parameter matrix A is also a necessary condition.

Proposition 3. If the conditional multivariate distribution of the vector x_t is continuous with unbounded support and H_t is positive semidefinite a.s. for all t, then it is necessary that the parameter matrix A be positive semidefinite.

Proof of Proposition 3. Again we make the proof by contradiction. Let mineig $(\cdot)$ denote the smallest eigenvalue of a matrix. Suppose that A has at least one negative eigenvalue, that is, mineig $(A) = \lambda < 0$. Recall the multivariate GARCH(1,1) recursion $H_t = C + A * (x_{t-1}x'_{t-1}) + B * H_{t-1}$. We work conditionally on H_{t-1}. If x_{t-1} is equal to the unit vector, then mineig $(A * (x_{t-1}x'_{t-1})) = \lambda$. Therefore, by continuity, there exists a neighborhood N_1 of the unit vector such that $\forall x_{t-1} \in N_1$, mineig $[A * (x_{t-1}x'_{t-1})] < \lambda/2$. Let $\mu =$ maxeig $(C + B * H_{t-1})$. In the event that $x_{t-1} = \sqrt{-2\mu/\lambda}\, z$ for some $z \in N_1$, we have mineig $(C + A * (x_{t-1}x'_{t-1}) + B * H_{t-1}) < 0$. Therefore, by the recursion formula, the conditional covariance matrix H_t is not positive semidefinite if this event occurs. Because the conditional multivariate distribution of the vector x_{t-1} is continuous with unbounded support, the event has a positive probability of happening, which leads to a contradiction. This proves that A cannot have any strictly negative eigenvalue. □

For B, the situation is less clear. Technically speaking, the only necessary condition is that $B^{\wedge k} * A$ is positive semidefinite for all $k \geq 1$. It is possible to construct a counterexample with a matrix B that satisfies this necessary condition but is not positive semidefinite. In the univariate case, just take $A = 0$, $B = -1$. This counterexample is mathematically correct but economically degenerate, and we have not been able to construct a more realistic one. In general, after extensive numerical experiments, our overall feeling is that the pairs (A, B) that satisfy the necessary condition and where B is not positive semidefinite are extremely rare, and can perhaps be ruled out on economic grounds. However, we have not been able to prove any formal result along these lines. Hence, we will make the positive definiteness of B an assumption rather than a conclusion.

Assumption 1. The true coefficient matrix B in the multivariate GARCH(1,1) model is positive semidefinite.

Covariance Stationarity: Another common concern in the application of GARCH models to financial returns is that the fitted model be covariance stationary. Hence, we want to make sure that $a_{ij} + b_{ij} < 1$ $\forall i, j = 1, \ldots, N$. The following proposition shows that it is only necessary to verify this on the diagonal, as long as the coefficient matrices are positive semidefinite.

Proposition 4. If A and B are positive semidefinite and if

$$a_{ii} + b_{ii} < 1 \qquad \forall\, i = 1, \ldots, N,$$

then

$$a_{ij} + b_{ij} < 1 \qquad \forall\, i, j = 1, \ldots, N.$$

Proof of Proposition 4.

$$a_{ij} + b_{ij} = |a_{ij}| + |b_{ij}| \leq \sqrt{a_{ii}a_{jj}} + \sqrt{b_{ii}b_{jj}} \leq \sqrt{a_{ii} + b_{ii}}\sqrt{a_{jj} + b_{jj}},$$

where the second to last inequality is a consequence of A and B being positive semidefinite, and the last inequality is a consequence of the Hölder inequality. □

In conclusion, the definitive version of our set of compatibility constraints is: $C \div (1 - B)$, A and B positive semidefinite, and $a_i + b_i < 1$ $\forall i = 1, \ldots, N$.

C. Transformation of Coefficient Matrices

We now build the matrices $\hat{C} = [\hat{c}_{ij}]_{i,j=1,\ldots,N}$, $\hat{A} = [\hat{a}_{ij}]_{i,j=1,\ldots,N}$, and $\hat{B} = [\hat{b}_{ij}]_{i,j=1,\ldots,N}$ by placing the esti-

mators from the first part of section IIA on the diagonal and placing the estimators from the second part on the appropriate positions off the diagonal. For convenience, we also define $D = C \div (1 - B)$ and $\hat{D} = \hat{C} \div (1 - \hat{B})$. Note that $\hat{D}$ thus defined *is* the quasi maximum likelihood estimator of D.

$\hat{D}$, $\hat{A}$, and $\hat{B}$ are consistent estimators of D, A, and B respectively, but they are generally not positive semidefinite. To be precise, $\hat{D}$, $\hat{A}$, and $\hat{B}$ converge to positive semidefinite matrices (under assumption 1), but in a finite sample there is no guarantee that they are positive semidefinite. Practically speaking, our experience has been that, for reasonable sample sizes, finding positive semidefinite estimates is extremely rare. In other words, this decentralized procedure yields parameters that are not compatible with one another. This is why it has not been used in the existing literature, and why further restrictions are commonly imposed on the diagonal-vech model.

Our central innovation is to transform the estimators $\hat{D}$, $\hat{A}$, and $\hat{B}$ to positive semidefinite matrices $\tilde{D}$, $\tilde{A}$, and $\tilde{B}$, which we then take to be the estimates of D, A, and B. These matrices $\tilde{D}$, $\tilde{A}$, and $\tilde{B}$ are chosen to be the closest to $\hat{D}$, $\hat{A}$, and $\hat{B}$, respectively, according to a certain norm, but forcing the diagonal parameters obtained from univariate GARCH(1,1) estimation to remain unchanged. This can be formalized as:

$$\min_{\tilde{D}} \|\tilde{D} - \hat{D}\| \quad \text{s.t.} \quad \tilde{D} \text{ is positive semidefinite} \quad \text{and } \tilde{d}_{ii} = \hat{d}_{ii} \ \ \forall\, i = 1, \ldots, N, \tag{9}$$

$$\min_{\tilde{A}} \|\tilde{A} - \hat{A}\| \quad \text{s.t.} \quad \tilde{A} \text{ is positive semidefinite} \quad \text{and } \tilde{a}_{ii} = \hat{a}_{ii} \ \ \forall\, i = 1, \ldots, N, \tag{10}$$

$$\min_{\tilde{B}} \|\tilde{B} - \hat{B}\| \quad \text{s.t.} \quad \tilde{B} \text{ is positive semidefinite} \quad \text{and } \tilde{b}_{ii} = \hat{b}_{ii} \ \ \forall\, i = 1, \ldots, N. \tag{11}$$

Once we have $\tilde{D}$ and $\tilde{B}$, we can recalculate $\tilde{C} = \tilde{D} * (1 - \tilde{B})$.

One appealing property of this transformation is that it guarantees that the multivariate GARCH(1,1) process will not explode, that is, $|\tilde{a}_{ij} + \tilde{b}_{ij}| < 1 \ \forall i, j = 1, \ldots, N$. As shown in proposition 4, it is sufficient to check the diagonal, because the transformed matrices are by construction positive semidefinite. Because we preserve the diagonal elements of $\hat{A}$ and $\hat{B}$, which come from covariance-stationary univariate GARCH(1,1) processes due to the constraint imposed in equation (3), this condition is automatically verified. This also explains why no similar constraint has to be imposed in equation (4).

Another useful property is that the conditional covariance matrix H_t is in general invertible. The parameter matrices $\tilde{D}$, $\tilde{A}$, and $\tilde{B}$ are not invertible, because, by construction, they lie on the frontier of the convex set of positive semidefinite matrices, and only the interior of this set is made of invertible matrices. Nonetheless, combining $\tilde{D}$, $\tilde{A}$, and $\tilde{B}$ according to equation (6) is sufficient to pull the resulting H_t into the interior of this set, thereby making it invertible, except in some degenerate special cases.

In order to measure closeness, different matrix norms are possible. We choose the Frobenius norm $\|U\|_F = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} u_{ij}^2}$, because it is intrinsically compatible with the usual quadratic formulation of consistency results. Unfortunately, there does not appear to be any closed-form solution for the minimization problems (9)–(11). We use a numerical algorithm due to Sharapov (1997, Section 3.2). For convenience, this algorithm is explained in the appendix.

It is important to understand that this transformation makes no difference asymptotically, since the limits of $\hat{D}$, $\hat{A}$, and $\hat{B}$ are positive semidefinite (under assumption 1). Therefore, the consistency of $\hat{D}$, $\hat{A}$, and $\hat{B}$ guarantees that of $\tilde{D}$, $\tilde{A}$, and $\tilde{B}$.

A disadvantage of our method is that it does not yield straightforward standard errors of the parameter estimates, as the transformation of the first-step matrices to positive semidefinite matrices is nonlinear and not available in closed form. At the expense of greater computational cost, however, standard errors can be obtained by using an appropriate bootstrap method. A natural choice would be a semiparametric bootstrap based on the fitted model. It generates bootstrap data $x_t^*, \ldots, x_T^*$ in the following way:

$$h_{ij,t}^* = \tilde{c}_{ij} + \tilde{a}_{ij} x_{i,t}^* x_{j,t-1}^* + \tilde{b}_{ij} h_{ij,t-1}^*, \tag{12}$$

$$x_t^* = (H_t^*)^{1/2} \epsilon_t^*. \tag{13}$$

Here, the ϵ_t^* are resampled from the fitted standardized residuals, properly transformed to have sample mean equal to zero and sample covariance matrix equal to the identity. This is done as follows:

- Compute $\hat{\epsilon}_t = \hat{H}_t^{-1/2} x_t$, $t = 1, \ldots, T$.
- Denote by $\bar{\hat{\epsilon}}$ the sample mean of the $\hat{\epsilon}_t$.
- Denote by $\Sigma_{\hat{\epsilon}}$ the sample covariance matrix of the $\hat{\epsilon}_t$.
- Let $\tilde{\epsilon}_t = \Sigma_{\hat{\epsilon}}^{-1/2} (\hat{\epsilon}_t - \bar{\hat{\epsilon}})$, $t = 1, \ldots, T$.
- The ϵ_t^* are then resampled (with replacement) from the $\tilde{\epsilon}_t$.

As a starting value for H_1^* in equation (13) one can use the sample covariance matrix, for example. (To make negligible the choice of the starting value, one can actually start the generation of bootstrapped data at time $t = -M$, with $M = 100$ say, and then discard the first $M + 1$ values.) The following algorithm describes how to compute bootstrap standard errors for the individual parameter estimates $\tilde{c}_{ij}$, $\tilde{a}_{ij}$, and $\tilde{b}_{ij}$. Choosing $K \geq 100$ in this algorithm should be sufficient in practice; see Efron and Tibshirani (1993).

TABLE 1.—SUMMARY STATISTICS OF LOG RETURNS

Country	U.S.	U.K.	France	Germany	Japan	Canada	Switzerland
Mean	15.16	17.85	15.91	12.14	11.69	11.27	15.23
SD	15.10	19.71	20.57	17.66	20.44	15.62	16.48
Skewness	−0.60	0.25	−0.49	−0.31	0.11	−0.47	−0.60
Kurtosis	6.82	9.01	5.29	4.72	4.80	6.54	6.95

This table presents the summary statistics for the weekly percentage log returns of seven different stock markets. The local currency returns were transformed into U.S. dollar returns by the appropriate exchange rate; they correspond to the returns obtained by a U.S. investor who does not hedge currency risk. The sample includes 1,356 observations from January 1, 1975 to December 31, 2000, obtained from Datastream. The numbers for the mean and the standard deviation are annualized.

Algorithm 1 (Bootstrap Standard Errors).

1. For $k = 1, \ldots, K$, generate bootstrap data $x^*_{1,k}, \ldots, x^*_{T,k}$ as described in equations (12)–(13).
2. Compute the estimators $\tilde{C}$, $\tilde{A}$, and $\tilde{B}$ on each data set to obtain bootstrap estimates $\tilde{C}^*_k$, $\tilde{A}^*_k$, and $\tilde{B}^*_k$ for $k = 1, \ldots, K$.
3. The sample standard deviations of $\tilde{c}^*_{ij,k}$, $\tilde{a}^*_{ij,k}$, and $\tilde{b}^*_{ij,k}$, $k = 1, \ldots, K$, are the respective bootstrap standard errors of $\tilde{c}_{ij}$, $\tilde{a}_{ij}$, and $\tilde{b}_{ij}$.

III. Application to International Stock Markets

In this section, we compare the performance of several multivariate GARCH(1,1) covariance estimators using historical stock return data. Additionally, we compare the multivariate GARCH(1,1) estimators with other, less sophisticated estimators. (Note that a less sophisticated estimator is not necessarily an inferior estimator.) The multivariate GARCH(1,1) estimator that we developed in the previous section will be called FlexM (for flexible multivariate GARCH) in the remainder of the paper.

A. Data

We use weekly stock market data from the United States, the United Kingdom, France, Germany, Japan, Canada, and Switzerland, as captured by the major, broad market indices in each of these countries. The sample goes from January 1, 1975 to December 31, 2000, yielding 1,356 weekly returns (from the close of Wednesday to the close of next Wednesday).

We take the point of view of a U.S. investor who does not hedge any currency risk. For each country, we thus convert weekly index prices to U.S. dollars (using the exchange rate of the appropriate date) and then compute log returns. To ease interpretation, the log returns are multiplied by 100, so they can be read as percentage returns. All data were obtained from Datastream. Summary statistics of the return data are presented in table 1; note that the numbers for the mean and the standard deviation have been annualized.

B. Competing Estimators

For comparison, we include two popular multivariate GARCH(1,1) estimators and two other widely used estimators of conditional covariance matrices in the study.

Constant-Conditional-Correlation GARCH: Bollerslev (1990) suggested a multivariate GARCH(1,1) model where the conditional correlations are constant over time. To be more specific, each conditional variance $h_{ii,t}$ is modeled by a separate univariate GARCH(1,1) model with parameters c_{ii}, a_{ii}, and b_{ii}, respectively, and the conditional covariance between variables x_i and x_j at time t is given by $h_{ij,t} = \rho_{ij}\sqrt{h_{ii,t}h_{jj,t}}$. Hence, there are a total of $N(N + 5)/2$ free parameters. This model gives positive definite and stationary conditional covariance matrices provided that the ρ_{ij} make up a well-defined correlation matrix and the parameters c_{ii}, a_{ii}, and b_{ii} are all nonnegative satisfying $a_{ii} + b_{ii} < 1\ \forall i = 1, \ldots, N$. The estimation is done by maximizing the quasi likelihood, assuming conditional normality. In the remainder of the paper, this estimator will be called CCC.

A problem with this model is the assumption of a constant conditional correlation, which conceivably will not always hold.

Diagonal BEKK GARCH: Engle and Kroner (1995) proposed a class of multivariate GARCH models that are guaranteed to produce positive definite conditional covariance matrices. In its full generality, the corresponding GARCH(1,1) model includes all positive definite diagonal-vech models and suffers from its intractability for higher dimensions. The model most commonly used in practice is the more restrictive first-order diagonal BEKK GARCH(1,1) model given by

$$H_t = G'G + E'x_{t-1}x'_{t-1}E + F'H_{t-1}F,$$

where H_t denotes the conditional covariance matrix at time t, x_t denotes the (column) vector of residuals at time t, G is a triangular matrix, and E and F are diagonal matrices. Again, there are a total of $N(N + 5)/2$ free parameters, and the conditional covariance matrices (which are positive semidefinite by construction) are guaranteed to be stationary if $e^2_{ii} + f^2_{ii} < 1\ \forall i = 1, \ldots, N$. The estimation is done by maximizing the quasi likelihood, assuming conditional normality. In the remainder of the paper, this estimator will be called BEKK.

A problem of this model, in the notation of the general diagonal-vech model, is the implied constraints $a_{ij} = \sqrt{a_{ii}a_{jj}}$ and $b_{ij} = \sqrt{b_{ii}b_{jj}}$, which could easily be violated for certain data.

TABLE 2.—PARAMETER ESTIMATES OF THE FLEXM MODEL

U.S.	U.K.	France	Germany	Japan	Canada	Switzerland
			$\tilde{C}$			
0.1448						
0.1292	0.1641					
0.1560	0.2120	0.4410				
0.0744	0.1270	0.1752	0.1783			
0.0627	0.1119	0.1666	0.1160	0.1858		
0.1813	0.1840	0.1876	0.0950	0.0732	0.2929	
0.1495	0.2062	0.2798	0.2575	0.1958	0.1817	0.5271
			$\tilde{A}$			
0.0861						
0.0499	0.0710					
0.0533	0.0700	0.1037				
0.0641	0.0680	0.0940	0.0918			
0.0475	0.0405	0.0565	0.0600	0.0772		
0.0839	0.0474	0.0600	0.0768	0.0565	0.1162	
0.0517	0.0629	0.0876	0.0820	0.0521	0.0567	0.0783
			$\tilde{B}$			
0.8835						
0.8886	0.9066					
0.8606	0.8714	0.8426				
0.8802	0.8880	0.8595	0.8808			
0.8896	0.9018	0.8700	0.8873	0.9012		
0.8509	0.8616	0.8319	0.8497	0.8601	0.8233	
0.8482	0.8588	0.8292	0.8469	0.8573	0.8198	0.8179

This table presents the estimated parameters of the flexible multivariate (FlexM) GARCH(1,1) model based on the entire sample. The model is developed and described in section II. As the matrices are symmetric, only the lower triangular parts are displayed to enhance readability.

Rolling Window: The ever popular rolling-window estimator simply estimates the covariance matrix at time t, conditional on the information available at time $t - 1$, as the sample covariance matrix of the observations $\mathbf{x}_{t-k}, \ldots, \mathbf{x}_{t-1}$, where k is some predetermined integer. A common choice for weekly data is $k = 104$, which corresponds to a 2-year window. In the remainder of the paper, this model will be called Window.

Exponential Smoothing: The exponential smoothing estimator is given by

$$\hat{H}_t = \lambda \mathbf{x}_{t-1}\mathbf{x}'_{t-1} + (1 - \lambda)\hat{H}_{t-1},$$

where λ is a small, positive constant. Note that this prescription requires some suitable starting values. A common approach is to use the rolling-window estimator at time $k + 1$ for $\hat{H}_1, \ldots, \hat{H}_{k+1}$.

The exponential smoothing estimator corresponds to a multivariate integrated GARCH(1,1) model with a unique autoregressive coefficient $(1 - \lambda)$ and a unique moving-average coefficient (λ) for all variances and covariances. This specification is the basis of many risk measurement systems currently in use and, for example, is advocated by RiskMetrics. A commonly used value for λ is 0.06. In the remainder of the paper, this model will be called RiskM.

C. Estimation of the Models

When estimating the three multivariate GARCH(1,1) models from the entire set of 1,356 weekly data, the estimation of the FlexM model took less than three minutes, using a proprietary optimization routine in Matlab. In contrast, the estimation of both the CCC and the BEKK model took over one hour, using off-the-shelf optimization routines available in Matlab. Tables 2–5 present the estimates of the parameters of the various models. Table 3 displays bootstrap standard errors for the FlexM model.

However, we do not use these estimated models in our comparisons, as this strategy would focus on the in-sample performance of the various estimators. In-sample comparisons are not ideal for our purposes, for at least two reasons. First, they are too optimistic, because the entire sample is used in the fitting process before the fitted models are then applied in hindsight. Second, they tend to favor models with more degrees of freedom, so FlexM might have an unfair advantage.

We will therefore use out-of-sample comparisons in what follows. In general, the forecasts for time t are made using information available up to time $t - 1$ only. The parameter estimates of the multivariate GARCH(1,1) models are updated every four weeks to reduce the computational burden for BEKK and CCC. All forecasts start at time $t = 601$.

D. Forecast Criteria

The real test for a multivariate GARCH(1,1) model is to compare its estimated, or forecasted, conditional covariance matrix with the true, realized matrix. The latter is unobservable, but a proxy can be constructed. A common and successful approach, termed *integrated volatility,* is to use

TABLE 3.—STANDARD ERRORS OF THE FLEXM MODEL

U.S.	U.K.	France	Germany	Japan	Canada	Switzerland
			$\tilde{C}$			
0.0862						
0.0497	0.1286					
0.0554	0.0903	0.2221				
0.0354	0.0550	0.0675	0.0869			
0.0360	0.0575	0.0610	0.0456	0.0948		
0.0759	0.0755	0.0860	0.0516	0.0439	0.1406	
0.0957	0.1505	0.1898	0.1508	0.1512	0.1068	0.5059
			$\tilde{A}$			
0.0271						
0.0183	0.0220					
0.0186	0.0164	0.0294				
0.0169	0.0171	0.0206	0.0264			
0.0177	0.0146	0.0149	0.0138	0.0191		
0.0245	0.0179	0.0197	0.0183	0.0172	0.0369	
0.0212	0.0228	0.0255	0.0284	0.0177	0.0244	0.0458
			$\tilde{B}$			
0.0379						
0.0319	0.0337					
0.0419	0.0339	0.0461				
0.0362	0.0288	0.0313	0.0332			
0.0390	0.0319	0.0350	0.0283	0.0254		
0.0428	0.0420	0.0537	0.0460	0.0524	0.0551	
0.0890	0.0873	0.0848	0.0853	0.0890	0.0913	0.1320

This table presents bootstrap standard errors corresponding to the parameter estimates of table 2. The standard errors were computed as outlined in algorithm 1, using $K = 100$. As the matrices are symmetric, only the lower triangular parts are displayed to enhance readability.

cumulative cross-products of intraday return residuals over the forecast horizon; for example, see Andersen, Bollerslev, and Lange (1999) (henceforth ABL) or Andersen, Bollerslev, Diebold, and Labys (2001). Unfortunately, we only have daily return data available, but the same methodology can be applied to them; this results in a less precise but still useful proxy. We consider forecast horizons of 1, 2, and 4 weeks. Note that there are standard formulas to compute the 2-week and 4-week forecasts for multivariate GARCH models, given the 1-week forecast and the estimated model at time $t - 1$; for example, see ABL. To compute the 2-week and 4-week forecasts for RiskM and Window, we simply multiply the 1-week forecasts by the forecast horizon. Denote by $\hat{H}_{t,k}$ the estimated conditional covariance matrix, based on the information available at time $t - 1$, for the k-week forecast horizon; in this notation $\hat{H}_{t,1}$ corresponds to $\hat{H}_t$, the 1-week forecast. Also, let $\Sigma_{t,k}$ be the cumulative cross-products of daily return residuals during that period. The typical elements of these two matrices are denoted by $\hat{h}_{ij,t,k}$ and $\sigma_{ij,t,k}$, respectively. As do ABL, we

TABLE 4.—PARAMETER ESTIMATES OF THE CCC MODEL

U.S.	U.K.	France	Germany	Japan	Canada	Switzerland
			c_{ii}			
0.1850	0.0733	0.3495	0.1514	0.2356	0.2992	0.5743
			a_{ii}			
0.0700	0.0520	0.0731	0.0446	0.0727	0.0898	0.0467
			b_{ii}			
0.8872	0.9379	0.8819	0.9278	0.8977	0.8443	0.8367
			Correlation Matrix			
1.0000						
0.4392	1.0000					
0.3572	0.4888	1.0000				
0.3302	0.4623	0.5643	1.0000			
0.2469	0.3369	0.3373	0.3688	1.0000		
0.6753	0.4352	0.3459	0.3267	0.2244	1.0000	
0.3655	0.4852	0.5299	0.7009	0.3891	0.3450	1.0000

This table presents the estimated parameters of the constant-conditional-correlation (CCC) GARCH(1,1) model based on the entire sample. The model is described in the first part of section IIIB.

TABLE 5.—PARAMETER ESTIMATES OF THE BEKK MODEL

U.S.	U.K.	France	Germany	Japan	Canada	Switzerland
			G			
0.3034						
0.1004	0.1662					
0.1508	0.2863	0.2634				
0.1291	0.2558	0.0215	0.2696			
0.0842	0.1154	0.0374	0.0561	0.3300		
0.2408	0.0831	−0.0329	−0.0270	−0.0060	0.2485	
0.2039	0.3923	−0.0519	0.0905	0.0557	−0.0620	0.2302
			Diag (*E*)			
0.1446	0.1283	0.2141	0.2050	0.1709	0.1567	0.1872
			Diag (*F*)			
0.9786	0.9885	0.9657	0.9655	0.9769	0.9729	0.9557

This table presents the estimated parameters of the diagonal BEKK GARCH(1,1) model based on the entire sample. The model is described in the second part of section IIIB. Note that G is a lower triangular matrix, so the elements not displayed are equal to zero.

consider the following two criteria to judge the quality of the volatility forecasts:

$$RMSE_k = \left(\frac{1}{N^2} \sum_{i,j} E(\hat{h}_{ij,t,k} - \sigma_{ij,t,k})^2 \right)^{1/2}, \tag{14}$$

$$MAD_k = \frac{1}{N^2} \sum_{i,j} E|\hat{h}_{ij,t,k} - \sigma_{ij,t,k}|. \tag{15}$$

$RMSE_k$ and MAD_k are multivariate versions of the root-mean-square error and mean absolute deviation, respectively. Criteria based on absolute deviations are sometimes preferred (as, for example, in ABL), because they are more robust and less affected by a few large outliers.

Table 6 reports estimates of the two criteria at the different forecast horizons. There are six comparisons altogether (two criteria and three horizons). FlexM is best four times (for both criteria at the 1-week and 2-week horizons), and CCC is best two times (for both criteria at the 4-week horizon). RiskM and Window are always worse than the multivariate GARCH models.

TABLE 6.—FORECAST CRITERIA FOR COVARIANCE MATRICES

Model	*RMSE*	*MAD*
	1-Week Horizon	
FlexM	9.73	2.96
CCC	9.88	3.01
BEKK	9.90	3.09
RiskM	9.98	3.31
Window	10.02	3.42
	2-Week Horizon	
FlexM	15.48	5.13
CCC	15.70	5.22
BEKK	15.74	5.37
RiskM	16.07	5.88
Window	16.03	6.09
	4-Week Horizon	
FlexM	17.42	8.90
CCC	17.09	8.71
BEKK	17.48	9.37
RiskM	26.14	10.82
Window	25.61	11.10

This table compares the forecasted conditional covariance matrices with the realized integrated volatility covariance matrices computed from daily data that serve as a proxy for the true but unobservable conditional covariance matrices. The criteria *RMSE* and *MAD* are defined in equations (14)–(15). All forecasts are out of sample. Forecasts start at week $t = 601$.

E. Standardized Residuals

Consider the standardized residuals $\epsilon_t = H_t^{-1/2} x_t$, where H_t is the true conditional covariance matrix at time t. Obviously, the ϵ_t have constant conditional covariance matrix equal to the identity, and the cross-products $\epsilon_t \epsilon_t'$ are uncorrelated over time. It is therefore natural to test for any left-over autocorrelation in the cross-products $\hat{\epsilon}_t \hat{\epsilon}_t'$, where $\hat{\epsilon}_t = \hat{H}_t^{-1/2} x_t$ and $\hat{H}_t$ is the estimated conditional covariance matrix at time t.

A standard test for serial correlation in a univariate time series $\{y_t\}$ is the Ljung-Box test. The test statistic is

$$LB(k) = T \sum_{l=1}^{k} \frac{T+2}{T-l} \hat{\rho}(l),$$

where $\hat{\rho}(l)$ is the sample autocorrelation of order l, and k is an integer which is small compared to the sample size T. The commonly used asymptotic null distribution is χ_k^2, the *chi*-squared distribution with k degrees of freedom.

There are, however, two problems with applying this test for our purposes. A general problem is that the asymptotic null distribution is only correct under the additional assumption of i.i.d. data. If the series $\{y_t\}$ is uncorrelated but dependent, the χ_k^2 approximation can be arbitrarily misleading (Romano and Thombs, 1996). Another problem is that the test is designed for univariate series and not series of $N \times N$ matrices. We address these two problems simultaneously by suggesting a combined test statistic that takes into account all cross-product elements at once and by constructing a test that, under the null, only requires that the cross-products be uncorrelated rather than i.i.d.

TABLE 7.—STANDARDIZED RESIDUALS

Model	Test Statistic	*P*-value
FlexM	275.4	0.72
CCC	640.1	0.00
BEKK	760.1	0.00
RiskM	386.7	0.01
Window	845.3	0.00

This table presents test results for left-over autocorrelation in the fitted standardized residuals. The test statistic is the combined Ljung-Box statistic defined in equation (16) using the first $k = 12$ sample autocorrelations. The *P*-value for the null hypothesis of no autocorrelation is obtained by applying the subsampling method with block size $b = 100$, as detailed in the discussion following equation (16). All fitted standardized residuals are out of sample and are computed starting at week $t = 601$.

The combined test statistic we suggest is

$$LB_{\text{comb}}(k) = \sum_{1 \le i \le j \le N} LB_{ij}(k), \tag{16}$$

where $LB_{ij}(k)$ is the univariate Ljung-Box test statistic computed from the series $\{\hat{\epsilon}_{i,t}\hat{\epsilon}_{j,t}\}$. To assess the evidence against the null hypothesis, we compute the *P*-value based on the subsampling method. To this end, let $LB_{\text{comb},t,b}(k)$ be the combined test statistic based on the stretch of data $\{\hat{\epsilon}_t, \ldots, \hat{\epsilon}_{t+b-1}\}$, for $t = 1, \ldots, T - b + 1$. Here, the block size b is an integer smaller than T. The subsampling *P*-value is then given as

$$PV_{\text{Sub}} = \frac{\#\{LB_{\text{comb},t,b}(k) \ge LB_{\text{comb}}(k)\}}{T - b + 1}.$$

By arguments analogous to the ones of Romano and Thombs (1996), it can easily be shown that this test is consistent if the cross-products are uncorrelated but dependent. For more details about the general use of subsampling tests with dependent data, the reader is referred to Politis, Romano, and Wolf (1999, chapter 3). The block size b needs to satisfy the asymptotic conditions $b \to \infty$ and $b/T \to 0$; some methods for choosing b in practice are given in Politis, Romano, and Wolf (1999, chapter 9).

Table 7 presents the test statistic and corresponding *P*-value for the five models, using $k = 12$ and $b = 100$; the results are similar for other values of k and b. FlexM has the smallest test statistic and is the only model that is not rejected at any conventional level; its *P*-value is 0.72, the one for RiskM is 0.01, and all the others are 0.

F. Value at Risk

An important use of the conditional covariance matrix is in calculations of the value at risk (VaR) of a portfolio of assets. A large number of methods to compute the VaR have been suggested and are currently employed, such as historical simulation, RiskMetrics, Monte Carlo, GARCH, nonparametric quantile regressions, and methods based on extreme-value theory. We certainly do not aim to settle the dispute as to which method is best, and it stands to reason that a uniformly best method does not exist. However, GARCH methods are very popular among practitioners and tend to perform well. (In particular, recent claims that they are dominated by methods based on extreme-value theory do not seem to be substantiated; for example, see Lee and Saltoğlu, 2001.)

If a single portfolio is considered, it makes more sense to fit a univariate GARCH(1,1) model to the corresponding return series and base any VaR calculations on this model. On the other hand, if a number of different portfolios based on the same universe of N assets are considered (as is the case with different traders of an investment bank, say), it is common practice to base the individual VaR calculations on a single estimate of the conditional covariance matrix of all N assets. This also allows computing the marginal contributions to risk of each position and evaluating the effect of hedges. Hence, multivariate GARCH is certainly relevant to risk management applications.

In our tests, we consider the following four portfolios based on the seven market indices that make up our data:

- U.S. portfolio: United States only.
- North American portfolio: United States and Canada equally weighted.
- European portfolio: United Kingdom, France, Germany, and Switzerland equally weighted.
- World portfolio: all seven countries equally weighted.

We use the estimated conditional covariance matrix to compute the one-week-ahead VaR at levels 1% and 5%. In order to try to fit the tails of the return distributions and to match the theoretical VaR levels, we assume a conditional t-distribution. To be more specific, let the portfolio be represented by the vector of weights, w. The estimated conditional variance of the portfolio at time t is then given by

$$\hat{h}_{w,t} = w'\hat{H}_t w.$$

At time $t - 1$, we condition on the past portfolio returns and their corresponding estimated conditional variances to choose the number of degrees of freedom, ν^*, that maximizes the likelihood

$$\prod_{s=1}^{t-1} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)\hat{h}_{w,s}}\,\Gamma(\nu/2)} \left(1 + \frac{(w'x_s)^2}{(\nu-2)\hat{h}_{w,s}}\right)^{-(\nu+1)/2}$$

over ν, where $\Gamma(\cdot)$ denotes the gamma function. Note that the standard formula for the t-distribution has been modified by the scale factor $\hat{h}_{w,s}(\nu - 2)/\nu$, where the degree-of-freedom adjustment is designed so that $\hat{h}_{w,s}$ is exactly equal to the conditional variance of $w'x_s$. Having thus found ν^*, the 1% VaR at time t is finally computed as

$$t_{\nu^*,0.01}\sqrt{\hat{h}_{w,t}(\nu^* - 2)/\nu^*}.$$

Here, $t_{\nu^*,0.01}$ denotes the 0.01 quantile of the t-distribution with ν^* degrees of freedom. An analogous computation yields the 5% VaR.

TABLE 8.—VALUE AT RISK

Model	Hit Rate		*P*-value		
	For 1%	For 5%	For 1%	For 5%	Av. ν^*
			U.S. Portfolio		
FlexM	0.027	0.068	0.71	0.82	85.3
CCC	0.027	0.068	0.64	0.93	22.7
BEKK	0.018	0.068	0.36	0.99	12.7
RiskM	0.030	0.066	0.85	0.83	184.6
Window	0.029	0.068	0.78	0.24	119.3
			North American Portfolio		
FlexM	0.015	0.070	0.90	0.86	22.0
CCC	0.015	0.070	0.00	0.68	10.9
BEKK	0.015	0.068	0.07	0.79	9.0
RiskM	0.027	0.066	0.61	0.44	151.3
Window	0.028	0.062	0.01	0.51	115.3
			European Portfolio		
FlexM	0.010	0.054	0.01	0.18	11.8
CCC	0.016	0.060	0.08	0.01	7.6
BEKK	0.011	0.054	0.04	0.09	7.2
RiskM	0.015	0.058	0.55	0.30	34.5
Window	0.017	0.052	0.00	0.00	20.7
			World Portfolio		
FlexM	0.016	0.063	0.03	0.44	19.9
CCC	0.015	0.069	0.00	0.00	10.1
BEKK	0.016	0.053	0.03	0.00	9.2
RiskM	0.020	0.065	0.18	0.26	103.2
Window	0.021	0.061	0.02	0.00	55.4

This table compares VaR calculations at levels 1% and 5%, assuming a conditional t-distribution (suitably normalized). The hit rate is the sample mean of the hit series defined in equation (17) and should be close to the nominal level. The P-value corresponds to the null hypothesis of no autocorrelation in the hit series and is obtained from the usual *chi*-squared approximation of the univariate Ljung-Box test statistic using the first $k = 12$ sample autocorrelations. The last column shows the average optimal number of degrees of freedom, ν^*, for the conditional t-distribution. All VaR calculations are out of sample and start at week $t = 601$.

For a certain portfolio and for a given level, define the hit variable

$$hit_t = I\{w' x_t < \widehat{VaR}_t\}, \tag{17}$$

where $I\{\cdot\}$ is the indicator function and $\widehat{VaR}_t$ is the estimated VaR at time t. If the model to calculate the VaR is correctly specified, the series $\{hit_t\}$ should be uncorrelated over time and have expected value equal to the desired nominal level.

Table 8 presents the sample means (or hit rates) and the Ljung-Box P-values for autocorrelation of the hit series for the various methods, portfolios, and VaR levels. The P-values are based on the first $k = 12$ sample autocorrelations. Because the hit series are univariate and because a (stationary) {0, 1} series is uncorrelated if and only if it is independent, it is safe to use the asymptotic *chi*-squared approximation to compute the P-values here. (The table also presents the average optimal number of degrees of freedom, ν^*, of the conditional t-distribution.)

The hit rates are all reasonably close to the target levels, although they tend to be a bit larger on average. There is no clear winner or loser in terms of the hit rates. Judging the serial correlation of the hit series $\{hit_t\}$, it is seen that RiskM performs best: all its P-values are above 0.1. FlexM is somewhat better than the other GARCH models.

G. *Portfolio Selection*

Another important application of the conditional covariance matrix is as an input to the Markowitz (1952) portfolio selection method. Hence, we examine the gains from international diversification obtained by taking into account the time-changing nature of the covariance matrix. In order to avoid having to specify the vector of conditional expected returns, which is more a task for the portfolio manager than a statistical problem, we focus on constructing the (global) minimum-variance portfolio, allowing for short sales.

Table 9 shows the realized (annualized) standard deviation of the returns of the conditional-minimum-variance

TABLE 9.—STANDARD DEVIATION OF PORTFOLIO RETURNS

Portfolio	Standard Deviation
U.S.	15.87
Equal-weighted world	13.33
Unconditional minimum-variance	12.91
FlexM minimum-variance	12.32
CCC minimum-variance	12.53
BEKK minimum-variance	12.54
RiskM minimum-variance	13.37
Window minimum-variance	12.89

This table presents (annualized) standard deviations of various portfolios. The minimum-variance portfolios are those with globally minimum variance; no restriction on the expected return is made, and short sales are allowed. All portfolios are constructed starting at week $t = 601$.

portfolio over the entire sample period, obtained from the three GARCH models, the RiskMetrics method, and the rolling-window method. It compares them with the standard deviation of the U.S. stock market, of the equal-weighted portfolio of the seven stock markets, and of the unconditional minimum-variance portfolio obtained from the sample covariance matrix at $t = 1{,}356$. (The last portfolio would be infeasible, but we include it nevertheless.) Not surprisingly, fully investing in the U.S. stock market yields the highest standard deviation, followed by the equal-weighted world portfolio and the unconditional minimum-variance portfolio. All three GARCH models provide a significant improvement, with FlexM being the best. Window is comparable to the unconditional minimum-variance portfolio, and RiskM is worse than even the equal-weighted portfolio.

IV. Conclusion

In this paper, we have developed an estimation procedure for the general diagonal-vech formulation of the multivariate GARCH(1,1) model. Our procedure is the first to be computationally feasible for dimensions $N > 5$, without constraining the coefficient matrices. Our method proceeds in two steps: first, we decentralize the problem by estimating separately N univariate and $N(N-1)/2$ bivariate GARCH models, all of which are computationally feasible problems; second, we bring together these results to form N-dimensional matrices of parameter estimates, which we transform in order to ensure the positive semidefiniteness of the conditional covariance matrices. In doing so, we avoid having to impose additional restrictions, which has been the common approach so far in the multivariate GARCH literature. In addition, our method is computationally far less demanding than traditional multivariate models, which is an important advantage if the sample size is large, as would be the case with high-frequency data.

We apply our procedure to 25 years of weekly data on seven major national stock markets and compare it with two popular traditional multivariate GARCH(1,1) models, namely the constant-conditional-correlation model and the diagonal BEKK model, and with two widely-used, albeit less sophisticated, estimators, namely the rolling-window estimator and the exponential smoothing estimator. Using a number of criteria, such as forecast accuracy, persistence of standardized residuals, precision of value-at-risk estimates, and optimal portfolio selection, we find that the flexible multivariate GARCH method does indeed offer improved performance. The use of high-frequency data, which undoubtedly will increase in the future, should make our procedure even more attractive.

Direct applications of this method involve portfolio selection and tests of asset pricing models such as the international CAPM, and risk measurement uses such as the value at risk. An interesting topic left for future research is an extension to asymmetric multivariate GARCH(1,1).

REFERENCES

Andersen, Torben G., Tim Bollerslev, Frank X. Diebold, and Paul Labys, "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association* 96 (2001), 42–55.

Andersen, Torben G., Tim Bollerslev, and Steve Lange, "Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon," *Journal of Empirical Finance* 6 (1999), 457–477.

Bollerslev, Tim, "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," this REVIEW 72 (1990), 498–505.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge, "Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy* 96 (1988), 116–131.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, *The Econometrics of Financial Markets* (Princeton: Princeton University Press, 1997).

Ding, Zhuanxin, and Robert F. Engle, "Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing," University of California mimeograph, San Diego (1994).

Efron, Bradley, and Robert J. Tibshirani, *An Introduction to the Bootstrap* (New York: Chapman & Hall, 1993).

Engle, Robert, "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business and Economic Statistics* 20:3 (2002), 339–350.

Engle, Robert F., and Kenneth Kroner, "Multivariate Simultaneous GARCH," *Econometric Theory* 11 (1995), 122–150.

Engle, Robert F., and Joseph Mezrich, "GARCH for Groups," *Risk* 9 (1996), 36–40.

Lee, Tae-Hwy, and Burak Saltoğlu, "Evaluating the Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check," University of California mimeograph, Riverside (2001).

Markowitz, Harry, "Portfolio Selection," *Journal of Finance* 7 (1952), 77–91.

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf, *Subsampling* (New York: Springer, 1999).

Romano, Joseph P., and Lori A. Thombs, "Inference for Autocorrelations under Weak Assumptions," *Journal of the American Statistical Association* 91 (1996), 590–600.

Sharapov, Ilya, "Advances in Multigrid Optimization Methods with Applications," University of California, PhD dissertation, Los Angeles (1997).

Styan, George P. H., "Hadamard Products and Multivariate Statistical Analysis," *Linear Algebra and Its Applications* 6 (1973), 217–240.

APPENDIX

Minimization of the Frobenius Norm

1. Problem Formulation

Given a symmetric matrix A with the property diag $(A) > 0$, find a symmetric, positive semidefinite matrix M with diag $(M) =$ diag (A) that minimizes $\|A - M\|_F$, where $\|\cdot\|_F$ is the Frobenius norm.

2. Numerical Solution

Write the matrix A and the current approximation M to the solution of the above problem as

$$A = \begin{pmatrix} a_{11} & a^T \\ a & \bar{A} \end{pmatrix}, \qquad M = \begin{pmatrix} a_{11} & m^T \\ m & \bar{M} \end{pmatrix},$$

and let the conditions of the problem be satisfied [that is, diag $(M) =$ diag (A) and $M = M^T \geq 0$]. For a matrix of the form

$$P = \begin{pmatrix} \rho & x^T \\ 0 & I_{n-1} \end{pmatrix}$$

(where I_{n-1} is identity), we can introduce the next iterate by

$$\breve{M} = PMP^T = \begin{pmatrix} \rho^2 a_{11} + 2\rho x^T m + x^T M x & \rho m^T + x^T \bar{M} \\ \rho m + \bar{M} x & \bar{M} \end{pmatrix}. \quad \text{(A-1)}$$

If we enforce the condition

$$\rho^2 a_{11} + 2\rho x^T m + x^T M x = a_{11}, \quad \text{(A-2)}$$

then the new approximation $\breve{M}$ satisfies the conditions of the problem: $\breve{M} = \breve{M}^T \geq 0$ and diag $(\breve{M})$ = diag (A).

If equation (A-2) is satisfied, we have

$$\|A - \breve{M}\|_F - \|A - M\|_F = 2\|a - (\rho m + \bar{M} x)\|_2^2 - 2\|a - m\|_2^2;$$

therefore, choosing x and ρ that minimize $\|a - (\rho m + \bar{M} x)\|_2^2$, from equation (A-2) we get $\breve{M}$ that minimizes $\|A - \breve{M}\|_F$, satisfies the conditions of the problem, and is obtained from the previous approximation M by changing its first row and column. The extension to the i^{th} column and row is obvious.

Remark 1. The convexity of the problem implies that the solution matrix M is singular, that is, lies on the boundary of the feasible region. Since det $(\breve{M})$ = ρ^2 det (M), we can make the iterates stay within the interior of the feasible region by initializing the process with a nonsingular matrix and choosing ρ to be bounded away from 0. Later on we treat ρ as a chosen constant between 0 and 1, so the iterates become singular no faster then exponentially. In numerical examples, ρ is chosen to be 0.5.

One step of the iterative procedure becomes

$$\min_x \|a - (\rho m + \bar{M} x)\|_2^2$$

subject to equation (A-2); introducing

$$b = a - \rho m,$$

it becomes

$$\min_x \|\bar{M} x - b\|_2^2,$$

still subject to equation (A-2).

The Lagrangian of this subproblem is

$$L(x, \lambda) = \|\bar{M} x - b\|_2^2 + \lambda(\rho^2 a_{11} + 2\rho x^T m + x^T M x - a_{11}),$$

and the optimality conditions are

$$F(x) = \rho^2 a_{11} + 2\rho x^T m + x^T M x - a_{11} = 0 \quad \text{(A-3)}$$

and

$$\nabla_x L(x, \lambda) = \bar{0},$$

which can be written as

$$\bar{M}^2 x - \bar{M} b + \lambda\rho m + \lambda \bar{M} x = 0. \quad \text{(A-4)}$$

For any λ, equation (A-4) can be solved for x:

$$x(\lambda) = (\bar{M}^2 + \lambda \bar{M})^{-1}(\bar{M} b - \lambda\rho m); \quad \text{(A-5)}$$

and

$$F(\lambda) = F(x(\lambda)) = 0$$

can be solved by the Newton's method:

$$\lambda \Leftarrow \lambda - \frac{F(\lambda)}{F_\lambda(\lambda)}. \quad \text{(A-6)}$$

The analytic expression for $F_\lambda(\lambda)$ can be obtained from

$$F_\lambda(\lambda) = \nabla_x F(x) \cdot x_\lambda = 2(\rho m + \bar{M} x)^T x_\lambda. \quad \text{(A-7)}$$

By differentiating equation (A-4) with respect to λ we get

$$\bar{M}^2 x_\lambda + \rho m + \bar{M} x + x \bar{M} x_\lambda = 0;$$

therefore

$$x_\lambda = -(\bar{M}^2 + \lambda \bar{M})^{-1}(\rho m + \bar{M} x).$$

Inserting this in equation (A-7), we get

$$F_\lambda(\lambda) = -2(\rho m + \bar{M} x)^T (\bar{M}^2 + \lambda \bar{M})^{-1}(\rho m + \bar{M} x). \quad \text{(A-8)}$$

We can summarize the solution of the subproblem as:

1. Initialize λ (say $\lambda = 0$).
2. Compute x by equation (A-5).
3. Compute $F(\lambda)$ and $F_\lambda(\lambda)$ using equations (A-3) and (A-8).
4. Update λ using Newton's step (A-6).
5. Recur Newton's procedure.

Remark 2. The steps (A-3) and (A-8) involve the inverse of $\bar{M}^2 + \lambda \bar{M}$, which is singular if $\bar{M}$ is. Restricting ρ to be a nonzero constant results in nonsingular $\bar{M}$ unless it is a solution; see remark 1.

A Matlab routine implementing this procedure has been written by Ilya Sharapov and is available from the authors upon request.

3. Numerical Tests

Extensive simulation tests of the numerical routine have been implemented. Numerical convergence is typically obtained after one or two iterations with minimal error. The results of these tests are omitted for brevity but can be obtained from the authors upon request.