

# Large Time-Varying Covariance Matrices with Applications to Finance

Petros Dellaportas and Mohsen Pourahmadi

petros@aueb.gr      pourahm@math.niu.edu

*Department of Statistics, Athens University of Economics and Business, Greece*

*Division of Statistics, Northern Illinois University, DeKalb, IL 60115, USA*

Summary: Correlations among the asset returns are the main reason for the computational and statistical complexities of the full multivariate GARCH models. We rely on the variance-covariance separation strategy and introduce a broad class of multivariate models in the spirit of Engle's (2002) dynamic conditional correlation models, that is univariate GARCH models are used for variances of individual assets coupled with parsimonious parametric models either for the time-varying correlation matrices or the components of their spectral and Cholesky decompositions. Numerous examples of structured correlation matrices along with structured components of the Cholesky decomposition are provided. This approach, while reducing the number of correlation parameters and severity of the positive-definiteness constraint, leaves intact the interpretation and magnitudes of the coefficients of the univariate GARCH models as if there were no correlations. This property makes the approach more appealing than the existing GARCH models. Moreover, the Cholesky decompositions, unlike their competitors, decompose the normal likelihood function as a product of univariate normal likelihoods with independent parameters resulting in fast estimation algorithms. Gaussian maximum likelihood methods of estimation of the parameters are developed. The methodology is implemented for a real financial dataset with one hundred assets, and its forecasting power is compared with other existing models. Our preliminary numerical results show that the methodology can be applied to much larger portfolios of assets and it compares favorably with other models developed in quantitative finance.

*Some key words:* Autoregressive conditional heteroscedastic models; latent factor models; time-varying ARMA coefficients; Cholesky decomposition; principal components; spectral decomposition, stochastic volatility models; maximum likelihood estimation.

## 1 Introduction

Many tasks of modern financial management including portfolio selection, option pricing and risk assessment can be reduced to the prediction of a sequence of large  $N \times N$  covariance matrices  $\{\Sigma_t\}$  based on the (conditionally) independently  $N(0, \Sigma_t)$ -distributed data  $r_t, t = 1, 2, \dots, T$ , where

$r_t$  is the shock (innovation) at time  $t$  of a multivariate time series of returns of  $N$  assets in a portfolio. Since the parameters in  $\Sigma_t$  are constrained by the positive-definiteness requirement and their number grows quadratically in  $N$ , the problem of parsimonious modeling of  $\{\Sigma_t\}$  is truly challenging and has been studied earnestly in the literature of finance in the last two decades (Engle, 1982, 2002). The key idea is to write difference equations for  $\{\Sigma_t\}$  similar to the univariate autoregressive and moving average (ARMA) models (Box *et al.*1994). More precisely, with  $\mathcal{F}_t$  standing for the past information up to and including the time  $t$ , it is assumed that  $r_t|\mathcal{F}_{t-1} \sim N(\mu_t, \sigma^2)$ . This model with constant-variance restriction is usually not supported by many financial series and was relaxed in the pioneering work of Engle (1982) who defined the class of *autoregressive conditional heteroscedastic* (ARCH) models and Bollerslev (1986) who introduced the generalized ARCH (GARCH) models by

$$\begin{cases} r_t|\mathcal{F}_{t-1} \sim N(\mu_t, \sigma_t^2), \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \end{cases} \quad (1)$$

where the constraints  $\alpha_0 > 0$  and  $\alpha_i \geq 0, \beta_i \geq 0$ , ensure a positive variance. Fortunately, many properties of GARCH models can be understood by viewing them as exact ARMA models for the squared return series  $\{r_t^2\}$ , so that one can bring the full force of ARMA model-building process to bear on the new class of GARCH models for the unobserved time-varying variances  $\{\sigma_t^2\}$  (Tsay, 2002, Chap.3).

Emboldened by the ease of use and success of univariate GARCH models, many early variants of multivariate GARCH models (Engle and Kroner, 1995) were defined simply as difference equations of the form (1) either for the vectorized sequence of covariance matrices  $\{vec \Sigma_t\}$  or the sequence  $\{\Sigma_t\}$  itself with suitable matrix coefficients. The number of free parameters of such

models is known to grow profligately (Sims, 1980) and are proportional to  $N^4$  and  $N^2$ , respectively. Simplification occurs (Alexander, 2001, Chap.7) when the coefficients are diagonal matrices, in which case, each variance/covariance term in  $\Sigma_t$  follows a univariate GARCH model with the lagged variance/covariance terms and squares and cross products of the data (Ledoit, Santa-Clara and Wolf, 2003), but complicated restrictions on the coefficient parameters are needed to guarantee their positive-definiteness. These restrictions are often too difficult to satisfy in the course of iterative optimization of the likelihood function even when the number of assets is about five. Consequently, for large covariance matrices the use of full multivariate GARCH models has proved impractical (Engle, 2002). Meanwhile, alternative classes of more practical multivariate GARCH models *generated by univariate* GARCH models are becoming popular. For example, the class of  $k$ -factor GARCH models, see (4) in Section 2, allows the individual asset volatilities and correlations to be generated by  $k + 1$  univariate GARCH models of the  $k$  latent series and the specific (idiosyncratic) errors.

In this paper, we show that separating the time-varying variances  $\{D_t\}$  and correlations  $\{R_t\}$  of the vector of return  $\{r_t\}$ , i.e.

$$\Sigma_t = D_t R_t D_t, \tag{2}$$

is ideal for resolving some of these complications. We model the volatility of the  $j$ th asset  $\{\sigma_{jt}^2\}$ , or the  $j$ th diagonal entry of  $\{D_t\}$ ,  $j = 1, 2, \dots, N$ , using the univariate GARCH models (1), and introduce parsimonious models for the time-varying correlations  $\{R_t\}$  of the  $N$  assets. Highly desirable and practical features of this approach are that, (i) we work with the original returns instead of latent factors constructed from them, (ii) the multivariate and univariate forecasts are

consistent with each other, in the sense that, when new assets are added to the portfolio, the volatility forecasts of the original assets will be unchanged and (iii) the estimation of the volatility and correlation parameters are separated. Recently, to reduce the high number of correlation parameters and to allow some dynamics for  $\{R_t\}$ , Engle (2002) and Tse and Tsui (2002) have introduced simple GARCH-type difference equations of the form

$$R_t = (1 - \alpha - \beta)\tilde{R} + \alpha R_{t-1} + \beta\psi_{t-1}, \quad (3)$$

where  $\tilde{R}$  is the sample correlation matrix of the vector of standardized returns and  $\psi_{t-1}$  is a positive-definite correlation matrix depending on the lagged data. The two parameters  $\alpha, \beta$  are nonnegative with  $\alpha + \beta \leq 1$ , so that  $R_t$  as a weighted average of positive-definite matrices with nonnegative coefficients is guaranteed to be positive-definite. Though such models are highly parsimonious, they may not be realistic in the sense that all pairwise correlations between assets are assumed to follow the same simple dynamics with identical coefficients. For example, it is implausible to think that the dynamics of the correlations of two technology stock returns and two utility returns are identical.

We provide some parsimonious models for the time-varying correlation matrices  $\{R_t\}$ , but instead of (3) we write difference equations either for its parameters or the parameters of the components of its spectral and Cholesky decompositions of  $R_t$  as well as those of its factor models. The new class of models are shown to be related to the standard and familiar factor models (Diebold and Nerlove, 1989; Vrontos et al. 2003), and orthogonal GARCH models (Alexander, 2001).

The outline of the paper is as follows. In Section 2 we review variants of multivariate GARCH and dynamic factor models for financial time series (Engle and Rothchild, 1990; Pitt and Shep-

hard, 1999a; Aguilar and West, 2000; Christodoulakis and Satchell, 2000; Vrontos et al. 2003). Many examples of structured and dynamic models for time-varying correlation matrices and their Cholesky factors are discussed in Section 3. These models are more parsimonious than Bollerslev’s (1990) constant correlation models and comparable to (3) and the multivariate GARCH models. It is shown that the problem of multivariate conditional covariance estimation can be reduced to estimating the  $3N$  parameters of univariate GARCH models and about 3 or 4 “dependence” parameters. Maximum likelihood procedure for the former are well-known (Vrontos *et al.*, 2000; Tsay, 2002) and will not be discussed here, such results for the “dependence” parameters being new are presented in Section 4, and an example of financial data with  $N = 100$  is presented in Section 5. Section 6 concludes the paper.

## 2 Dynamic Factor and Orthogonal GARCH Models

The close connection among hierarchical factor models, spectral and two Cholesky decompositions of covariance matrices are presented in this section. For generality, our coverage refers to the returns  $\{r_t\}$  with covariances  $\{\Sigma_t\}$ . However, most empirical work in Sections 4 and 5 will rely on the standardized returns and their correlation matrices  $\{R_t\}$ .

### 2.1 Hierarchical and Dynamic Factor GARCH Models

Of the many attempts to deal with the high-dimensionality and positive-definiteness problems in modeling covariance matrices, factor models seem to be the most promising. A  $k$ -factor model for

the returns is usually written as

$$r_t = Bf_t + e_t, \quad (4)$$

where  $f_t = (f_{1t}, \dots, f_{kt})'$  is a  $k$ -vector of time-varying common factors with a diagonal covariance matrix  $V_t = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{kt}^2)$ ,  $B_{N \times k}$  is a matrix of factor loadings and  $e_t$  is a vector of specific (idiosyncratic) errors with a diagonal covariance matrix  $\mathbf{W}_t = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{Nt}^2)$ . Using univariate GARCH models for the  $k$  time-varying common factor variances  $\{\sigma_{it}^2\}$  and the specific variances  $\{\sigma_{jt}^2\}$  will reduce their high number of parameters and allows generating  $N \times N$  time-varying covariance matrices in terms of only  $k + 1$  univariate GARCH models.

For  $k = 1$ , (4) is the capital asset pricing model (CAPM), where  $\{f_t\}$  stands for the market returns and the parameters of the univariate GARCH models can be interpreted easily (Diebold and Nerlove, 1989). However, for  $k > 1$  since  $Bf_t = BPP'f_t$  for any orthogonal matrix  $P$ , the matrix of factor loadings  $B$  and the common factors  $f_t$  are identifiable up to a rotation matrix. The nonuniqueness of the pair  $(B, f_t)$  is a source of some controversies and opportunities. Fortunately, the recent work in finance (Geweke and Zhou, 1996; Aguilar and West, 2000) shows that a unique  $k$ -factor model is possible if  $B$  is restricted to have full-rank  $k$  with a “hierarchical” factor structure,

i.e.

$$B = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ b_{2,1} & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ b_{k,1} & b_{k,2} & b_{k,3} & \cdots & 1 \\ b_{k+1,1} & b_{k+1,2} & b_{k+1,3} & \cdots & b_{k+1,k} \\ \vdots & & & & \vdots \\ b_{N,1} & b_{N,2} & b_{N,3} & \cdots & b_{N,k} \end{pmatrix}. \quad (5)$$

Of course, it is evident from (4) that such choice of  $B$  corresponds to an *a priori* ordering of the components of  $r_t$  in the sense that the first time series  $\{r_{1t}\}$  is essentially the first latent process  $\{f_{1t}\}$  save an additive noise, the second series  $\{r_{2t}\}$  is a linear combination of the first two latent factors plus a noise and so on. This is tantamount to introducing a tentative order among the components of  $r_t$ . While ordering variables is a challenging problem, lately there has been good progress in developing algorithms to arrive at “optimal” ordering that, for example, minimizes the bandwidth of the Cholesky factor of a positive-definite matrix.

The dynamic factor models of Aguilar and West (2000) and Christodoulakis and Satchell (2000) replaces the matrix  $B$  in (4) by the time-varying matrix of factor loadings  $\{B_t\}$ :

$$r_t = B_t f_t + e_t. \quad (6)$$

Moreover, assuming that  $\{f_t\}$  and  $\{e_t\}$  are independent, the factor model (6) leads to the decomposition

$$\Sigma_t = B_t V_t B_t' + W_t. \quad (7)$$

For identification purposes, the loading matrices  $B_t$  are constrained to be block lower triangular as in (5). A way to reduce the dimension of the parameters in  $\{B_t; 1 \leq t \leq n\}$ , is to write smooth evolution equations like (3) for the time-varying matrices of factor loadings. To this end, one may stack up the non-redundant entries of  $B_t$  in a  $d = Nk - k(k+1)/2$  dimensional vector  $\theta_t = (b_{21,t}, b_{31,t}, \dots, b_{Nk,t})'$ , and then write a first-order autoregression for  $\{\theta_t\}$  with scalar coefficients as in (3). Aspects of this approach are developed in Lopes, Aguilar and West (2002).

## 2.2 The Orthogonal GARCH Models

An approach closely related to (4)-(7) for estimating multivariate models is the orthogonal GARCH or principal component GARCH method, advocated independently by Klaassen (2000) and Alexander (2001). The key idea is to remove the instantaneous correlations in  $r_t$  through a linear transformation, that is for each  $t$  find a matrix  $A_t$  so that the components of  $Z_t = A_t r_t$  are uncorrelated. When univariate GARCH models are fitted to variances of the components of  $\{Z_t\}$ , then we say  $\Sigma_t = cov(r_t)$  has an orthogonal GARCH model. Since

$$A_t \Sigma_t A_t' = V_t, \tag{8}$$

is of the form (7) with  $W_t \equiv 0$ , it follows that orthogonal GARCH models are extreme examples of the more familiar factor models. Two important choices for  $A_t$  are the orthogonal and lower triangular matrices, corresponding to the spectral and Cholesky decomposition of  $\Sigma_t$ , respectively.

In the case of the spectral decomposition, the instantaneous linear transformations turn out to be the orthogonal matrices  $\{P_t\}$  consisting of the normalized eigenvectors of  $\Sigma_t$ . The time-invariant case  $P_t \equiv P$  has been studied extensively by Flury (1988) in the literature of multivariate

statistics and by Klaassen (2000) and Alexander (2001) in the literature of finance. However, the time-varying case is quite challenging due to the orthogonality of  $P_t$ 's whereby writing a suitable analogue of (3) is not easy.

The instantaneous transformations in the case of Cholesky decomposition turn out to be the unit lower triangular matrices  $\{T_t\}$  whose entries have interpretation as the regression coefficients, see Pourahmadi (1999); Christodoulakis and Satchell (2000); Tsay (2002, Chap.9). This case being newer and less familiar is discussed in the next subsection.

### 2.3 The Cholesky Decompositions: AR and MA Structures

We rely on the notion of regression to motivate the use of lower triangular matrices in (8). For the time being, we drop the subscript  $t$  in  $Y_t, \Sigma_t$  and focus on the contemporaneous covariance structure of a generic random vector  $Y = (y_1, \dots, y_N)'$ , by viewing  $y_1, y_2, \dots, y_j, \dots, y_N$  as a time series indexed by  $j$ . Consider regressing  $y_j$  on its predecessors  $y_1, \dots, y_{j-1}$ :

$$y_j = \sum_{k=1}^{j-1} \phi_{jk} y_k + \varepsilon_j, \quad j = 1, 2, \dots, N, \quad (9)$$

where  $\phi_{jk}$  and  $\sigma_j^2 = \text{var}(\varepsilon_j)$  are the unique regression coefficients and residual variances and by convention  $\sum_{k=1}^0 = 0$ . Indeed, with  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$  and  $\nu = \text{cov}(\varepsilon) = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$ , one can write (9) in the matrix form  $TY = \varepsilon$ , where  $T$  is a unit lower triangular matrix with  $-\phi_{jk}$  in the  $(j, k)$ th position, then it follows that the unit lower triangular matrix  $T$  diagonalizes  $\Sigma$ :

$$T\Sigma T' = \nu. \quad (10)$$

The pair of matrices  $(T, \nu)$  are the components of the modified Cholesky decomposition of  $\Sigma$  (Pourahmadi, 1999). For an unstructured covariance matrix, the nonredundant entries of  $T$  and  $\nu$

are referred to as its *generalized autoregressive parameters* (GARP) and *innovation variances* (IV), respectively.

Since  $T^{-1} = B = (\theta_{ij})$  is also a unit lower triangular matrix, it follows from (10) that

$$\Sigma = B\nu B' = \begin{pmatrix} \sigma_1^2 & \theta_{21}\sigma_1^2 & \theta_{31}\sigma_1^2 & \cdots & \theta_{N1}\sigma_1^2 \\ \theta_{21}\sigma_1^2 & \sum_{i=1}^2 \theta_{2i}^2\sigma_i^2 & \sum_{i=1}^2 \theta_{3i}\theta_{21}\sigma_i^2 & \cdots & \sum_{i=1}^2 \theta_{21}\theta_{Ni}\sigma_i^2 \\ \theta_{31}\sigma_1^2 & \sum_{i=1}^2 \theta_{2i}\theta_{3i}\sigma_i^2 & \sum_{i=1}^3 \theta_{3i}^2\sigma_i^2 & \cdots & \sum_{i=1}^3 \theta_{3i}\theta_{Ni}\sigma_i^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{N1}\sigma_1^2 & \sum_{i=1}^2 \theta_{2i}\theta_{Ni}\sigma_i^2 & \sum_{i=1}^3 \theta_{3i}\theta_{Ni}\sigma_i^2 & \cdots & \sum_{i=1}^N \theta_{Ni}^2\sigma_i^2 \end{pmatrix}. \quad (11)$$

In fact, regressing  $y_j$  on the past innovations  $\varepsilon_1, \dots, \varepsilon_{j-1}$  or from  $TY = \varepsilon$ , it follows that

$$y_j = \varepsilon_j + \sum_{k=1}^{j-1} \theta_{jk}\varepsilon_k, \quad j = 1, \dots, N, \quad (12)$$

$$Y = B\varepsilon. \quad (13)$$

Thus, as in the classical time series analysis, once  $T$  or the AR parameters are given, one can compute the MA parameters  $\theta_{jk}$  recursively and vice versa.

Note that the correlation coefficients computed from (12) depend on both the  $\theta_{ij}$ 's and  $\sigma_t^2$ 's, so that any error in modeling the IV's can have negative impact on the correlations. A way around this problem is to rescale the  $y_t$ 's and work with  $y_t/\sigma_t$  or use a slightly different Cholesky decomposition of the form

$$\Sigma = \nu^{1/2} \tilde{B} \tilde{B}' \nu^{1/2}, \quad (14)$$

where as before  $\tilde{B}$  is a unit lower triangular matrix which evidently determines the correlation

matrix  $R$ . More details on the applications and interpretations of entries of  $\tilde{B}$  can be found in Chen and Dunson (2003) and Pourahmadi (2007).

We show that the MA structure (13) is closely related to the factor models. To this end, we partition the innovation vector  $\boldsymbol{\varepsilon}$  and the matrix  $B$  so that (13) becomes

$$Y = B\boldsymbol{\varepsilon} = (B_1 \quad \vdots \quad B_2) \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \dots \\ \boldsymbol{\varepsilon}_2 \end{pmatrix} = B_1\boldsymbol{\varepsilon}_1 + B_2\boldsymbol{\varepsilon}_2. \quad (15)$$

Now, think of  $\boldsymbol{\varepsilon}_1$  as the  $k \times 1$  vector of latent factors,  $B_1$  the corresponding matrix of factor loadings, and  $B_2\boldsymbol{\varepsilon}_2$  as the vector of idiosyncratic errors. Then, the latent factors  $\boldsymbol{\varepsilon}_1$  have clear statistical interpretations as the first  $k$  innovations of  $Y$  and (15) has the appearance of the factor model (4). For  $k = N$ , the vector of idiosyncratic errors in (15) is zero and it reduces to the full-factor representation of  $Y$  (Aguilar and West, 2000; Vrontos *et al.* 2003).

## 2.4 Reparameterization Using Partial Autocorrelations

In this section we try to mimick the phenomenal success of the partial autocorrelation function (PACF) in model formulation and removing the positive-definiteness constraint on the autocorrelation function of a stationary time series. To this end, note that once an order is fixed among the entries of a random vector, then one can establish a one-to-one correspondence between a general correlation matrix  $R$  and its associated matrix of partial autocorrelations  $\boldsymbol{\Pi} = (\pi_{ij})$ , where  $\pi_{ii} = 1$  and for  $i < j$ ,  $\pi_{ij}$  is the partial correlation between  $y_i$  and  $y_j$  adjusted for the *intervening* variables (Joe, 2006). The matrix  $\boldsymbol{\Pi}$  is symmetric, but simpler than  $R$  since it is not required to be positive-definite, hence its entries are free to vary in the interval  $(-1, 1)$ . Furthermore, using

the Fisher z-transform, the matrix  $\mathbf{\Pi}$  can be mapped into a matrix  $\tilde{\mathbf{\Pi}}$  where its off-diagonal entries take values in the entire real line  $(-\infty, +\infty)$ .

An attractive feature of the above reparameterization is that using the *generalized partial correlogram*, i.e. the plot of  $\{\pi_{j,j+k}; j = 1, \dots, p-k\}$  versus  $k = 1, \dots, p-1$ , as a graphical tool, it is possible to formulate parsimonious models for  $\mathbf{\Pi}$  in terms of time lags and other covariates. Note that the partial autocorrelations  $\pi_{j,j+k}$  between successive variables  $y_j$  and  $y_{j+k}$  are grouped by their lags  $k = 1, \dots, p-1$ , and heuristically,  $\pi_{j,j+k}$  gauges the conditional (in)dependence between variables  $k$  units apart conditional on the intervening variables, so one expects it to be smaller for larger  $k$ . In the Bayesian framework, this intuition suggests putting shrinkage priors on  $\pi_{j,j+k}$  that shrinks the matrix  $\mathbf{\Pi}$  toward certain simpler structures (Daniels and Kass, 2001).

### 3 Structured Covariance and GARP Matrices

In this section we provide a few examples of structured covariance matrices with a small number of parameters denoted by the vector  $\rho$ . These matrices can be used, for example, in the Bollerslev's (1990) constant-correlation models to reduce the number of correlation parameters from the maximum of  $N(N-1)/2$  to as low as one, their time-varying analogue  $\{\rho_t\}$  with a smooth dynamic model like (3) will provide more realistic and flexible models than the dynamic correlation models of Engle (2002) and Tse and Tsui (2002). Similar structures for the lower triangular matrix  $T$  containing the GARPs in (9)-(10) will reduce the high number of parameters in the AR structures.

### 3.1 Examples of Structured $T$ , $B$ and $\nu$

Some of the most natural and familiar choices of  $T$ ,  $B$  and  $\nu$ , in increasing order of generality are given below and used later in the empirical work in Section 5. These structures are motivated by the commonly used exchangeable, AR and MA correlation matrices.

**Example 1. The Exchangeable GARPs:** Here all the nonredundant entries of  $T$  are equal,

$$\phi_{ij} \equiv \phi_0, j = 1, 2, \dots, N - 1; i = j + 1, \dots, N. \quad (16)$$

This matrix can be inverted easily to give the GMAPs:

$$\theta_{i,i-j} = \phi_0(1 + \phi_0)^{j-1}, j = 1, \dots, N - 1; i = j + 1, \dots, N.$$

Exploiting the inverse relationship between  $T$  and  $B$  one step further, reveals that the choice of the *exponential function*

$$\phi_{i,i-j} = \phi_0(1 + \phi_0)^{j-1}, j = 1, \dots, N - 1; i = j + 1, \dots, N.$$

for the GARPs, leads to the constant GMAPs;  $\theta_{ij} \equiv \phi_0$ . Fortunately, unlike the stationary case, the parameter  $\phi_0$  here is unconstrained so that  $(1 + \phi_0)^j$ , for  $j$  large, could decay or grow exponentially fast. For example, it decays exponentially fast if  $-2 < \phi_0 < 0$ , the case of  $\phi_0 = 1$  relates to nonstationary random walks (Zellner, 1979).

**Example 2. Toeplitz GARPs:** Here the entries along each subdiagonal of  $T$  are constant:

$$\phi_{i,i-j} = \phi_j, j = 1, 2, \dots, N - 1; i = j + 1, \dots, N. \quad (17)$$

where the  $\phi_j$ 's are unconstrained parameters. If needed, one could further reduce the dimension of

$(\phi_1, \phi_2, \dots, \phi_{N-1})$  via parametric models as in Pourahmadi (1999) by setting, for example,

$$\phi_j = \gamma_0 + \gamma_1 j^{\pm k}, j = 1, \dots, N - 1, \quad (18)$$

where  $\gamma_0, \gamma_1$  are the two new parameters and  $k$  is a known positive integer.

In applications where constancy along the subdiagonals of  $T$  is deemed inappropriate, one could exponentiate  $\phi_j$ 's by the Box-Cox transformation of the (time) index  $i$  along those subdiagonals. Namely, a non-Toeplitz  $T$  can be obtained by setting

$$\phi_{i,i-j} = \phi_j^{f(i;\lambda_j) - f(i-j;\lambda_j)}, j = 1, 2, \dots, N - 1; \quad i = j + 1, \dots, N, \quad (19)$$

where

$$f(x; \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \log x & \text{if } \lambda = 0. \end{cases}$$

For example, if  $0 < \phi_j < 1$ , then the entries along the  $j$ th subdiagonal of  $T$  are monotone increasing if  $\lambda_j < 1$ , monotone decreasing if  $\lambda_j > 1$ , or constant if  $\lambda_j = 1$ . For other range of values of  $\phi_j$ , similar nonconstant patterns could be prescribed depending on the values of the exponent  $\lambda_j$ . Here, the number of parameters  $(\phi_1, \dots, \phi_{N-1}, \lambda_1, \dots, \lambda_{N-1})$  could be as small as 2 or as large as  $2(N - 1)$ , see Example 4 for an important practical case where the number of parameters in  $T$  is reduced to 2.

**Example 3. Tensor-Product GARPs:** Here the subdiagonal entries of  $T$  are the tensor-product of  $\phi_1, \phi_2, \dots, \phi_{N-1}$  appearing in Example 2, i.e.

$$\phi_{ij} = \phi_i \phi_j, i = 2, \dots, N; j = 1, \dots, N - 1. \quad (20)$$

In general, there are  $N - 1$  possibly distinct parameters, their numbers can be reduced by relying on the specific models for  $\phi_j$  as in (18).

**Example 4. Non-Toeplitz GARPs of order  $p$ :** For a positive integer  $p$ , the last  $N - p - 1$  subdiagonals of  $T$  are set to zero and the rest are arbitrary. This is reminiscent of the nonstationary  $AR(p)$  or antedependence models of order  $p$ , see Pourahmadi (2001, Sec. 3.6) for more details and references. Note that the number of distinct parameters in  $T$ , is  $p \left( N - \frac{p+1}{2} \right)$ , selection of  $p$  and further reduction of the parameters can be achieved by relying either on the regressogram which provides a graphical tool for formulating models for the unconstrained parameters  $\{\phi_{jk}, \log \sigma_j^2\}$ , see Pourahmadi (1999, 2001) and Section 5 here.

Alternatively, one could use specific parametric models for the non-zero entries of  $T$  introduced in Examples 1-3. For the purpose of illustration, consider the simple case of  $p = 1$  where only the first subdiagonal of  $T$  is nonzero with entries

$$\phi_{i,i-1} = \phi_{i-1}, i = 2, \dots, N. \quad (21)$$

The corresponding GMAPs turn out to be of the form

$$\theta_{i,i-j} = \prod_{k=1}^j \phi_{i-k}, \quad j = 1, \dots, N-1; i = j+1, \dots, N, \quad (22)$$

and the correlation matrix from (11) has the remarkable property that all its entries are determined by the lag-one correlations  $\rho_1, \dots, \rho_{N-1}$ :

$$\rho_{i,i-j} = \prod_{k=1}^j \rho_{i-k}, \quad j = 1, \dots, N-1; i = j+1, \dots, N. \quad (23)$$

Even in this simple case, one has to deal with  $N-1$  lag-one GARPs or correlation parameters which could be large in some applications. Their numbers can be reduced considerably by employing the idea of Box-Cox power transformation and writing

$$\phi_{i,i-1} = \phi^{f(i;\lambda) - f(i-1;\lambda)}, i = 2, \dots, N, \quad (24)$$

in terms of only two parameters  $(\phi, \lambda)$ .

Finally, the number of innovation variances  $\sigma_1^2, \dots, \sigma_N^2$  in  $\mathcal{V}$  can also be reduced using parametric models such as a low-order polynomial of  $j$  for  $\log \sigma_j^2$ . Since for a correlation matrix  $R$  these variances must be decreasing in  $j$  with  $\sigma_1^2 = 1$ , for most of our work here we rely on the following simple function,

$$\sigma_j^2 = \exp(\lambda_0 - \lambda_1 j), j = 2, \dots, N; \quad (25)$$

where  $\lambda_1 \geq 0$  and  $\lambda_0$  arbitrary, are its two parameters.

### 3.2 Some Time-Varying Correlation Matrices

More flexible and realistic time-varying correlation matrices than the Engle's (2002) and Tse and Tsui (2002) dynamic conditional correlation matrices generated by (3) are introduced in this section. This can be done by writing difference equations either for the parameter vector of the correlation matrices or for the GARPs associated to their Cholesky decompositions. Of course, one may go beyond the examples mentioned above.

**Example 5.** The single parameter  $\rho$  of an  $AR(1)$  correlation matrix is in  $(-1, 1)$  but its Fisher- $z$  transform  $\log \frac{1-\rho}{1+\rho}$  takes values in  $(-\infty, \infty)$ . The corresponding time-varying correlation matrix has a parameter sequence  $\{\rho_t\}$  satisfying the difference equation

$$\log \frac{1 - \rho_t}{1 + \rho_t} = \alpha \log \frac{1 - \rho_{t-1}}{1 + \rho_{t-1}} + e_t, \quad (26)$$

where  $|\alpha| < 1$  and  $\{e_t\}$  is a white noise with mean 0 and variance  $\sigma_e^2$ .

Now, shifting attention to the Cholesky decomposition of correlation matrices  $\{R_t\}$ , and allow-

ing their GARPs and IVs in (10) to be time-varying,

$$T_t R_t T_t' = \nu_t = \text{diag}(1, \sigma_{2t}^2, \dots, \sigma_{Nt}^2), \quad (27)$$

one could select  $\{T_t\}$  and  $\{\nu_t\}$  to have simple parametric forms as in Examples 1-4 and (25), and write analogue of (26) for the time-varying vector of parameters. The same idea can be applied to the components of the slightly different Cholesky decomposition in (14). Alternatively, in analogy with Bollerslev's (1990) constant-correlation models one could start by assuming constant-GARP models, i.e.  $T_t \equiv T$ , but allow  $\{\nu_t\}$  to remain time-varying or vice versa. More details in implementing an approach close to this and the relevant empirical results can be found in Lopes et al. (2002).

## 4 Estimation

In this section we present some of the conceptual underpinnings for estimation of the correlation parameters. More specifically, using the normality assumption and (2), the log-likelihood function, up to ignoring some irrelevant constants, is the sum of a volatility part ( $L_V$ ) and a correlation part

$(L_C)$ :

$$\begin{aligned}
-2L &= \sum_{t=1}^n (\log |\Sigma_t| + r_t' \Sigma_t^{-1} r_t) \\
&= \sum_{t=1}^n (\log |D_t R_t D_t| + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t) \\
&= \sum_{t=1}^n (\log |D_t|^2 + Y_t' Y_t) + \sum_{t=1}^n [\log |R_t| + Y_t' (R_t^{-1} - I) Y_t] \\
&= L_V(\theta) + L_C(\theta, \rho)
\end{aligned} \tag{28}$$

where  $Y_t = D_t^{-1} r_t$  is the vector of standardized returns,  $\theta$  the  $3N$ -vector of univariate GARCH parameters of  $\{D_t\}$  and  $\rho$  the parameters of  $\{R_t\}$ . Thus, the variance-correlation separation strategy allows us to maximize  $L$  by handling each term separately over the volatility and correlation parameters. The earliest and simplest example of (2) and (28) is the constant-correlation models of Bollerslev (1990) where the correlation matrices  $\{R_t\}$ , i.e.  $R_t \equiv R$  with  $N(N-1)/2$  parameters. Its maximum likelihood estimate (MLE) turns out to be the sample correlation of the vector of standardized returns  $\{Y_t\}$ , which is always positive-definite and the optimization of the likelihood function will not fail so long as the estimated variances are positive. In view of (28), we need to focus only on the estimation of the parameters  $\rho$  of the correlation part  $L_C$  (for a given vector of volatility parameters  $\hat{\theta}$ ).

From (28), ignoring the term  $Y_t'Y_t$  and  $\varepsilon_{1t}^2$  which do not depend on  $\rho$ , we have

$$\begin{aligned}
L_C(\hat{\theta}, \rho) &= \sum_{t=1}^n \left( \sum_{j=1}^N \log \sigma_{jt}^2 + Y_t' T_t' \nu_t^{-1} T_t Y_t \right) \\
&= \sum_{t=1}^n \sum_{j=2}^N \left( \log \sigma_{jt}^2 + \frac{\varepsilon_{jt}^2}{\sigma_{jt}^2} \right) \\
&= \sum_{j=2}^N \left\{ \sum_{t=1}^n \left( \log \sigma_{jt}^2 + \frac{\varepsilon_{jt}^2}{\sigma_{jt}^2} \right) \right\}.
\end{aligned} \tag{29}$$

The last representation is the most convenient to use since it can be viewed as a sum of  $(N - 1)$ , univariate Gaussian likelihoods for the mutually uncorrelated returns  $\{\varepsilon_{jt}\}, j = 2, \dots, N$ . Note that from the matrix form of (9) we have  $\varepsilon_t = T_t Y_t$ , so that  $\varepsilon_{jt}$ 's do depend on the GARPs in  $T_t$ . Though we could begin MLE computation with the simple case of constant GARPs ( $T_t \equiv T$ ) and GARCH (1,1) for time-varying innovation variances  $\{\nu_t\}$  of the standardized residuals, still the number of parameters to be estimated is  $N(N - 1)/2 + 3N$  which is large for  $N = 100$  or  $500$ . This can be reduced considerably by selecting  $T$  from among Examples 1-4 and  $\{\sigma_{jt}^2\}$  a time-varying version of (25); for example, in Section (5) we reduce a full lower diagonal matrix with 630 entries to a parsimonious representation with two parameters based on model (18).

However, it is important to examine closely a rather appealing property of the constant GARPs model that is not available in any other multivariate time series models with time-varying innovation variances. The decomposition (10) leads to univariate Gaussian likelihoods where the estimation of the  $N(N - 1)/2$  parameters in  $T$  can be achieved by estimating independently  $N$  univariate regression models with time-varying innovation variances. For example, if a GARCH(1,1) process is assumed, the estimation problem consists of just estimating  $N$  regression GARCH(1,1) models. This property is not available in any other model in the context of multivariate time-varying correlations. In the Orthogonal GARCH models the estimation of the  $A_t \equiv A$  and  $V_t$  matrices in (8)

is achieved in two steps, the matrix  $A$  being estimated first as if  $V_t$  was time-independent, followed by estimation of  $V_t$  which is based on the resulting  $Z_t = AY_t$ . Similarly, Vrontos *et al.*(2004) focused on representation (11) and they estimated simultaneously the GMAP and IV parameters, but they did not exploit the fact that the GARP parameters  $T = B^{-1}$  could be estimated faster via (29). From a practical perspective, constant GARP structures provide an easy and very competitive alternative to the existing multivariate models presented in Section 2, allowing quick and easy estimation for larger values of  $N$ .

Of course, in view of (27), the correlation parameters  $\rho$  in (29) can be partitioned into two parts, the first corresponding to  $\{\nu_t\}$  is denoted by the vector  $\lambda$  and other corresponding to  $\{T_t\}$  is denoted by the vector  $\phi$ . From the definition of  $\varepsilon_{jt}$ , it is evident that the likelihood (29) is a quadratic function of the GARPs  $\phi$ . Thus, for given IV parameters  $\lambda$ , MLE of  $\phi$  has a closed-form. Details of an algorithm for finding MLE of  $\rho = (\lambda, \phi)$  and the asymptotic properties of the estimators can be found in Pourahmadi (2000, Sec.2).

## 5 A real data example

### 5.1 Large datasets

We report results from fitting AR structures (9) with unstructured GARP parameters and GARCH(1,1) dynamics for innovation variances to a dataset with large  $n$  and  $N$ . The first consists of 2780 daily returns of 36 stocks; for a detailed description of these stocks see Han (2002). His extensive empirical study illustrates the economic advantages of employing multivariate rather than univariate modelling of returns; the importance of considering conditional covariance matrices, rather than

univariate conditional moments, in order to improve standard asset market-pricing theory such as the capital asset pricing model, has been discussed by Bollerslev, Engle & Wooldridge (1988). The second consists of 2274 daily returns of 100 sector indices; for details see Engle & Sheppard (2001).

Both the exploratory and inferential parts of our analysis require an ordering of the  $N$  assets. Since these models are primarily useful for forecasting, we will not implement ways to solve the ordering problem based on model fit criteria. In Bayesian literature this exact problem has been attacked by Webb and Forster (2008) who presented an efficient reversible jump algorithm that searches over all models with different orderings. This could be readily implemented here by applying a Laplace approximation to each row of (9) as in Vrontos *et al.*(2003). A searching algorithm based on some criterion such as AIC or BIC can be also readily constructed by noting that the number of comparisons required is of order  $N^2$  and not of order  $N!$ : There are  $N$  possible ways to write (9) for  $j = N$ , and conditional on the best model out of  $N$ , there are  $N - 1$  ways to write (9) for  $j = N - 1$ , and so forth.

The ordering used here is based on the sample marginal volatility, with the less variable stock/index being the first, we also explore the impact of ordering on forecasting. An interesting aspect of the AR structures with unstructured GARP is that estimates of  $T$  and  $\{\nu_t\}$  can be obtained readily as follows: First perform the modified Cholesky decomposition (10) of the sample covariance matrix of the data to obtain  $T$  or the initial estimates of the  $\phi_{jk}$  parameters. Then, construct  $N$  innovation processes  $\{T_j Y_t\}_{t=1}^2$ ,  $j = 1, \dots, N$ , where  $T_j$  is the  $j$ th row of  $T$ , and model each as a univariate GARCH(1,1) to estimate  $(\alpha_{j0}, \alpha_{j1}, \beta_{j1})$ . In our examples, we found that the  $\phi_{jk}$ 's estimated in this way are quite close to their final maximum likelihood estimates; see Figures

1 and 2. These estimates are used to construct regressograms providing visual insight into the order  $p$  of the AR structure (see Example 4). Figure 3 depicts the subdiagonals and rows of  $T$  and  $B$  versus their indices for the 36-dimensional dataset. These simple graphs are capable of suggesting parsimonious models like polynomials to be fitted to each subdiagonal or row to reduce further the number of parameters in  $T$  and  $B$ . For example, the regressogram (Pourahmadi, 1999) on the upper left suggests a linear model of the form (18) with only two parameters reducing the parameters of  $T$  from  $N(N - 1)/2 = 630$  to only 2. Estimation of these parameters can be achieved by maximizing the log-likelihood function iteratively between the GARCH(1,1) and  $(\gamma_0, \gamma_1)$  parameters. The resulting estimated equation is

$$\phi_j = -0.0754 + 0.0025j, j = 1, \dots, 35.$$

To be more realistic, we also added a  $N \times 1$  parameter vector  $\mu$ , for the mean so our final model was  $T(Y - \mu) = \varepsilon$ , with  $N(N - 1)/2 + 4N$  parameters. The program took 6.3 and 85 minutes to run on a PENTIUM 4 PC at 1.7 MHz for the two datasets, respectively. The program was written in MATLAB; it is expected that use of a lower-level language can speed the program by a factor of at least 10.

Formal tests, such as AIC and BIC can be adopted to test whether certain subdiagonals or rows of  $T$  are zero (Wu and Pourahmadi, 2003). In our datasets, all AIC tests rejected the hypothesis that a row  $T_j$  is zero, but some BIC tests provided evidence, at the 5% significance level, that 5 rows of the  $36 \times 36$  matrix of the 36-dimensional dataset and 7 rows of the  $100 \times 100$  matrix of the 100-dimensional dataset can be set to zero. For illustration, Figure 4 depicts the last 20 rows of  $T$  in the larger dataset; the rows that BIC indicated to be set to zero are 89, 90, 94, 97, 98, 99, 100,

achieving a model with 660 fewer parameters.

## 5.2 Forecasting power

The importance of multivariate time-varying volatility models can be only judged by examining closely their forecasting power. We concentrate here on both the forecasting variability caused by (order) permuting the response vector, and the comparison against 5 existing multivariate volatility models. The 3 designed experiments that follow focus, respectively, on comparison of forecasts against some reliable proxy, on their ability to calculate Value at Risk (VaR), and on how well they can be used to construct a portfolio. The data and models used are as follows.

We obtained (source: DATASTREAM) 7 daily exchange rates of the USA dollar against UK pound, EURO, Swedish korona, Australian dollar, Canadian dollar, swiss franc and Japanese YEN, recorded from 2/1/1999 up to 28/10/2003. We also obtained (source: ROYTERS) two-minute intraday data from the same exchange rates for the following 3 days 29-31/10/2003. From the 7 series, we created all  ${}^7C_5 = 21$  combinations of 5 exchange rates, and used them as replications of our experiment. Finally, we used for the purposes of comparison the following five widely used multivariate models:

- (i) The multivariate diagonal-vec model of order (1,1); see Bollerslev, Engle and Wooldridge (1988).
- (ii) The matrix-diagonal model of order (1,1); see Bollerslev, Engle and Nelson (1994).
- (iii) An exponentially weighted moving average model of the form

$$\Sigma_t = \alpha(\varepsilon_t \varepsilon_t') + (1 - \alpha)\Sigma_{t-1},$$

where  $0 \leq \alpha \leq 1$  and  $\varepsilon_t$  is the vector of shocks. Note that RiskMetrics uses this model where the smoothing parameter  $\alpha$  is not estimated, but set to 0.06.

(iv) The constant conditional correlation of order (1, 1); see Borreslev (1990).

(v) The orthogonal GARCH model of order (1, 1); see Alexander (2001).

### 5.2.1 Comparison against a proxy

For each of the 21 replications of the exchange rates, and for all  $5! = 120$  possible orderings within each replication, we predicted the conditional covariance matrices for the next three days of 29-31/10/2003 using the AR structure with unconstrained GARCHs and the five multivariate models (i)-(v) mentioned above.

Although the true realized covariance matrix is unavailable, recent developments in the analysis of realized covariation (Andersen, Bollerslev and Lange, 1999; Barndorff-Nielsen and Shephard, 2004) allow us to replace it by a reliable proxy, the realized covariation matrix. The realized covariation matrix with elements  $\sigma_{ij}$ , is calculated for each of the three days as the cumulative cross-products of intraday returns over each day. Following Ledoit, Santa-Clara and Wolf (2003), for corresponding forecasts  $\sigma_{ij}^*$  derived from the daily data series the following two measures of forecasting performance will be used here:

**Mean absolute deviation;**

$$MAD = N^{-2} \sum_{i,j} E|\sigma_{ij}^* - \sigma_{ij}|.$$

**Root mean square error;**

$$RMSE = \left[ N^{-2} \sum_{i,j} E \left( \sigma_{ij}^* - \sigma_{ij} \right)^2 \right]^{1/2} .$$

Figures 5-10 present the results of our experimental study. Our suggested ordering based on the sample marginal volatility, performs well compared to the other existing models. When we average MAD and RMSE over all days and all 21 datasets, only the constant conditional correlation model beats our model in both MAD and RMSE, whereas the matrix-diagonal models beats RMSE but loses to MAD; all other models have larger MAD and RMSE than our model. We feel that these results are very promising, since in portfolios that contain financial products more diverse than just exchange rates the constant correlation model will fail to capture the empirical dynamics of all series, whereas in larger dimensions only the exponential moving average and the orthogonal GARCH models have the ease of estimation offered by the full AR structures. Moreover, the range of predictions obtained by all orderings is similar to the range of the competing forecasts.

Surprisingly, the AR structures attaining the minimum MAD and RMSE provide forecasts that outperform the 5 competing models. This inevitably calls attention to developing feasible computational techniques for specifying this optimal ordering when  $N$  is sufficiently large like 500 or more. A method to choose the “best” among a series of models based on MCMC model searching techniques for such an application in multivariate GARCH modelling for moderate  $N$  is given in Vrontos, Dellaportas and Politis (2000).

### 5.2.2 Value at risk

Another way to assess the forecasting power of our proposed models is to measure their ability to calculate Value-at-Risk (VaR). Assume that at time  $t$  we obtain an estimate of the conditional covariance matrix  $\Sigma_{t+1}$  and that we have portfolio of  $N$  exchange rates with weights  $w_N$ , so the portfolio's estimated conditional variance at time  $t + 1$  is  $\Sigma_{t+1}^* = w_N' \Sigma_{t+1} w_N$ . We follow closely Ledoit, Santa-Clara and Wolf (2003) and calculate one-day-ahead VaR at 1% level as follows: first fit a t-distribution to the past portfolio returns and their estimated conditional variances by estimating the degrees of freedom of the t-distribution that maximizes the resulting likelihood function. Then, calculate the 1% VaR at time  $t + 1$ ,  $VaR_{t+1}$ , and define the hit variable  $hit_{t+1} = I\{w_N' Y_{t+1} < VaR_{t+1}\}$ . Finally, we test using the asymptotic Chi-square approximation whether the series  $hit_t$  is uncorrelated over time and has expected value equal to the desired confidence level; for more details see Ledoit, Santa-Clara and Wolf (2003). We used the first four exchange rates and predicted VaR for days 1001 – 1224 using an AR structure and the models (i)-(v) described at the start of Section 5.2. All p-values and sampled means (hit rates) are shown in Table 1. The results indicate that all models except the orthogonal GARCH do a good job in predicting the VaR.

### 5.2.3 Portfolio selection

Finally, an important application of time-varying covariance matrix estimation is in portfolio selection. We calculated, for days 1001 – 1224, the time-varying weights of the minimum variance portfolio for all models and data of Section 5.2.2. If at time  $t$  the predicted conditional variance is

	Mean hit rate	p-value	SD of portfolio
AR structure	0.013	0.52	0.0047
Diagonal-vec	0.009	0.50	0.0047
Matrix-diagonal	0.013	0.52	0.0048
Moving-average	0.018	0.54	0.0047
Constant conditional correlation	0.009	0.50	0.0047
Orthogonal GARCH	0.094	0.83	0.0053

Table 1: Performance of various models for calculating Value-at-Risk and standard deviation (SD) of portfolio returns

$\Sigma_{t+1}$ , the weights are given by

$$w_t = \frac{\Sigma_{t+1}^{-1} \iota}{\iota' \Sigma_{t+1}^{-1} \iota}$$

where  $\iota$  is a  $N \times 1$  vector of ones. Based on these weights, we can then compare the realized standard deviation (SD) of the returns of the conditional minimum variance portfolio over the period of the last 224 days. The results are presented in Table 1 where it is clear that the AR structure performs well against the competitors.

It is interesting to investigate the trade-off between parsimony and performance of the volatility predictions using unstructured covariance matrices. For the 36-dimensional data set of Section 5.1 and the estimated model (18) with  $k = 1$ , Figure 11 depicts the realized standard deviation of the returns of the conditional minimum variance portfolio over the period of the last 20 days, based on estimates for the first 2760 days. It is surprising that the portfolio based on the (highly!)

parsimonious model produces an equally good, if not better, minimum variance portfolio, when compared with the model with the extra 628 parameters.

## 6 Conclusions

Our preliminary empirical work on a portfolio of 100 stocks shows the promise of the proposed methodology in providing parsimonious models for conditional covariances while guaranteeing the positive-definiteness of their estimators. Detailed empirical results from an experimental study based on exchange rates indicates that ordering of the stocks does not alter the forecasting power of our proposed models. More work is needed to compare the performance of our methodology with, say, the recent work of Engle and Sheppard (2001) on dynamic conditional correlation multivariate GARCH models.

## Acknowledgements

We would like to thank Y. Han and Kevin Sheppard who provided the datasets used in Section 5.1. Research of the second author was supported in part by the NSF grant DMS-0307055.

## References

- Aguilar, O. & West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, **18**, 338-357.
- Alexander C. (2001). *Market Models: A Guide to Financial Data Analysis*. Wiley, New York.

- Andersen T.G., Bollerslev and Lange S. (1999). Forecasting Financial Market Volatility: Sample Frequency Vis-à-vis Forecast Horizon. *Journal of Empirical Finance*, **6**, 457-477.
- Barndorff-Nielsen O.E. and Shephard N. (2004). Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. To appear in *Econometrica*.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, **72**, 498-505.
- Bollerslev, T., Engle, R. & Nelson, D. (1994). ARCH Models. In *Handbook of Econometrics*, R. Engle and D. McFadden (editors), pp. 2959-3038.
- Box, G.P., Jenkins, G.M. & Reinsel, G.C. (1994). *Time Series Analysis - Forecasting and Control* (Revised 3rd Edition). Holden-Day, San Francisco, California.
- Brockwell, P.J. & Davis, R.A. (1991). *Time Series: Theory and Methods*, Second Ed., Springer-Verlag, New York.
- Campbell, J., Lo, A. & MacKinlay, A. (1997). *The Econometrics of Financial Markets*. New Jersey, Princeton University Press.
- Chib, S., Nardari, F. & Shephard, N. (2001). Analysis of high dimensional multivariate stochastic volatility. Preprint.

- Christodoulakis, G.A. & Satchell, S.E. (2000), Evolving systems of financial returns: Autoregressive conditional beta. <http://www.staff.city.uk/gchrist/research/research.html>
- Diebold, F.X. & Nerlove, M. (1989). The dynamics of exchange rate volatility: A multivariate latent factor ARCH model, *Journal of Applied Econometrics*, **4**, 1-21.
- Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987-1008.
- Engle, R.F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, **20**, 339-350.
- Engle, R.F. & Kroner, K.F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, **11**, 122-150.
- Engle, R.F., NG, V.K. & Rothchild, M. (1990). Asset pricing with a factor-ARCH covariance structure: Empirical estimates for Treasury bills. *Journal of Econometrics*, **45**, 213-237.
- Engle, R.F. & Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. preprint.
- Geweke, J.F. & Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies*, **9**, 557-587.
- Golub, G.H., Van Loan, C.F. (1990), *Matrix Computations*. Johns Hopkins, 2nd Ed.

- Han, Y. (2002). The economic value of volatility modeling: asset allocation with a high dimensional dynamic latent factor multivariate stochastic volatility model. preprint.
- Harvey, A.C., Ruiz, E. & Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247-264.
- Joe, H. (2006). Generating random correlation matrices based on partial correlations. *J. of Multivariate Analysis*, **97**, 2177-2189.
- Kim, S., Shephard, N. & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **65**, 361-393.
- Ledoit, O., Santa-Clara, P., and Wolf, M. (2003). Flexible multivariate GARCH modeling with an application to international stock markets. *Review of Economics and Statistics*, **85**, 735-747.
- Lopes, H.F., Aguilar, O. and West, M. (2002). Time-varying covariance structures in currency markets. Preprint.
- O'Hagan, A. (1994). *Bayesian Inference*. Edward Arnold, Great Britain.
- Pitt, M.K. & Shephard, N. (1999a). Time-varying covariances: A factor stochastic volatility approach (with discussion), in *Bayesian Statistics 6*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith, Oxford, UK, Oxford University Press, 547-570.
- Pitt, M.K. & Shephard, N. (1999b). Filtering via simulation based on auxiliary particle filters. *Journal of American Statist. Assoc.*, **94**, 590-599.

- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, **86**, 677-690.
- Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, **87**, 425-435.
- Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*, John Wiley, New York.
- Pourahmadi, M. (2002). Graphical diagnostics for modeling unstructured covariance matrices. *International Statistical Review*, **70**, 395-417.
- Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters. *Biometrika*, **94**, 1006-1013.
- Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, **48**, 1-48.
- Smith, M. & Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *J. of Amer. Statist. Assoc.*, **97**, 1141-1153.
- Tsay, R. (2002). *Analysis of Financial Time Series*. John Wiley, New York.
- Tse, Y.K. & Tsui, A.K. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *J. of Business and Economics Statistics*, **20**, 351-362.
- Vrontos, I.D., Dellaportas, P. & Politis, D.N. (2003). A full-factor multivariate GARCH model. *Econometrics Journal*, **6**, 312-334.

- Vrontos, I.D., Dellaportas, P. & Politis, D. (2000). Full Bayesian inference for GARCH and EGARCH models. *Journal of Business and Economics Statistics*, **18**, 187-198.
- Webb E.L. and Forster J.J. (2008). Bayesian model determination for multivariate ordinal and binary data. *Computational Statistics and Data Analysis*, **52**, 2632-2649.
- Wu, W.B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data, *Biometrika*, **90**, 831-844.
- Zellner, A. (1979). An error-components procedures (ECP) for introducing prior information about covariance matrices and analysis of multivariate regression models. *International Economic Review*, **20**, 201-214.

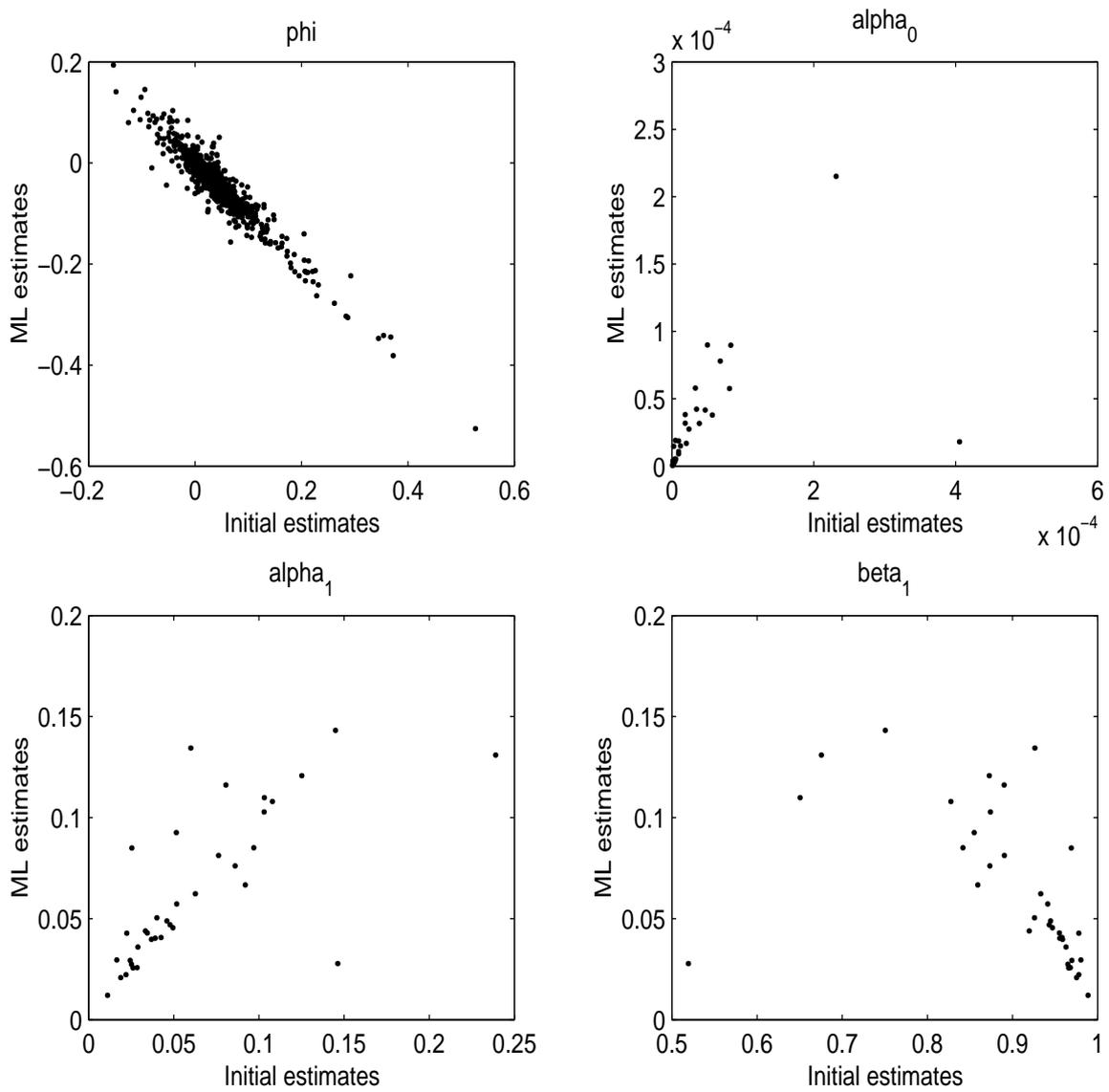


Figure 1: 36-dim data: Comparison of initial and maximum likelihood estimates

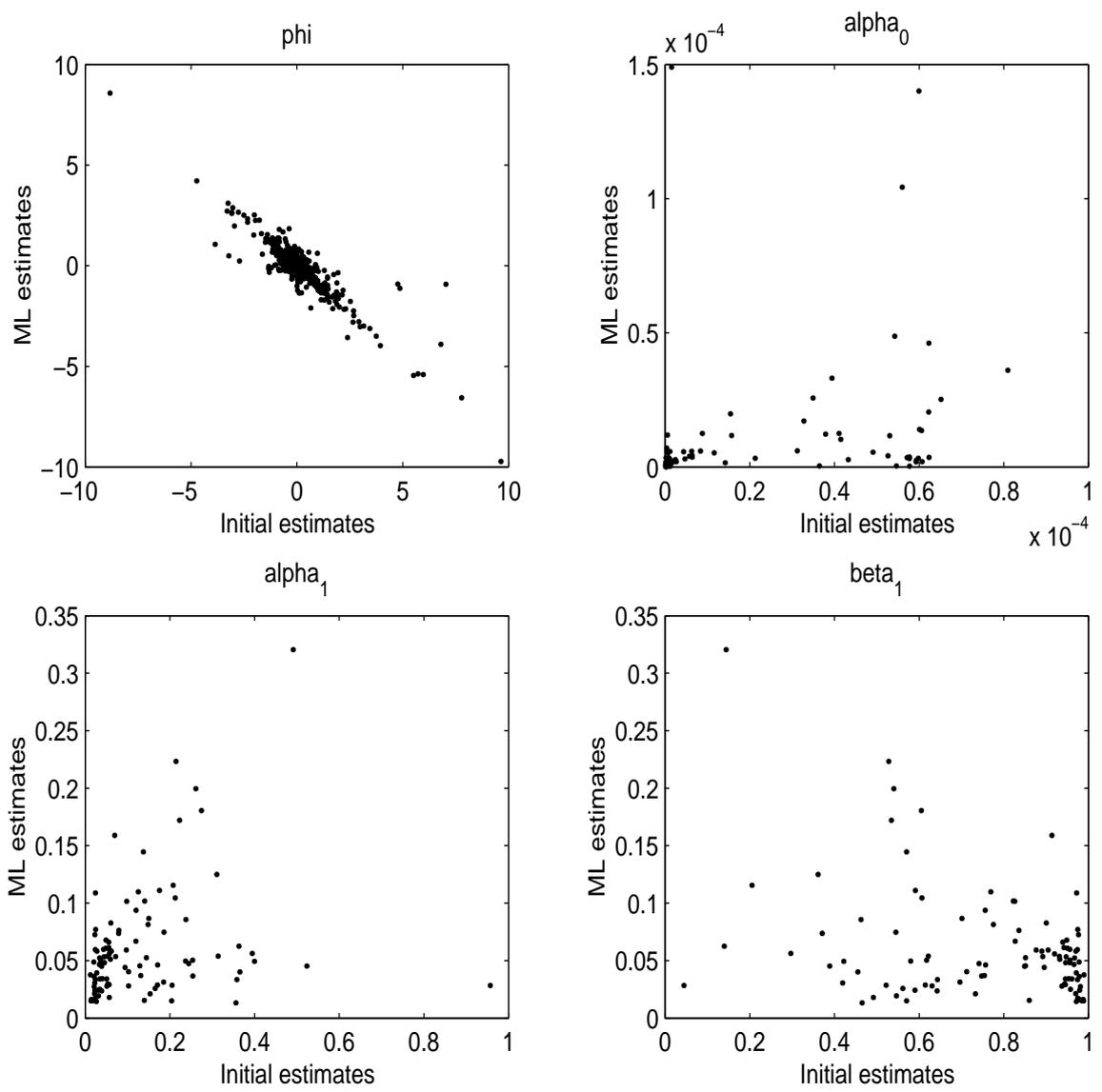


Figure 2: 100-dim data: Comparison of initial and maximum likelihood estimates

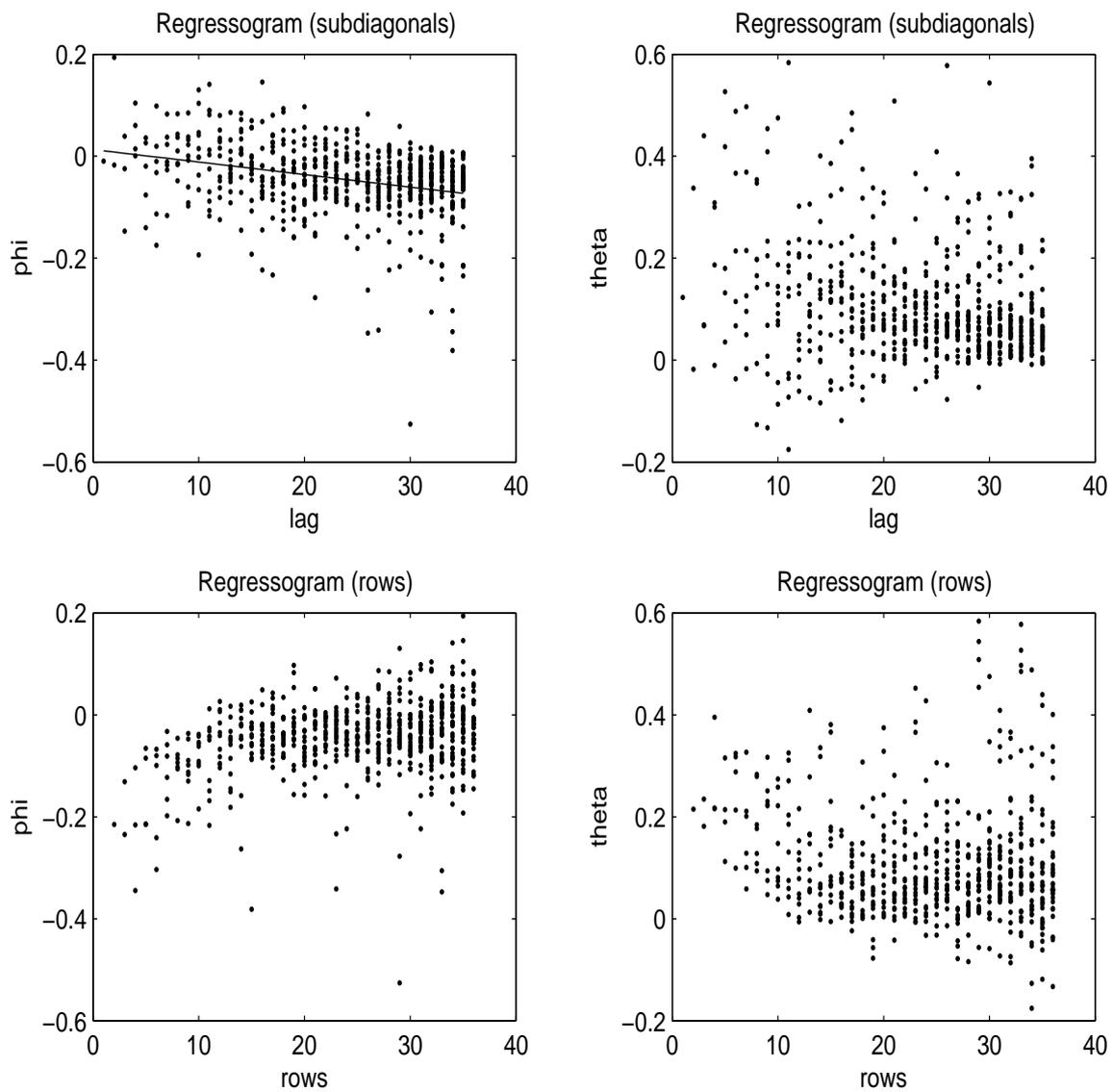


Figure 3: 36-dim data: Regressograms: Plots of subdiagonals and rows of  $T$  and  $L$  vs their index

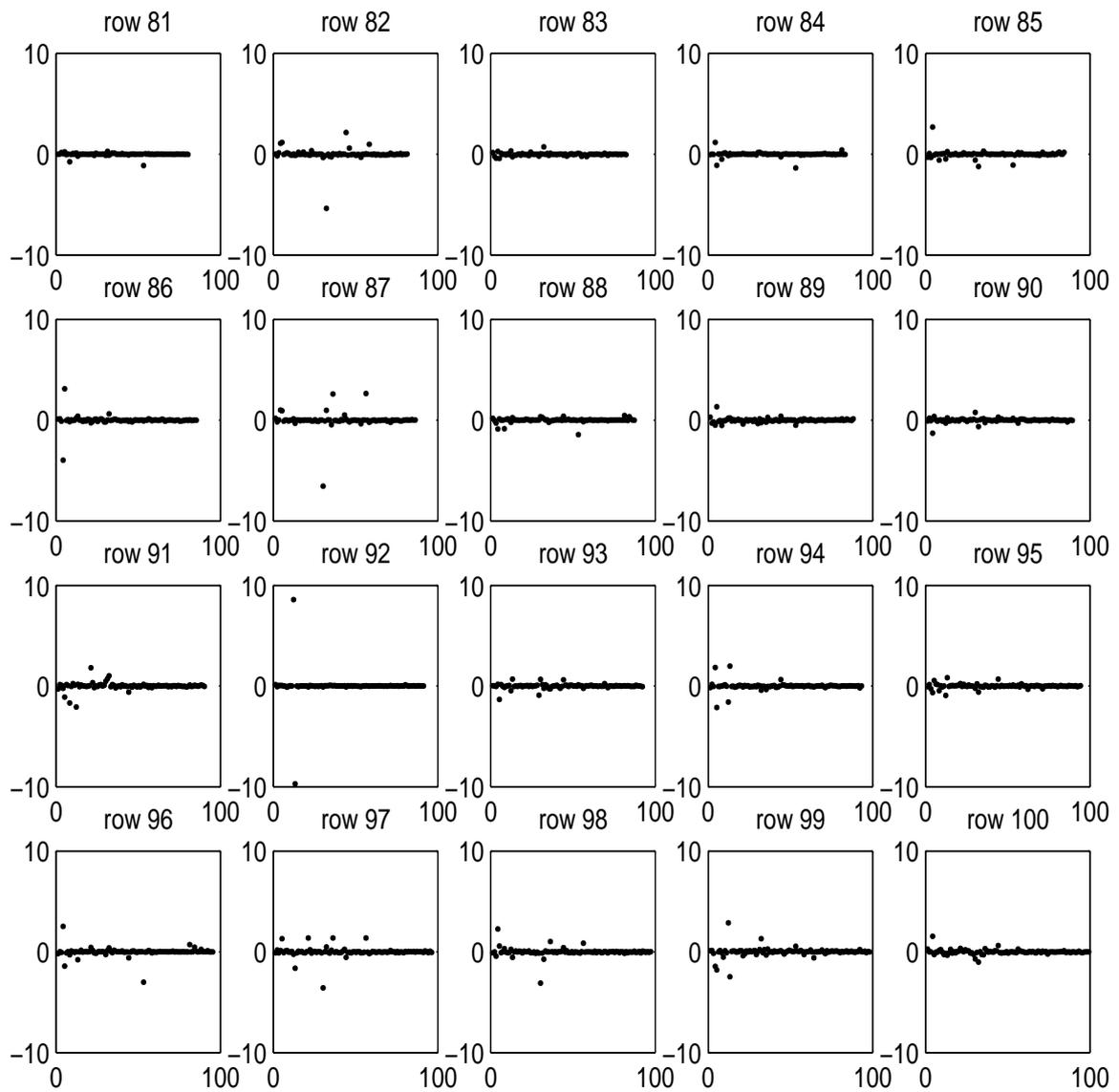


Figure 4: 100-dim data: Plots of the last 20 rows of  $T$

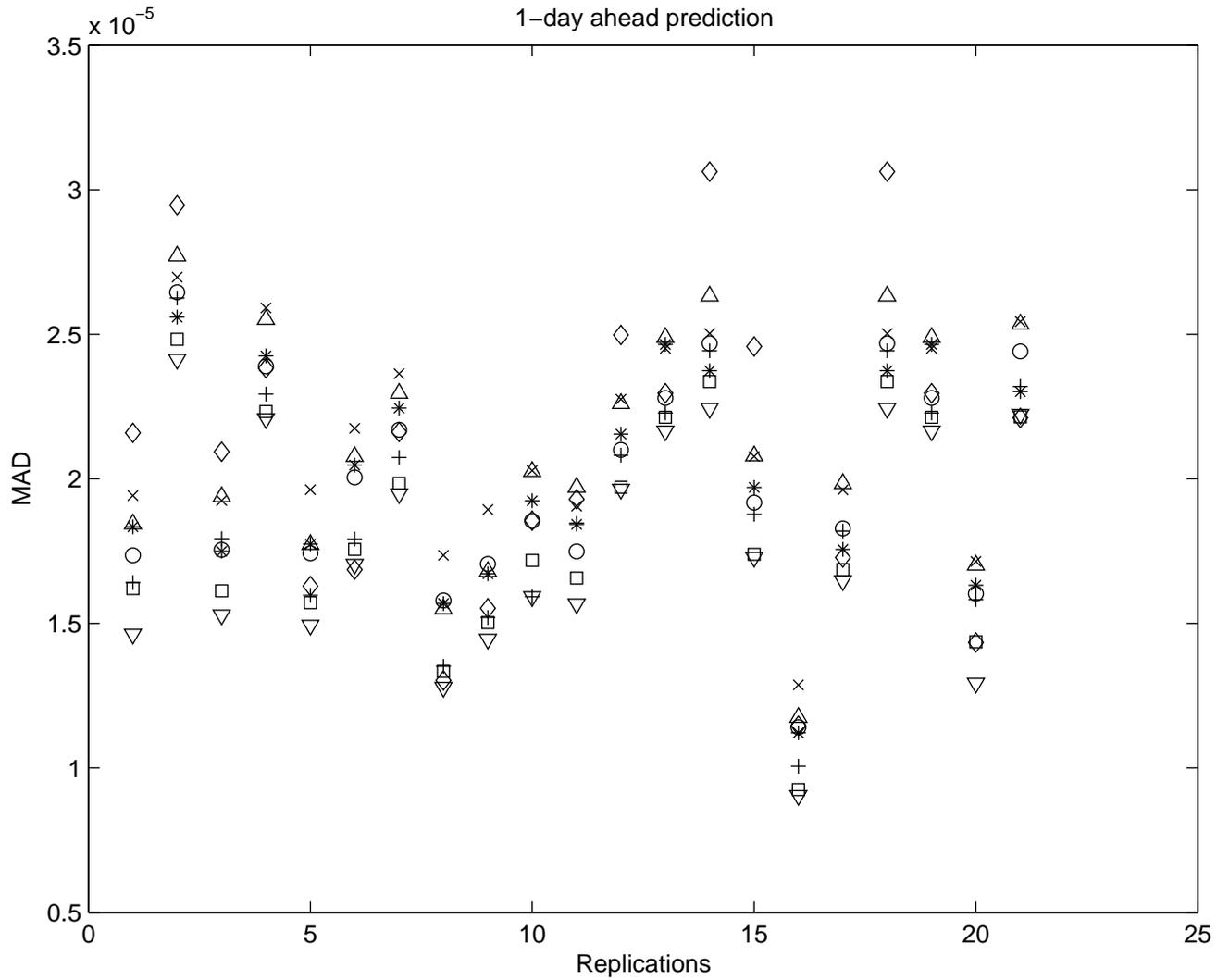


Figure 5: Mean absolute deviation (MAD) for 5-dimensional exchange rates. *circle*: diagonal-vec; *asterisk*: matrix-diagonal; *cross*: weighted moving average; *square*: constant conditional correlation; *diamond*: orthogonal GARCH; *upward and downward pointing triangles*: maximum and minimum MAD over the 120 full structure AR models produced by all possible orderings of the exchange rates; *plus sign*: our suggested ordering based on ordering the unconditional variance.

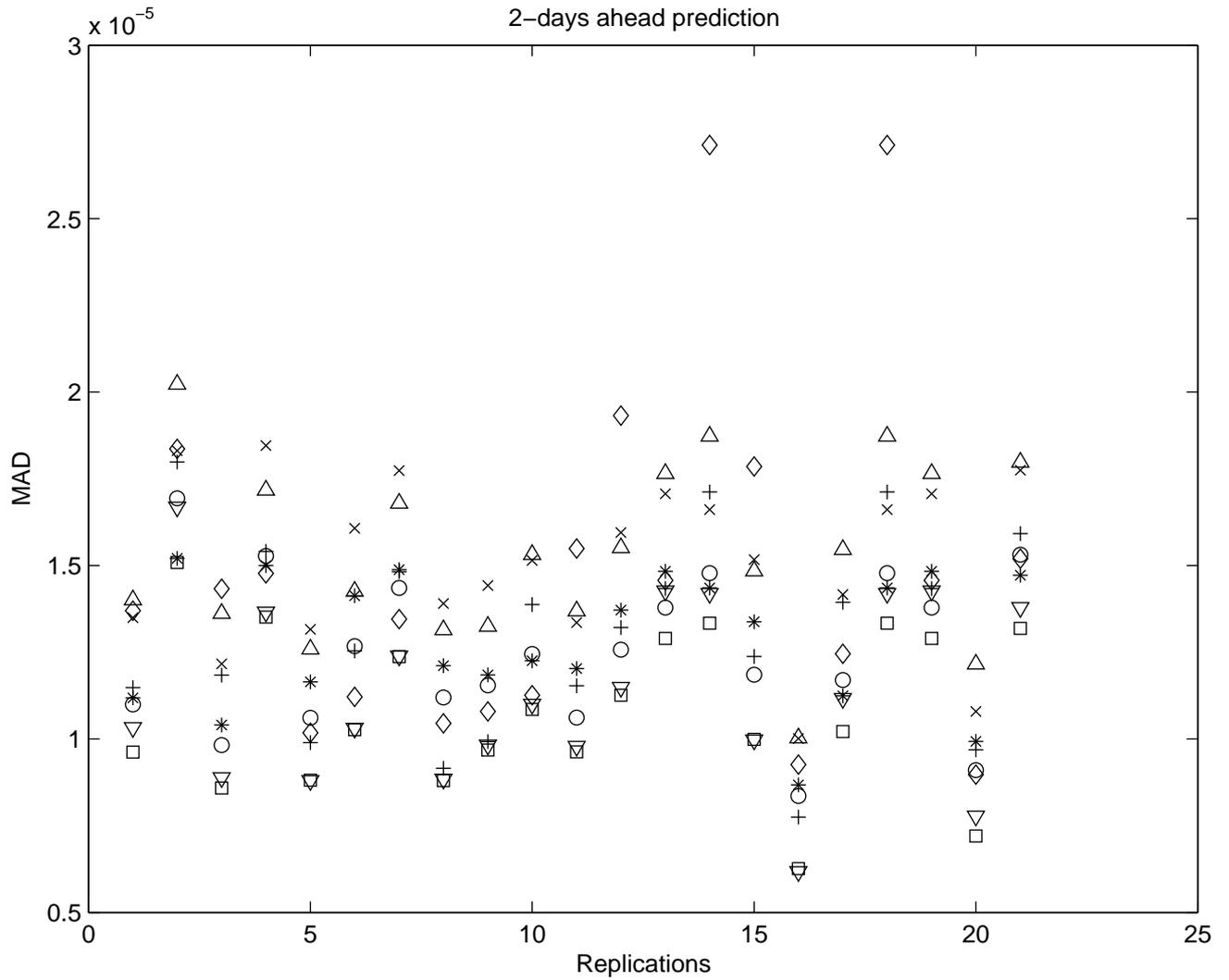


Figure 6: Mean absolute deviation (MAD) for 5-dimensional exchange rates. *circle*: diagonal-vec; *asterisk*: matrix-diagonal; *cross*: weighted moving average; *square*: constant conditional correlation; *diamond*: orthogonal GARCH; *upward and downward pointing triangles*: maximum and minimum MAD over the 120 full structure AR models produced by all possible orderings of the exchange rates; *plus sign*: our suggested ordering based on ordering the unconditional variance.

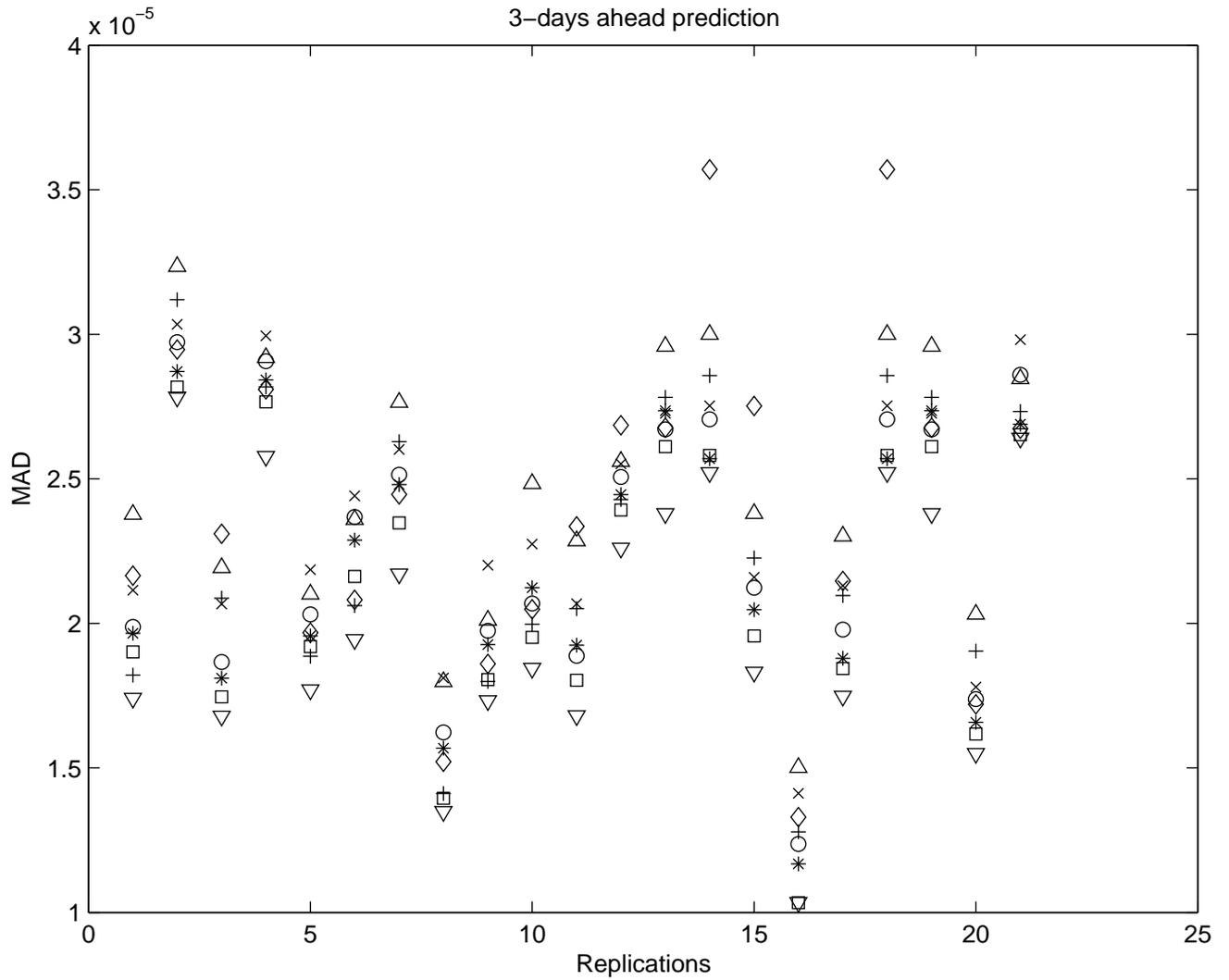


Figure 7: Mean absolute deviation (MAD) for 5-dimensional exchange rates. *circle*: diagonal-vec; *asterisk*: matrix-diagonal; *cross*: weighted moving average; *square*: constant conditional correlation; *diamond*: orthogonal GARCH; *upward and downward pointing triangles*: maximum and minimum MAD over the 120 full structure AR models produced by all possible orderings of the exchange rates; *plus sign*: our suggested ordering based on ordering the unconditional variance.

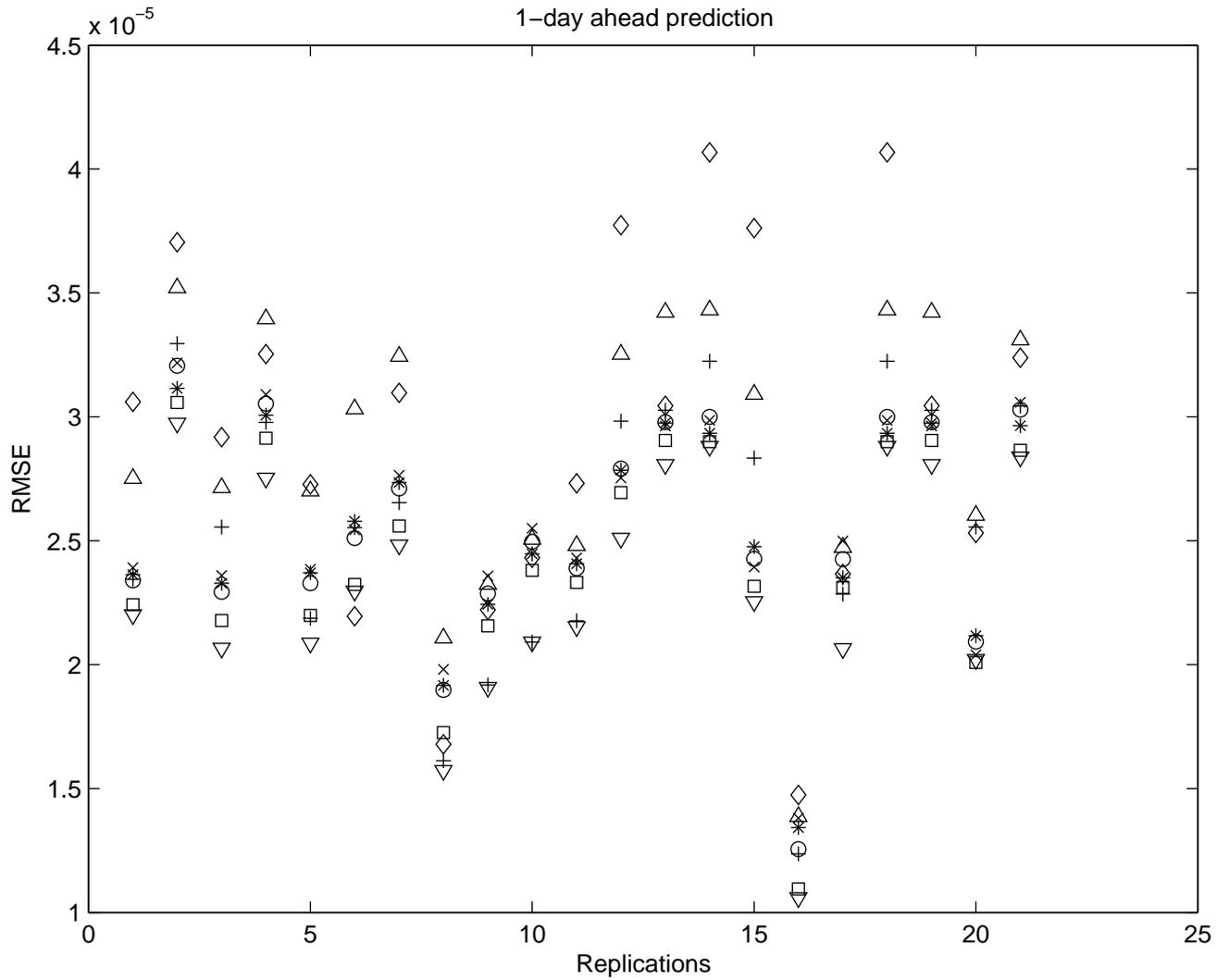


Figure 8: Root mean square error (RMSE) for 5-dimensional exchange rates. *circle*: diagonal-vec; *asterisk*: matrix-diagonal; *cross*: weighted moving average; *square*: constant conditional correlation; *diamond*: orthogonal GARCH; *upward and downward pointing triangles*: maximum and minimum RMSE over the 120 full structure AR models produced by all possible orderings of the exchange rates; *plus sign*: our suggested ordering based on ordering the unconditional variance.

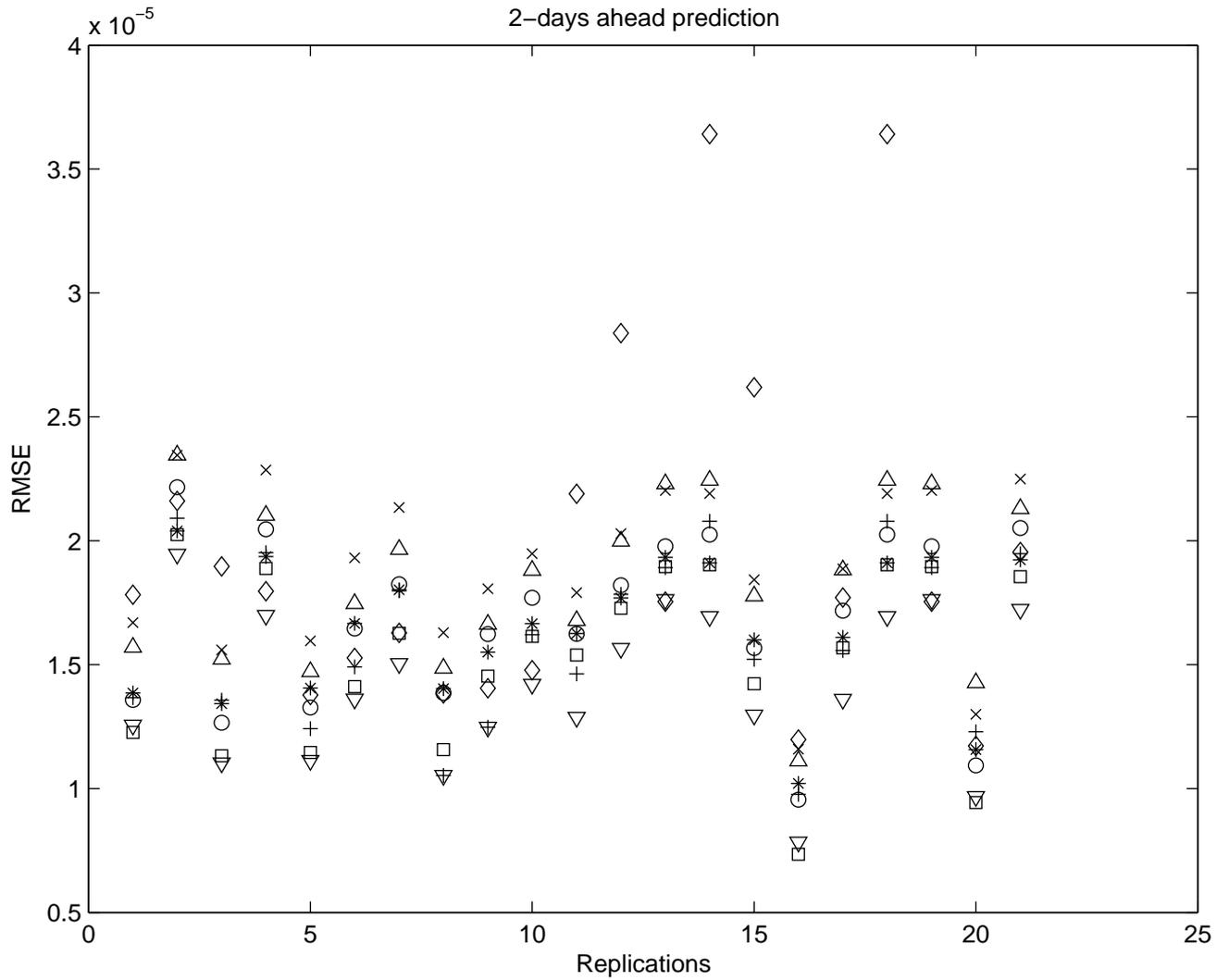


Figure 9: Root mean square error (RMSE) for 5-dimensional exchange rates. *circle*: diagonal-vec; *asterisk*: matrix-diagonal; *cross*: weighted moving average; *square*: constant conditional correlation; *diamond*: orthogonal GARCH; *upward and downward pointing triangles*: maximum and minimum RMSE over the 120 full structure AR models produced by all possible orderings of the exchange rates; *plus sign*: our suggested ordering based on ordering the unconditional variance.

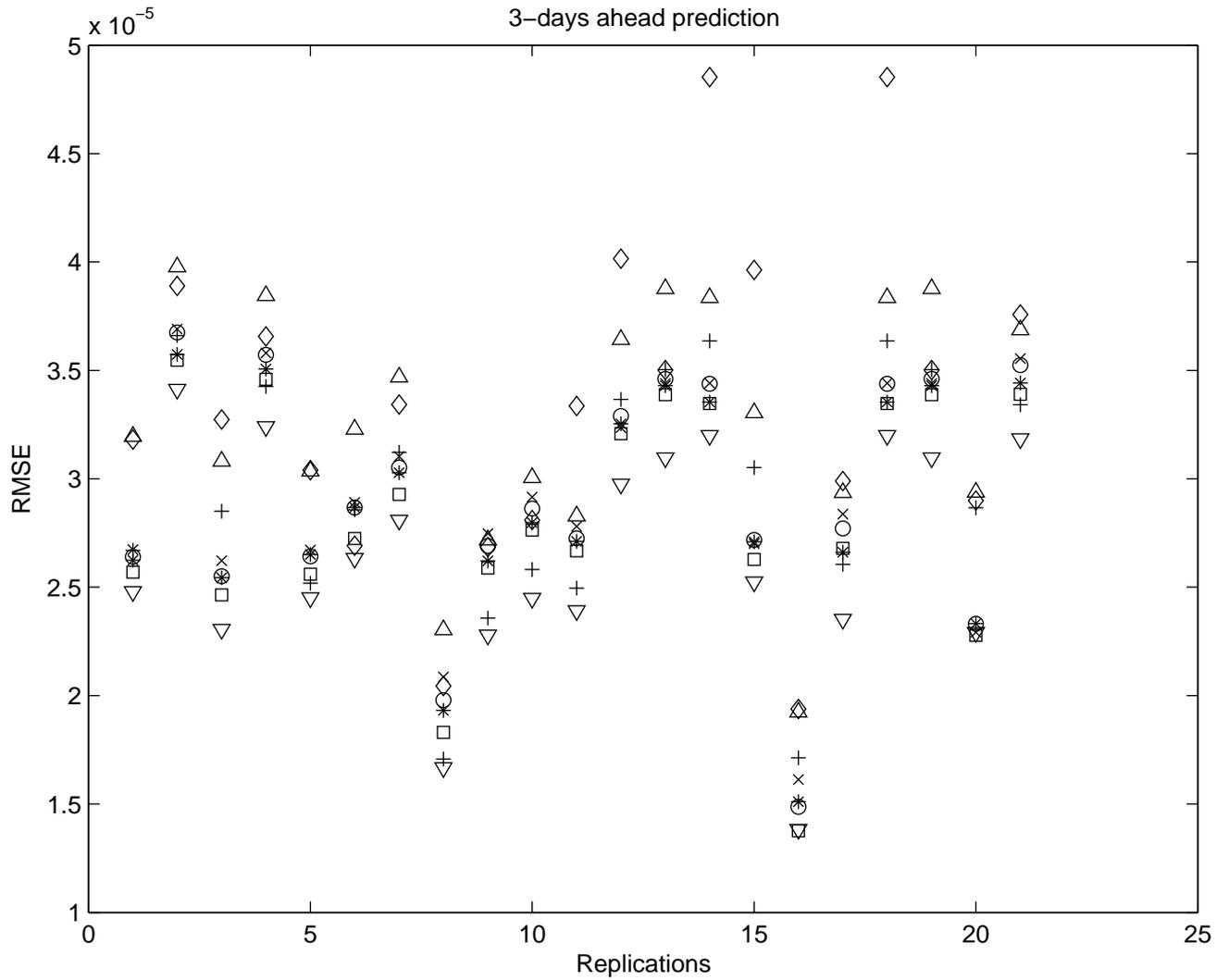


Figure 10: Root mean square error (RMSE) for 5-dimensional exchange rates. *circle*: diagonal-vec; *asterisk*: matrix-diagonal; *cross*: weighted moving average; *square*: constant conditional correlation; *diamond*: orthogonal GARCH; *upward and downward pointing triangles*: maximum and minimum RMSE over the 120 full structure AR models produced by all possible orderings of the exchange rates; *plus sign*: our suggested ordering based on ordering the unconditional variance.

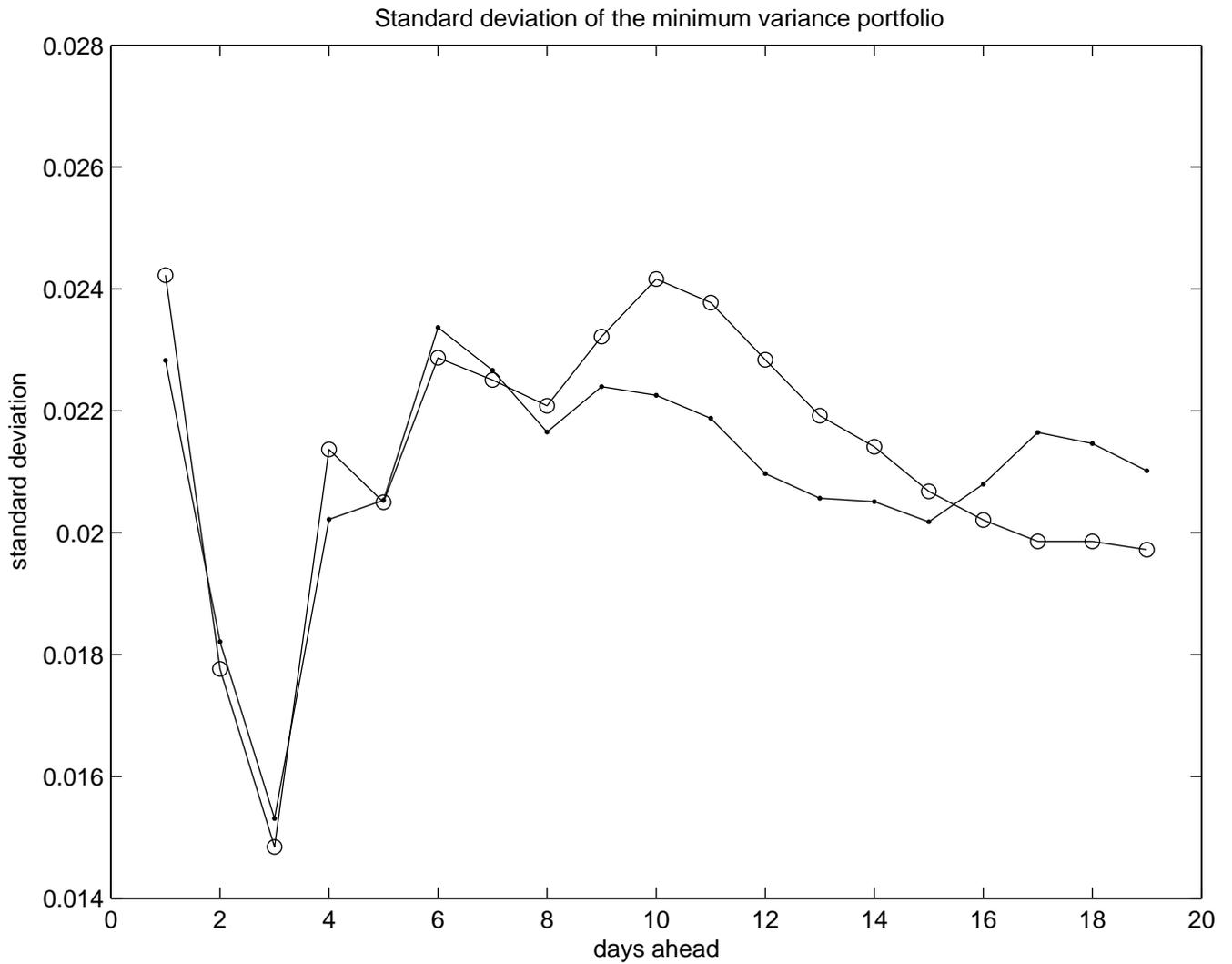


Figure 11: Realized standard deviation of the returns of the conditional minimum variance portfolio over the period of the last 20 days for the 36-dim data; *dots*: A parsimonious Toeplitz GARP model with constant parameters  $\phi_j$  in each subdiagonal  $j$  given by the model of the form  $\phi_j = \alpha + \beta j$ ; *circles*: A GARP model.