Constrained Nonlinear Programming for Volatility Estimation with GARCH Models*

Aslıhan Altay-Salih[†]
Mustafa Ç. Pınar[‡]
Sven Leyffer[§]

Abstract. This paper proposes a constrained nonlinear programming view of generalized autoregressive conditional heteroskedasticity (GARCH) volatility estimation models in financial econometrics. These models are usually presented to the reader as unconstrained optimization models with recursive terms in the literature, whereas they actually fall into the domain of nonconvex nonlinear programming. Our results demonstrate that constrained nonlinear programming is a worthwhile exercise for GARCH models, especially for the bivariate and trivariate cases, as they offer a significant improvement in the quality of the solution of the optimization problem over the diagonal VECH and the BEKK representations of the multivariate GARCH model.

Key words. time series econometrics, constrained nonlinear programming, multivariate GARCH, volatility estimation, maximum likelihood estimation

AMS subject classifications. 90C30, 90C90, 62M10, 91B84

DOI. 10.1137/S0036144501400115

1. Introduction. Volatility plays an important role in several areas of current finance literature. It is central to portfolio selection models as efficient portfolios are formed by computing the maximum return for a given level of volatility. Equilibrium models like the capital asset pricing model (CAPM) require the estimation of market variance as well as the covariance of risky assets with the market portfolio. Prices of options are also expressed as functions of volatility. As a result, volatility and covariance estimation is an important research area for both academicians and practitioners.

ARCH (autoregressive conditional heteroskedasticity; Engle (1982)) and GARCH (generalized ARCH; Bollerslev (1986)) volatility forecasting models have been the major tool for characterizing volatility, by using past unpredictable changes in the returns of an asset to predict the future time-varying second-order moments. Volatility clustering phenomena (Mandelbrot (1963), Fama (1965)) are the driving force for the

GARCH family of models. The success of these models in the univariate case for volatility estimation has inspired an interest in covariance estimation, which is a harder problem, and has led to the development and application of the multivariate extensions.[1] The major difficulty in the multivariate case stems from the highly nonlinear and nonconvex nature of the resulting optimization problem.

The first attempt to solve the multivariate GARCH model was the diagonal VECH model of Bollerslev, Engle, and Wooldridge (1988), who assumed constant covariances for solvability. This extension can be thought of as a trade-off between estimation intractability and practical applicability. Later, statistical tests were developed to check the empirical validity of the assumption of constant covariances; see Bera and Kim (1996) and Tse (2000). Their results for national stock markets show that the covariances are in fact time varying. Therefore, other solutions that can deal with the complexity of the multivariate estimation problem need to be developed.

The factor ARCH model of Engle, Ng, and Rothschild (1990) and the BEKK model of Baba, Engle, Kraft, and Kroner (1989) were attempts to solve the same problem by ensuring positive definiteness of the variance-covariance matrices in the process of optimization, which is an important constraint in multivariate GARCH models. All of these specifications impose very different restrictions on the variance-covariance matrix for computational tractability. For example, Schoenberg (1998), in his GAUSS-based commercial software FANPAC, claims to impose constraints on the eigenvalues of the variance-covariance matrices, although the details are not revealed.

The purpose of the present paper is to solve the multivariate GARCH optimization problem in which we follow a more general approach by taking a constrained nonlinear programming view of GARCH volatility estimation models without imposing artificial restrictions for tractability. This is made possible by recent advances in numerical optimization algorithms and software. ARCH and GARCH models are usually presented to the reader as unconstrained optimization models with recursive terms in econometrics and finance texts (see, e.g., Hamilton (1987) and Gourieroux (1997)), whereas they actually fall into the domain of nonconvex, nonlinearly constrained nonlinear programming. They are usually solved by extensions of Newton or quasi-Newton methods that take into account the recursive nature of terms defining the objective function. Against this background a major goal of this paper is to test the practical solvability (i.e., computing a Karush–Kuhn–Tucker point) of these models as nonlinearly constrained nonconvex programs using the AMPL modeling language (Fourer, Gay, and Kernighan (1993)) and the state-of-the-art optimization packages available through the recently developed NEOS[2] interface at the Argonne National Laboratory.

We believe this research effort is a worthwhile undertaking, as the current financial econometrics literature does not use these valuable sources of optimization software, to the best of our knowledge. Second, we establish through our computational results that the bivariate and trivariate GARCH volatility estimation models for which relatively few software systems exist in the market are solved very effectively by our approach, thus contributing a new tool to the econometric finance literature.

---

[1]See Engle (1987); Bollerslev, Engle, and Wooldridge (1988); Giovannini and Jorion (1989); Engle, Ng, and Rothschild (1990); Bollerslev (1990); Ng, Engle, and Rothschild (1991); Conrad, Gültekin, and Kaul (1991); Kroner and Claesens (1991); Kroner and Sultan (1993); Lien and Luo (1994); Karolyi (1995); Park and Switzer (1995); Tse (2000).

[2]http://www-neos.mcs.anl.gov; see Cyzik, Mesnier, and Moré (1998)

Furthermore, our empirical results for the DAX, FTSE, and S & P 500 indices demonstrate that this approach tracks the variability in realized volatility better than both the diagonal VECH and the BEKK representations. However, we should stress that the major contribution of this paper lies in the proposed general approach and its documented superior solution quality from an optimization point of view. Although a visual inspection of the results and mean-square errors of the trivariate application is promising, a thorough empirical investigation of the forecasting accuracy is a topic for further research.

We organize the rest of this paper as follows. In section 2, we review the univariate GARCH model. Section 3 is devoted to a review and discussion of the multivariate and, in particular, the bivariate and trivariate GARCH models on which we concentrate. In section 4, we illustrate our approach by applying it to daily returns of the S & P 500, FTSE 100, and DAX indices, report our results, and compare them with the diagonal VECH and BEKK representations. Section 5 concludes the paper.

2. Univariate GARCH Model. The analysis of time series dynamics of economic data is usually based on observations of relevant processes, e.g., the behavior of short- and long-term interest rates, rate of inflation, stock prices, etc. In general terms, an observed time series is viewed as a realization of a stochastic process, i.e., a sequence of random variables that are defined on some state space Ω. These random variables may be unidimensional, leading to univariate econometric models, or multidimensional, in which case multivariate models are appropriate. Furthermore, the random variables are indexed by time, where we assume that observations are recorded at regularly spaced intervals, which allows one to consider time indices taking only integer values. The stochastic process is denoted by

$$Y = (Y_t, t \in \mathcal{T}),$$

where the index set $\mathcal{T}$ is the set of nonnegative integers or the set of natural numbers. In the present paper we consider the following autoregressive process for stock index returns, which explains the behavior of the random variable in terms of its past values as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_m Y_{t-m} + \varepsilon_t,$$

where $\varepsilon = (\varepsilon_t)$ is a weak white noise satisfying the martingale difference sequence condition

$$E(\varepsilon_t | \underline{\varepsilon_{t-1}}) = 0,$$

where the notation $E(.)$ denotes mathematical expectation and $\underline{\varepsilon_{t-1}} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ represents the vector of past values. It is important to model the level of financial time series $\{Y_t\}$, but sometimes it might be even more important to model the volatility of the series to quantify the risks involved in a specific trading strategy, especially when the empirical evidence suggests that the level process $\{Y_t\}$ shows no particular time dependence, whereas the volatility process exhibits a certain time dependence. Instead of assuming that the conditional variance of the noise, i.e., $E(\varepsilon_t^2 | \underline{\varepsilon_{t-1}})$, is time independent, we allow for time dependence through an autoregressive equation for the squared error terms (innovations) as follows:

$$(2.1) \qquad E(\varepsilon_t^2 | \underline{\varepsilon_{t-1}}) \equiv h_t = c + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}.$$

The above model is referred to as GARCH(p,q).[3] In the case $p = 0$, we have the ARCH(q) model:

$$(2.2) \qquad E(\varepsilon_t^2|\underline{\varepsilon_{t-1}}) \equiv h_t = c + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2.$$

An important consideration in the study of time series is the stationarity properties of the time series in the interest of forecasting ability. Imposing stationarity is a vital part of modeling. In particular, if $\{Y_t\}$ is stationary, the mean, variance, and autocorrelations can usually be well approximated by sufficiently long time averages. Formally, a stochastic process with a finite mean and variance is called covariance (or second-order) stationary if for all t, $t - s$,

$$(2.3) \qquad E(Y_t) = E(Y_{t-s}) = \mu,$$

$$(2.4) \qquad E[(Y_t - \mu)^2] = E[(Y_{t-s} - \mu)^2] = \sigma_Y^2,$$

$$(2.5) \qquad E[(Y_t - \mu)(Y_{t-s} - \mu)] = E[(Y_{t-j} - \mu)(Y_{t-j-s} - \mu)] = \gamma_s,$$

where μ, σ_Y^2, and γ_s are all constants. Simply put, a time series is covariance stationary if its mean and all auto-covariances are unaffected by a change of time origin. In the above models, $\phi \in \Re^m$, $\alpha \in \Re_{++}^q$, $\beta \in \Re_{++}^p$ (the notation $\Re_{++}^p$ and $\Re_{++}^q$ represent the space of q- and p-dimensional real vectors with strictly positive components, respectively), c is a positive scalar, and

$$(2.6) \qquad \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$$

is sufficient to ensure second-order stationarity asymptotically. For further details the reader is referred to Property 3.19 of Gourieroux (1997).

An important tool in the estimation of the above parameters is the technique of maximum likelihood estimation. Assuming a normal distribution for Y_t given the past observations, application of the maximum likelihood technique in the case of GARCH(p,q) leads to the following optimization problem:

$$(2.7) \qquad \max -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log h_t - \frac{1}{2} \sum_{t=1}^{T} \frac{\varepsilon_t^2}{h_t}$$

subject to the stationarity condition (2.6), the specification of conditional variances h_t given by (2.1), and the nonnegativity condition on c, α, β.

Therefore, for the GARCH(p,q) case we can formulate the following optimization problem:

$$\max \quad -\frac{1}{2} \sum_{t=1}^{T} \log h_t - \frac{1}{2} \sum_{t=1}^{T} \frac{\varepsilon_t^2}{h_t}$$

$$\text{s.t.} \quad c + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} = h_t \quad \forall t = 1, \ldots, T,$$

---

[3]Excellent references are available on this important topic. The interested reader is referred to Droesbeke, Fichet, and Tassi (1994); Gourieroux (1997); and Hamilton (1987) for details.

$$\sum_{i=1}^{m} \phi_i Y_{t-i} + \varepsilon_t \;=\; Y_t \quad \forall t = 1, \ldots, T,$$

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i \;\leq\; 1,$$

$$h_t \;\geq\; 0 \quad \forall\, t = 1, \ldots, T,$$

$$c \;\geq\; 0,$$

$$\alpha_i \;\geq\; 0 \quad \forall\, i = 1, \ldots, q,$$

$$\beta_i \;\geq\; 0 \quad \forall\, i = 1, \ldots, p,$$

where we have replaced the strict inequality in (2.6) with a nonstrict inequality in the interest of computational tractability. This modification did not create any problems, as this constraint turned out to be inactive (satisfied as a strict inequality) at the reported solution in our computational tests (see values of α_1 and β_1 in Table 1, section 4).

Regarding issues of convexity in the above model, we notice that the function $\log h_t + \frac{\varepsilon_t^2}{h_t}$ is a quasi-convex function in (ε_t, h_t). Unfortunately, the sum of quasi-convex functions is not necessarily quasi-convex. Therefore, we do not expect to detect hidden convexity in the objective function of the above model. The constraints are also of a polynomial nature and obviously nonconvex. These observations imply that any attempt at numerical solution of the above model is bound to yield at best a Karush–Kuhn–Tucker point (not necessarily a local maximum).

3. Multivariate Model. When the error term ε_t is a multivariate process of dimension n, we can introduce the same formulation as in the univariate case for all the components of the conditional variance-covariance matrix. Now, for all $t = 1, \ldots, T$ we have $Y_t \in \Re^n$ and $\varepsilon_t \in \Re^n$ with components Y_{lt} and $\varepsilon_{lt}, l = 1, \ldots, n$, respectively. We denote the components of the $n \times n$ conditional variance-covariance matrix $H_t = E(\varepsilon_t \varepsilon_t^T | \varepsilon_{t-1})$ by h_{klt}. The log-likelihood function to be maximized in the multivariate case is given as

$$-\frac{1}{2} \sum_{t=1}^{T} (\log \det H_t + \varepsilon_t^T H_t^{-1} \varepsilon_t).$$

Following Kraft and Engle (1982) and Bollerslev, Engle, and Wooldridge (1988), a multivariate extension of univariate GARCH (2.1) is as follows:

$$(3.1) \qquad \text{vech}(H_t) = \text{vech}(C) + \sum_{i=1}^{q} A_i \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}^T) + \sum_{j=1}^{p} B_j \text{vech}(H_{t-j}),$$

where vech is the operator that consists in stacking up the lower triangular and the diagonal portions of the columns of a symmetric matrix into a vector, the matrices A_i and B_j are of size $\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}$, and C is a symmetric matrix of size $n \times n$. This general formulation is termed the VECH model by Engle and Kroner (1995).

Now, we consider the following estimation problem that we refer to as the constrained nonlinear programming (NLP) formulation:

$$\max \quad -\frac{1}{2}\sum_{t=1}^{T}(\log \det H_t + \varepsilon_t^T H_t^{-1}\varepsilon_t)$$

$$\text{s.t.} \qquad \text{vech}(H_t) \;=\; \text{vech}(C) + \sum_{i=1}^{q} A_i \text{vech}(\varepsilon_{t-i}\varepsilon_{t-i}^T) + \sum_{j=1}^{p} B_j \text{vech}(H_{t-j})$$

$$\forall\, t = 1,\ldots,T,$$

$$\sum_{i=1}^{m}\phi_{li}Y_{l,t-i} + \varepsilon_{lt} \;=\; Y_{lt} \quad \forall t = 1,\ldots,T, l = 1,\ldots,n,$$

$$H_t \;\succeq\; 0 \quad \forall\, t = 1,\ldots,T,$$

where the symbol $\succeq$ means "symmetric, positive semidefinite." The above mathematical program is the most general multivariate GARCH specification model, from which simplified specifications were obtained by imposing certain restrictions on matrices A_i and B_j. Below we briefly review the most important two from the literature in sections 3.1 and 3.2, respectively.

We obtained above a nonlinear programming problem with semidefiniteness constraints. In this case, the stationarity condition is not easy to incorporate into the above problem, as it requires that the roots of the determinant of $I - \sum_{i=1}^{q} A_i z^i - \sum_{j=1}^{p} B_j z^j$ be greater than 1. However, this condition considerably simplifies to an implementable constraint in the bivariate case. It is easy to verify that for $n = 2$, the stationarity condition is equivalent to

$$(3.2) \qquad\qquad\qquad I - A - B \succeq 0,$$

which can be incorporated as nonlinear constraint(s) into the model, where we take $A = A_1$ and $B = B_1$ to be symmetric for tractability.[4] Notice also that the function $\frac{1}{2}\sum_{t=1}^{T}(\log \det H_t + \varepsilon_t^T H_t^{-1}\varepsilon_t)$ is a difference of convex functions since the second component function is a convex function in H_t, ε_t (see Vanderbei and Benson (1999)), and the negative of the first component function is also known to be convex in H_t.

We now compare the above approach with the diagonal VECH and the BEKK representations, the two competing models used in the present paper.

3.1. The Diagonal VECH Model. The diagonal VECH representation was proposed by Bollerslev, Engle, and Wooldridge (1988), who took the matrices A_i and B_j to be diagonal. For a GARCH(1,1) process the entries h_{ijt} of the matrix are specified according to the recursion

$$(3.3) \qquad\qquad h_{ijt} = \omega_{ij} + \beta_{ij}h_{ij,t-1} + \alpha_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1},$$

where ε_t is a multivariate process of dimension n.

In matrix notation, we can cast the associated log-likelihood maximization model as follows:

$$\max \quad -\frac{1}{2}\sum_{t=1}^{T}(\log \det H_t + \varepsilon_t^T H_t^{-1}\varepsilon_t)$$

$$\text{s.t.} \qquad H_t \;=\; C + A \odot \varepsilon_{t-1}\varepsilon_{t-1}^T + B \odot H_{t-1} \quad \forall\, t = 1,\ldots,T,$$

$$\sum_{i=1}^{m}\phi_{li}Y_{l,t-i} + \varepsilon_{lt} \;=\; Y_{lt} \quad \forall t = 1,\ldots,T, l = 1,\ldots,n,$$

$$H_t \;\succeq\; 0 \quad \forall\, t = 1,\ldots,T,$$

---

[4]This is to be able to conveniently decompose $I - A - B$ into LDL^T factors. However, we also have computational results where this constraint was omitted. The results are similar to the results obtained when using the constraint, and they are available to the interested reader upon request.

where the notation $\odot$ is used to represent the componentwise product (Hadamard product) of two matrices of conformable dimensions, and C, A, and B are $n \times n$ symmetric matrices.

3.2. The BEKK Model. As the positive semidefiniteness conditions of the general VECH model were found hard to handle, Engle and Kroner (1995) proposed to model the variance and covariance functions with quadratic forms, which is called the BEKK representation. Now, the conditional variance-covariance matrices are represented in the form

$$(3.4) \qquad H_t = C^T C + B^T H_{t-1} B + A^T \varepsilon_{t-1} \varepsilon'_{t-1} A,$$

where A, B, and C are $n \times n$ (not necessarily symmetric) matrices. Clearly, this model ensures positive semidefiniteness of H_t at the expense of increasing the number of parameters to be estimated in comparison to the diagonal VECH model. From a numerical optimization point of view, the BEKK model also increases the nonlinearity of the constraints by utilizing a higher order polynomial representation in comparison to specification (3.1).

3.3. Bivariate and Trivariate Examples. The bivariate case is of special interest since we can give an explicit NLP formulation in this case using a simple formula for the determinant or a Cholesky-type decomposition. The trivariate case is also amenable to solution using an LDL^T representation that we discuss below. For ease of exposition let us consider the simpler ARCH(1) process. We have three distinct conditional variance-covariance components:

$$h_{11,t} = E(\varepsilon_{1t}^2 | \underline{\varepsilon_{t-1}}),$$

$$h_{12,t} = E(\varepsilon_{1t} \varepsilon_{2t} | \underline{\varepsilon_{t-1}}),$$

$$h_{22,t} = E(\varepsilon_{2t}^2 | \underline{\varepsilon_{t-1}}).$$

The recurrence relation (3.1) becomes

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{12} \\ c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1} \varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix}.$$

Hence, we have the following optimization problem:

$$\max \quad -\frac{1}{2} \sum_{t=1}^{T} \left(\log(h_{11,t} h_{22,t} - h_{12,t}^2) + \frac{\varepsilon_{1t}^2 h_{22,t} + \varepsilon_{2t}^2 h_{11,t} - 2\varepsilon_{1t}\varepsilon_{2t} h_{12,t}}{h_{11,t} h_{22,t} - h_{12,t}^2} \right)$$

$$\text{s.t.} \qquad \begin{aligned} h_{11,t} &= c_{11} + a_{11}\varepsilon_{1,t-1}^2 + a_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{13}\varepsilon_{2,t-1}^2 \\ & \quad \forall\, t = 1, \dots, T, \end{aligned}$$

$$\begin{aligned} h_{12,t} &= c_{12} + a_{21}\varepsilon_{1,t-1}^2 + a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{23}\varepsilon_{2,t-1}^2 \\ & \quad \forall\, t = 1, \dots, T, \end{aligned}$$

$$\begin{aligned} h_{22,t} &= c_{22} + a_{31}\varepsilon_{1,t-1}^2 + a_{32}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{33}\varepsilon_{2,t-1}^2 \\ & \quad \forall\, t = 1, \dots, T, \end{aligned}$$

$$\sum_{i=1}^{m} \phi_{1i} Y_{1,t-i} + \varepsilon_{1t} = Y_{1t} \quad \forall t = 1, \dots, T,$$

$$\sum_{i=1}^{m} \phi_{2i} Y_{2,t-i} + \varepsilon_{2t} = Y_{2t} \quad \forall t = 1, \ldots, T,$$

$$h_{11,t} h_{22,t} - h_{12,t}^2 \geq 0 \quad \forall t = 1, \ldots, T,$$

$$h_{11,t} \geq 0 \quad \forall t = 1, \ldots, T.$$

We refer to the above formulation as the *determinant*-constrained NLP formulation.

Note that the constraints can be rewritten as

$$h_{11,t} = c_{11} + \begin{pmatrix} \varepsilon_{1,t-1} & \varepsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{11} & \frac{a_{12}}{2} \\ \frac{a_{12}}{2} & a_{13} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix},$$

$$h_{12,t} = c_{12} + \begin{pmatrix} \varepsilon_{1,t-1} & \varepsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{21} & \frac{a_{22}}{2} \\ \frac{a_{22}}{2} & a_{23} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix},$$

$$h_{11,t} = c_{11} + \begin{pmatrix} \varepsilon_{1,t-1} & \varepsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{31} & \frac{a_{32}}{2} \\ \frac{a_{32}}{2} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix}.$$

More succinctly, the above constraints can be put as

$$H_t = C + \begin{pmatrix} \varepsilon_{1,t-1} & \varepsilon_{2,t-1} & 0 & 0 \\ 0 & 0 & \varepsilon_{1,t-1} & \varepsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{11} & \frac{a_{12}}{2} & a_{12} & \frac{a_{22}}{2} \\ \frac{a_{12}}{2} & a_{13} & \frac{a_{22}}{2} & a_{23} \\ a_{21} & \frac{a_{22}}{2} & a_{31} & \frac{a_{32}}{2} \\ \frac{a_{22}}{2} & a_{23} & \frac{a_{32}}{2} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} & 0 \\ \varepsilon_{2,t-1} & 0 \\ 0 & \varepsilon_{1,t-1} \\ 0 & \varepsilon_{2,t-1} \end{pmatrix}.$$

It suffices that the matrices C and

$$\underline{A_1} = \begin{pmatrix} a_{11} & \frac{a_{12}}{2} & a_{12} & \frac{a_{22}}{2} \\ \frac{a_{12}}{2} & a_{13} & \frac{a_{22}}{2} & a_{23} \\ a_{21} & \frac{a_{22}}{2} & a_{31} & \frac{a_{32}}{2} \\ \frac{a_{22}}{2} & a_{23} & \frac{a_{32}}{2} & a_{33} \end{pmatrix}$$

be positive semidefinite to guarantee positive semidefiniteness of H_t.

An alternative formulation to the determinant-VECH formulation is obtained by parameterizing the matrices H_t as $H_t = L_t D_t L_t^T$, $t = 1, \ldots, T$, where L_t is a unit-lower triangular matrix, and D_t is a diagonal matrix. Clearly, the requirement that H_t be positive (semi)definite is equivalent to the requirement that the entries of the diagonal matrix D_t be positive (nonnegative). More precisely, within the context of the above example, the LDL^T model would translate into

$$\max \quad -\frac{1}{2} \sum_{t=1}^{T} (\log(d_{1t}) + \log(d_{2t}) + \varepsilon_{1t} w_{1t} + \varepsilon_{2t} w_{2,t})$$

$$\text{s.t.} \quad d_{1t} = c_{11} + a_{11} \varepsilon_{1,t-1}^2 + a_{12} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{13} \varepsilon_{2,t-1}^2$$
$$\forall t = 1, \ldots, T,$$

$$d_{1t} l_{21t} = c_{12} + a_{21} \varepsilon_{1,t-1}^2 + a_{22} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{23} \varepsilon_{2,t-1}^2$$
$$\forall t = 1, \ldots, T,$$

$$d_{1t} l_{21t}^2 + d_{2,t} = c_{22} + a_{31} \varepsilon_{1,t-1}^2 + a_{32} \varepsilon_{1,t-1} \varepsilon_{2,t-1} + a_{33} \varepsilon_{2,t-1}^2$$
$$\forall t = 1, \ldots, T,$$

$$\sum_{i=1}^{m} \phi_{1i} Y_{1,t-i} + \varepsilon_{1t} = Y_{1t} \quad \forall t = 1, \ldots, T,$$

$$\sum_{i=1}^{m} \phi_{2i} Y_{2,t-i} + \varepsilon_{2t} = Y_{2t} \quad \forall t = 1, \ldots, T,$$

$$d_{1t}, \ d_{2t} \geq 0 \quad \forall \, t = 1, \ldots, T,$$

where w_{1t} and w_{2t}, $t = 1, \ldots, T$, are "implied" variables used to simplify the objective function that involves the inverse H_t^{-1} of H_t, $t = 1, \ldots, T$. These variables are incorporated into the model as definition-type AMPL constraints which simply implement the forward substitution, diagonal solve, and backward substitution steps to compute the term $H_t^{-1} \varepsilon_t$ in the objective function: $u_{1t} = -\sum_{i=1}^{m} \phi_{1i} Y_{1,t-i} + Y_{1t}$, $u_{2t} = -\sum_{i=1}^{m} \phi_{2i} Y_{2,t-i} + Y_{2t} - l_{21t} u_{1t}$, $v_{1t} = u_{1t}/d_{1t}$, $v_{2t} = u_{2t}/d_{2t}$, $w_{2t} = v_{2t}$, and $w_{1t} = v_{1t} - l_{21t} w_{2t}$ for all $t = 1, \ldots, T$.

We utilize both the LDL^T model and the determinantal model in our tests, wherever computationally appropriate. All our bivariate formulations also include the stationarity condition (3.2) as a constraint similar to the LDL^T decomposition of H_t's.

For the above example, the diagonal VECH representation takes the following form:

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \odot \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix}.$$

The BEKK model yields the following recursion for the bivariate example:

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

$$+ \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Notice that A and C no longer need to be symmetric.

When we have a trivariate model, we use the following mathematical program, which is a direct extension of the bivariate LDL^T representation to the trivariate case:

$$\max \quad -\frac{1}{2} \sum_{t=1}^{T} \left(\log(d_{1t}) + \log(d_{2t}) + \log(d_{3t}) + \varepsilon_{1t} w_{1t} + \varepsilon_{2t} w_{2,t} + \varepsilon_{3t} w_{3,t} \right)$$

s.t.

$$d_{1t} = c_{11} + a_{11}\varepsilon_{1,t-1}^2 + a_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{13}\varepsilon_{3,t-1}\varepsilon_{1,t-1} + a_{14}\varepsilon_{2,t-1}^2 + a_{15}\varepsilon_{3,t-1}\varepsilon_{2,t-1} + a_{16}\varepsilon_{3,t-1}^2,$$

$$d_{1t}l_{21t} = c_{21} + a_{21}\varepsilon_{1,t-1}^2 + a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{23}\varepsilon_{3,t-1}\varepsilon_{1,t-1} + a_{24}\varepsilon_{2,t-1}^2 + a_{25}\varepsilon_{3,t-1}\varepsilon_{2,t-1} + a_{26}\varepsilon_{3,t-1}^2,$$

$$d_{1t}l_{31t} = c_{31} + a_{31}\varepsilon_{1,t-1}^2 + a_{32}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{33}\varepsilon_{3,t-1}\varepsilon_{1,t-1} + a_{34}\varepsilon_{2,t-1}^2 + a_{35}\varepsilon_{3,t-1}\varepsilon_{2,t-1} + a_{36}\varepsilon_{3,t-1}^2,$$

$$d_{1t}l_{21t}^2 + d_{2,t} = c_{22} + a_{41}\varepsilon_{1,t-1}^2 + a_{42}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{43}\varepsilon_{3,t-1}\varepsilon_{1,t-1} + a_{44}\varepsilon_{2,t-1}^2 + a_{45}\varepsilon_{3,t-1}\varepsilon_{2,t-1} + a_{46}\varepsilon_{3,t-1}^2,$$

$$d_{1t}l_{21t}l_{31t} + d_{2,t}l_{32t} = c_{32} + a_{51}\varepsilon_{1,t-1}^2 + a_{52}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{53}\varepsilon_{3,t-1}\varepsilon_{1,t-1} + a_{54}\varepsilon_{2,t-1}^2 + a_{55}\varepsilon_{3,t-1}\varepsilon_{2,t-1} + a_{56}\varepsilon_{3,t-1}^2,$$

$$d_{1t}l_{31t}^2 + d_{2,t}l_{32t}^2 + d_{3t} = c_{33} + a_{61}\varepsilon_{1,t-1}^2 + a_{62}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{63}\varepsilon_{3,t-1}\varepsilon_{1,t-1} + a_{64}\varepsilon_{2,t-1}^2 + a_{65}\varepsilon_{3,t-1}\varepsilon_{2,t-1} + a_{66}\varepsilon_{3,t-1}^2$$
$$\forall \, t = 1, \ldots, T,$$

$$\sum_{i=1}^{m} \phi_{1i} Y_{1,t-i} + \varepsilon_{1t} = Y_{1t} \quad \forall t = 1, \dots, T,$$

$$\sum_{i=1}^{m} \phi_{2i} Y_{2,t-i} + \varepsilon_{2t} = Y_{2t} \quad \forall t = 1, \dots, T,$$

$$\sum_{i=1}^{m} \phi_{3i} Y_{3,t-i} + \varepsilon_{3t} = Y_{3t} \quad \forall t = 1, \dots, T,$$

$$d_{1t}, \, d_{2t}, \, d_{3t} \geq 0 \quad \forall \, t = 1, \dots, T,$$

where w_{1t}, w_{2t}, w_{3t}, $t = 1, \dots, T$, are "implied" variables used to simplify the objective function that involves the inverse H_t^{-1} of H_t, $t = 1, \dots, T$: $u_{1t} = -\sum_{i=1}^{m} \phi_{1i} Y_{1,t-i} + Y_{1t}$, $u_{2t} = -\sum_{i=1}^{m} \phi_{2i} Y_{2,t-i} + Y_{2t} - l_{21t} u_{1t}$, $u_{3t} = -\sum_{i=1}^{m} \phi_{2i} Y_{2,t-i} + Y_{2t} - l_{31t} u_{1t} - l_{32t} u_{2t}$, $v_{1t} = u_{1t}/d_{1t}$, $v_{2t} = u_{2t}/d_{2t}$, $v_{3t} = u_{3t}/d_{3t}$, $w_{3t} = v_{3t}$, $w_{2t} = v_{2t} - l_{32t} w_{3t}$, and $w_{1t} = v_{1t} - l_{31t} w_{3t} - l_{21t} w_{2t}$ for all $t = 1, \dots, T$.

4. Estimation and Empirical Results. To test our approach first we apply the constrained NLP formulation to the univariate case. In the univariate case our data consist of daily returns of the S & P 500 index with 2000 data points.[5] The data set covers the period from 25.4.1988 to 13.3.1996. Table 1 reports the coefficients, standard errors, and the log-likelihood values for the GARCH(1,1) model with the traditional univariate GARCH formulation and the constrained NLP formulation proposed in the present paper. The standard errors in this context refer to the variance-covariance matrix of the maximum likelihood estimates of the coefficients and are computed approximately using the second derivative matrix of the log-likelihood function. The traditional GARCH estimation is carried out using the S-PLUS GARCH module implementing the BHHH algorithm (see Hamilton (1987) for a discussion of the BHHH algorithm), and the NLP estimation is carried out using the FILTER software (see Fletcher and Leyffer (1998)) for constrained NLP. The results demonstrate that the coefficient values obtained by the two models are very close to each other with comparable standard errors. There is a slight improvement in the log-likelihood function for the constrained NLP approach. The value of this exercise is that it validates our formulation prior to an application to the multivariate setting.

For the multivariate application we start with the bivariate case. Our data consist of daily returns of two stock indices, the S & P 500 and the FTSE 100 with 1500 data points covering the period from 18.5.1990 to 12.3.1996. The time period was chosen using the Ljung–Box test, which is used to diagnose the presence of GARCH effects. We compare the constrained NLP approach with the most popular bivariate specifications available in the S-PLUS GARCH module, namely diagonal VECH and the BEKK specifications. To solve the constrained NLP for the bivariate case we use the SNOPT software (see Gill, Murray, and Saunders (1997)). The nonlinear programs resulting from this exercise have 4506 constraints and 4525 variables.

Table 2 reports the coefficients, standard errors, and log-likelihood values for these three specifications. As in the univariate case, the standard errors represent the variance-covariance matrix of the maximum likelihood estimates of the model coefficients. We would like to note here that the coefficients are not very easy to interpret intuitively for either constrained NLP or BEKK, compared to diagonal VECH.

---

[5] For GARCH diagnosis, autocorrelation functions and Ljung–Box statistics were checked. The data can be supplied upon request. All data were obtained from Salomon Brothers' database.

Table 1 *Results with the univariate model on S & P 500 data.*

Method	c	α_1	β_1	Log-likelihood value
Constrained NLP	0.00201931	0.978463	0.0180615	−2179.67
(St. Err.)	(0.0015)	(0.00784)	(0.00103)	
SPLUS	0.00285	0.97250	0.02204	−2181.8
(St. Err.)	(0.000762)	(0.003177)	(0.0034232)	

Table 2 *Results with the bivariate model on S & P 500 and FTSE 100 data (numbers in parentheses are standard errors).*

Coefficients	Constrained NLP	D-VECH	BEKK
c_{11}	−0.198775	0.021812	0.126516
	(0.00597)	(0.07542)	(0.026245)
c_{12}	1.24346	0.016743	0.005078
	(0.00471)	(0.010096)	(0.018835)
c_{22}	−0.121942	0.005688	0.059896
	(0.00211)	(0.001437)	(0.009138)
a_{11}	0.20436	0.04509	0.196017
	(0.00036)	(0.009925)	(0.024318)
a_{12}	−0.384304	0.026886	−0.013858
	(1.27×10^{-9})	(0.011565)	(0.024476)
a_{21}			−0.003001
			(0.016084)
a_{13}	0.17964		
	(0.000106)		
a_{13}	0.17964		
	(0.000106)		
a_{22}	0.959926	0.033912	0.171552
	(0.000824)	(0.005841)	(0.017128)
a_{23}	−0.382031		
	(0.000346)		
a_{33}	0.248888		
	(0.0001308)		
b_{11}	0.396459	0.930056	0.971880
	(0.01033)	(0.016520)	(0.007864)
b_{12}	2.11141	0.885738	0.001883
	(0.01133)	(0.062685)	(0.005981)
b_{21}			0.003817
			(0.004755)
b_{13}	−0.446092		
	(0.002658)		
b_{22}	−8.53698	0.954386	0.980089
	(0.11985)	(0.007181)	(0.004033)
b_{23}	1.62468		
	(0.007876)		
b_{33}	0.509248		
	(0.004097)		
Log-likelihood	−2572.48	−3453.05	−3461.91
AIC	5176.96	6924.1	6945.82
SIC	5261.01	6971.91	7004.26

However, log-likelihood values show that constrained NLP brings a substantial improvement over the diagonal VECH and BEKK representations in the solution of the multivariate GARCH formulation. One might be led to think that comparing likelihood values for different models with differing number of parameters may not be fair. Therefore, to make a fair comparison, we use the AIC and SIC statistics,

which are standard tests of comparison between GARCH models in the literature. AIC and SIC statistics are used for model selection purposes and enable us to compare models with different numbers of coefficients and different numbers of observations. They are calculated as AIC $= -2$ log-likelihood $+ 2$ number of coefficients, and SIC $= -2$ log-likelihood $+ 2\ln$ (number of observations) number of coefficients.[6] Based on the AIC and SIC tests, we can say that constrained NLP parameter estimates are superior to both diagonal VECH and BEKK specifications, although all three provide a solution to the same multivariate GARCH estimation problem. Furthermore, the diagonal VECH model seems to do slightly better than the BEKK representation. As explained in previous sections the log-likelihood function to be maximized is identical in all three approaches compared in the present paper. We believe this result is due to the following three factors: 1. The constrained NLP approach uses a more general representation compared to its competitors. 2. It incorporates the stationarity condition as a side constraint. 3. It employs state-of-the-art optimization software.

Although an empirical investigation of the forecasting accuracy of the so-called GARCH representations is beyond the scope of this paper, we include visual comparisons of our approach to the competing estimations of the same model. To this end, we plot conventional comparison measures of realized volatility[7] versus estimated volatilities from three different specifications. We use annualized volatility, as practitioners quote volatilities in annualized terms using 252 trading days. In Figures 1 and 2 we plot the annualized realized volatility, which is defined as $\sqrt{\text{daily-returns}^2 \times 252}$, and the conditional annualized volatility obtained from GARCH specifications, defined as

$$\sqrt{\text{conditional variances obtained from the estimations} \times 252}$$

for the last 500 data points. The solid lines in the figures are the model's conditional annualized volatilities, whereas the dotted lines represent the benchmark annualized realized volatility. In Figure 3 we plot realized covariances defined as daily return S & P 500 $\times$ daily return FTSE 100 and the conditional covariances obtained from the three different specifications. The better estimation should approximate the dotted lines more closely. A visual inspection of the figures shows that constrained NLP has higher variability, which seems to follow the variability of the realized volatility.

We observe from the figures that the diagonal VECH and BEKK results exhibit rather similar behavior in that the series tend to follow a certain mean value with very small variations. A possible explanation for this behavior can be given as follows. It is highly likely that the numerical optimization algorithm used in S-PLUS diagonal VECH and BEKK implementations (BHHH algorithm) lands on very close Karush–Kuhn–Tucker points. On the other hand, the constrained NLP results display series which seem to follow more closely the trends in realized volatility and covariances, although it has a tendency to overestimate at times. It is conceivable that the sequential quadratic programming algorithm used in SNOPT lands at a completely different Karush–Kuhn–Tucker point compared to the diagonal VECH and BEKK representations.

---

[6]Geweke and Meese (1981) report that asymptotically, SIC correctly identifies an ARMA model, whereas AIC tends to overfit the model. For completeness, we report both. The smaller the statistic, the better the model fit.

[7]There is still ongoing academic debate about the definition and proper calculation of realized volatility, as the observed daily return is just one realization out of many, and volatility is not an observed quantity in the market, but rather estimated. Therefore, the true volatility is unknown. However, we employ the most conventional realized volatility definition in our comparisons.
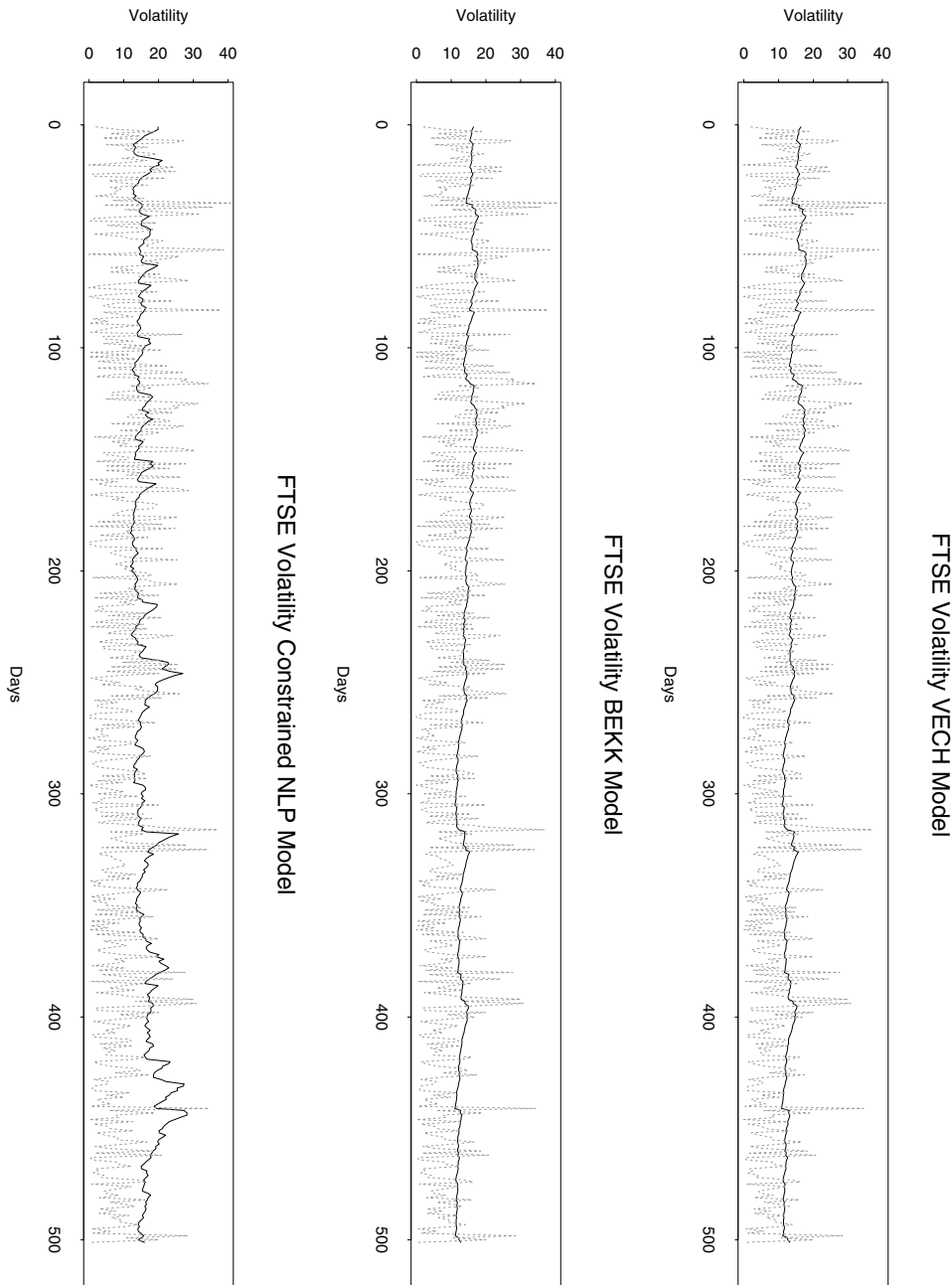
Fig. I *Volatility for the FTSE.*

In Figures 4–6 we report the results of our trivariate tests, where we used 500 data points from the S & P 500, FTSE, and DAX[8] indices for the period 5.7.1988 to

_____

[8]The results we obtained for the DAX were very similar to the results for the S & P 500. Therefore, to keep the paper a reasonable length, we do not report the DAX results here. They can be found at http://www.ie.bilkent.edu.tr/~mustafap/pubs/dax.gz.
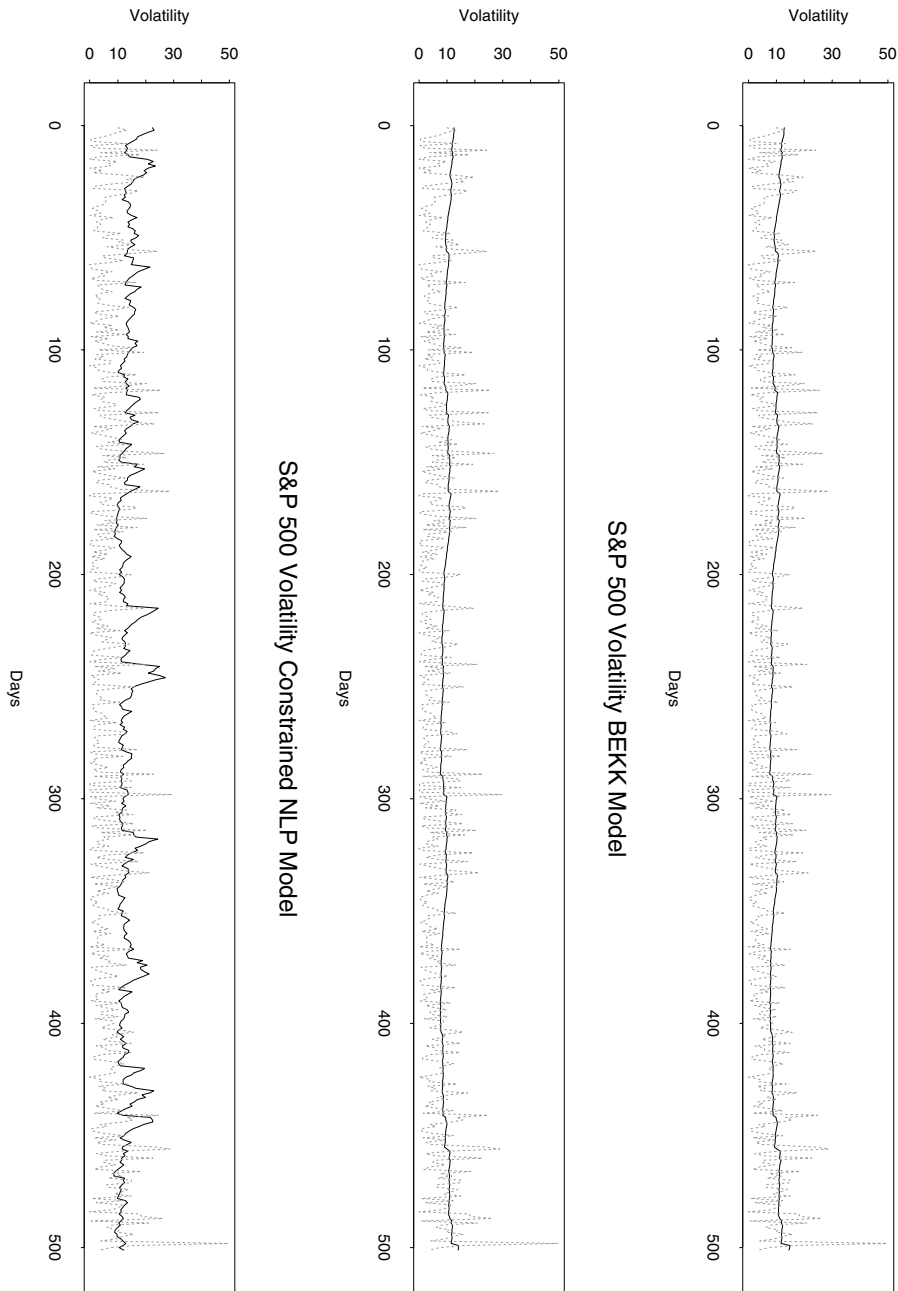
Fig. 2 *Volatility for the S & P 500.*

14.6.1990. As in the bivariate case the data were chosen using the Ljung–Box test. The constrained NLP approach (4575 variables and 4491 constraints) with the SNOPT solver in NEOS yields a log-likelihood objective function value of -437.02, while the competing diagonal VECH and the BEKK representations in the package S-PLUS
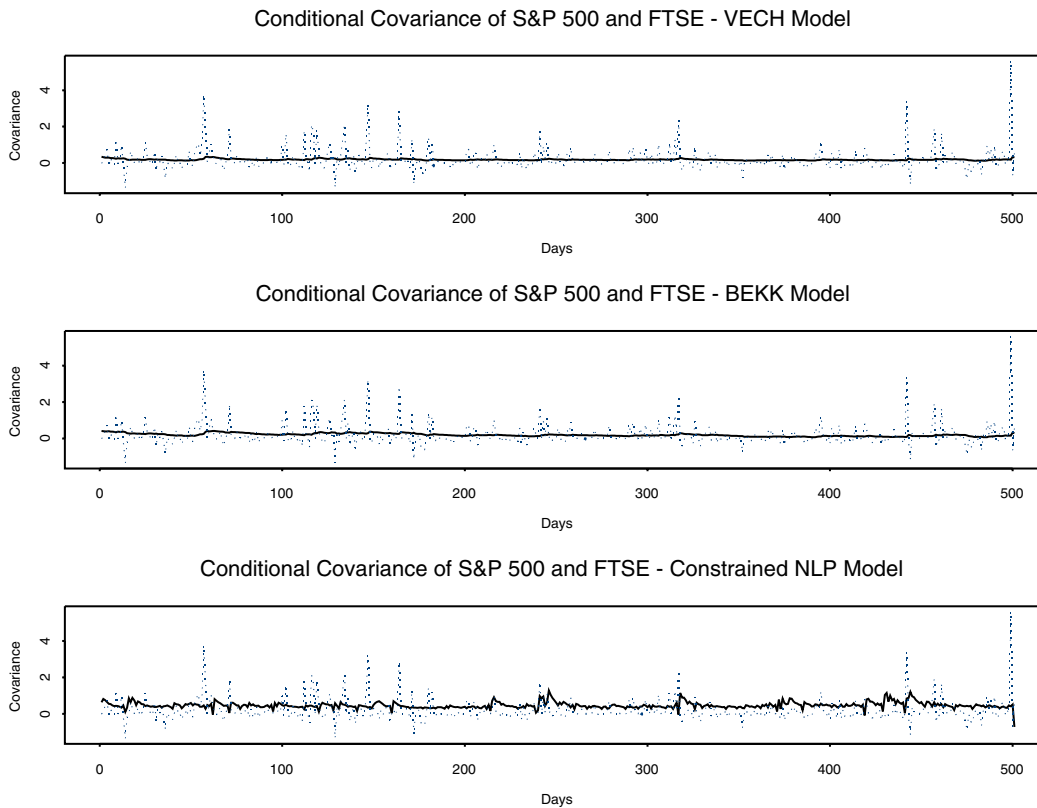
Conditional Covariance of S&P 500 and FTSE - VECH Model

Conditional Covariance of S&P 500 and FTSE - BEKK Model

Conditional Covariance of S&P 500 and FTSE - Constrained NLP Model

Fig. 3 *Conditional covariances of the S & P 500 and the FTSE.*

give the values -1910.51 and -1949.95. The AIC statistic is equal to 958.04 for constrained NLP, whereas it is equal to 3862.78 and 3953.9 for diagonal VECH and BEKK, respectively. The SIC statistic has a value equal to 1396.1 for constrained NLP, whereas it is equal to 4343.0 and 4421.9 for diagonal VECH and BEKK, respectively. According to log-likelihood, AIC, and SIC criteria, the constrained NLP approach performs far better than the other specifications. In Figures 4 and 5 the solid lines represent as usual the model's conditional annualized volatilities, whereas the dotted lines represent the annualized realized volatility. In Figure 6, the solid lines represent the model's conditional covariances, while the dotted lines are used to plot realized conditional covariances. We observe that the constrained NLP results follow the dotted lines more closely, especially in the case of the S & P 500 index, compared to competitive specifications, the diagonal VECH and BEKK.

In addition to visual inspection we also report in Table 3 the mean-square errors associated with the above figures. The mean-square error is calculated as the average of the squared differences of daily forecast volatility and realized volatility values for each figure. Table 3 shows that in the bivariate case constrained NLP mean-square errors are higher for volatilities and covariances. This could be due to the upward bias also observed in the figures. For the trivariate case, constrained NLP has smaller mean-square errors except for FTSE volatility, where all models are comparable. By
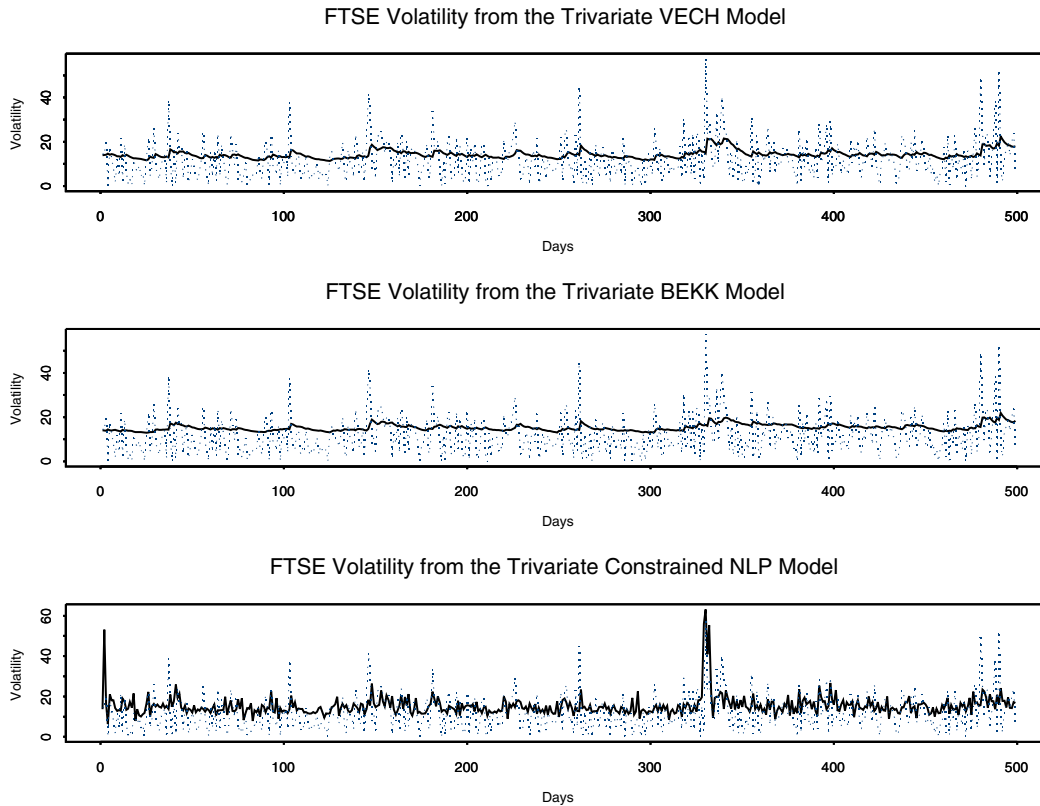
FTSE Volatility from the Trivariate VECH Model

FTSE Volatility from the Trivariate BEKK Model

FTSE Volatility from the Trivariate Constrained NLP Model

Fig. 4 *Volatility for the FTSE in the trivariate case.*

itself, the mean-square error is only a summary measure for forecasting accuracy. Moreover, there exists a variety of more sophisticated methods, both parametric and nonparametric, to judge the forecasting ability of a particular model. On the other hand, the forecasting ability of the multivariate GARCH models is an important empirical research question, and it should be addressed in a more elaborate fashion with different data sets for various time periods. This is obviously a topic for further research.

As a final remark, the computing times using SNOPT on NEOS platforms tend to be rather long for bivariate and trivariate examples, reaching as much as $2\frac{1}{2}$ hours for the bivariate tests and 23 hours for the trivariate tests (SNOPT uses four times as many iterations in the trivariate case as in the bivariate example), while the S-PLUS results were usually obtained within a few minutes, although the diagonal VECH model failed to converge to a solution and computation was stopped after 100 iterations. However, the purpose of our tests was not so much computational efficiency as the quality of the solution, which seems to be clearly superior in the constrained NLP approach to the competing specifications. Furthermore, the constrained NLP results were obtained on NEOS over the World Wide Web, whereas the S-PLUS solvers were available on our personal desktop. Therefore, the times are not directly comparable, and thus should give only a rough indication to the reader.
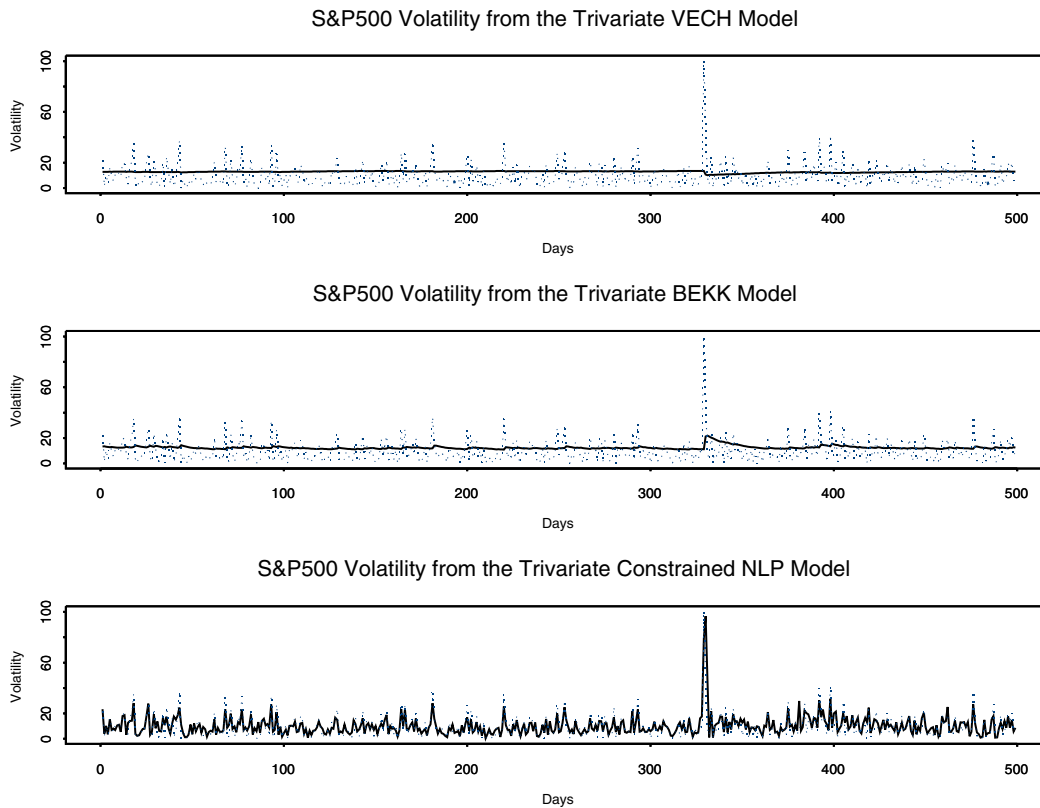
S&P500 Volatility from the Trivariate VECH Model

S&P500 Volatility from the Trivariate BEKK Model

S&P500 Volatility from the Trivariate Constrained NLP Model

Fig. 5 *Volatility for the S & P 500 in the trivariate case.*

In summary, the present paper shows that the early simplifications in the estimation of the multivariate GARCH model (e.g., diagonal VECH and BEKK specifications) proposed in the literature in the interest of solvability are unnecessary from an optimization point of view given the current state-of-the-art in optimization technology. Therefore, the additional investment in the constrained NLP approach seems to be paying off in terms of solution quality. However, there is certainly a need for further research to ascertain the comparative advantage of the constrained NLP approach, especially from a forecasting accuracy point of view.

5. Conclusions. This paper proposed a constrained nonlinear programming view of multivariate generalized autoregressive conditional heteroskedasticity (GARCH) volatility estimation models in financial econometrics. These models are usually presented to the reader as unconstrained optimization models consisting of the maximization of a nonconvex, nonlinear likelihood function defined through recursive terms in the literature, whereas they actually fall into the domain of nonconvex constrained NLP. Our results showed that constrained NLP is a worthwhile exercise for GARCH estimation problems as demonstrated by examples of bivariate and trivariate data, and that it is a significant competitor to the diagonal VECH and the BEKK representations popular in the literature.
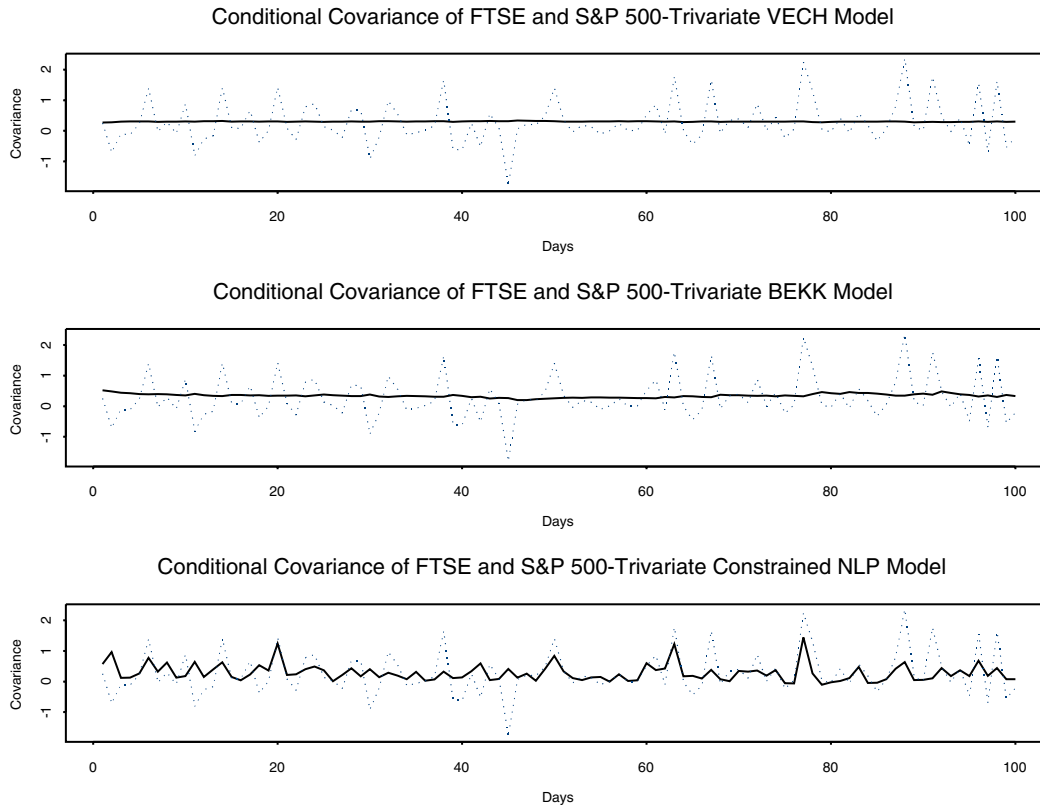
Conditional Covariance of FTSE and S&P 500-Trivariate VECH Model

Conditional Covariance of FTSE and S&P 500-Trivariate BEKK Model

Conditional Covariance of FTSE and S&P 500-Trivariate Constrained NLP Model

Fig. 6 *Conditional covariances for the S & P* 500 *and the FTSE in the trivariate case.*

Table 3 *Mean-square errors in bivariate (last* 500 *observations) and trivariate tests (last* 100 *observations).*

Index	Diag. VECH	BEKK	CNLP
Panel A: Bivariate			
FTSE	0.3064	0.3094	0.6706
S & P 500	0.1816	0.1826	0.4528
FTSE and S & P 500	0.3451	0.3495	0.4650
Panel B: Trivariate			
FTSE	0.3541	0.3708	0.3804
S & P 500	0.2551	0.2554	0.0783
FTSE and S & P 500	0.4769	0.4993	0.3741

REFERENCES

Y. Baba, R.F. Engle, D. Kraft, and K.F. Kroner (1989), *Multivariate Simultaneous Generalized ARCH*, manuscript, Department of Economics, University of California at San Diego.

A.K. Bera and S. Kim (1996), *Testing Constancy of Correlation with an Application to International Equity Returns*, Center for International Business Education and Research, working paper 96-107, University of Illinois, Urbana-Champaign.

T. Bollerslev (1986), *Generalized autoregressive conditional heteroskedasticity*, J. Econometrics, 31, pp. 307–327.

T. Bollerslev (1990), *Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach*, Rev. Econ. and Stat., 72, pp. 498–505.

T. Bollerslev, R.F. Engle, and J Wooldridge (1988), *A capital asset pricing model with time varying covariances*, J. Polit. Econ., 96, pp. 116–131.

J. Conrad, M. Gültekin, and G. Kaul (1991), *Asymmetric predictability of conditional variances*, Rev. Financ. Stud., 4, pp. 597–622.

J. Cyzik, M. Mesnier, and J. Moré (1998), *The NEOS server*, IEEE Comput. Sci. Engrg., 5, pp. 68–75.

J.-J. Droesbeke, B. Fichet, and Ph. Tassi, eds. (1994), *Modélisation ARCH: Théorie statistique et applications dans le domaine de la finance*, Editions de L'Université de Bruxelles.

R.F. Engle (1982), *Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation*, Econometrica, 50, pp. 987–1008.

R.F. Engle (1987), *Multivariate GARCH with Factor Structures–Cointegration in Variance*, manuscript, Department of Economics, University of California at San Diego.

R.F. Engle, V. Ng, and M. Rothschild (1990), *Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills*, J. Economet., 45, pp. 213–238.

R.F. Engle and K.F. Kroner (1995), *Multivariate simultaneous GARCH*, Economet. Theor., 11, pp. 122–150.

E.F. Fama (1965), *The behavior of stock market prices*, J. Bus., 38, pp. 34–105.

R. Fletcher and S. Leyffer (1998), *User manual for FILTER/SQP*, University of Dundee Numerical Analysis Report NA-181, Dundee, Scotland.

R. Fourer, D. Gay, and B. Kernighan (1993), *AMPL: A modeling language for mathematical programming*, Duxbury Press, Belmont, CA.

J.F. Geweke and R. Meese (1981), *Estimating regression models of finite but unknown order*, Int. Econ. Rev., 22, pp. 55–70.

P. Gill, W. Murray, and M. Saunders (1997), *User's guide for SNOPT*, ftp://sdna3.ucsd.edu/pub/peg/reports/sndoc.ps.Z.

A. Giovannini and P. Jorion (1989), *The time variation of risk and return in the foreign exchange and stock market*, J. Financ., 44, pp. 307–325.

C. Gourieroux (1997), *ARCH Models and Financial Applications*, Springer Ser. Statist., Springer-Verlag, New York.

J.D. Hamilton (1987), *Time Series Analysis*, Princeton University Press, Princeton, NJ.

G.A. Karolyi (1995), *A multivariate GARCH model of international transmissions of stock returns and volatility: The case of United States and Canada*, J. Bus. Econ. Stat., 13, pp. 11–25.

D.F. Kraft and R.F. Engle (1982), *Autoregressive Conditional Heteroskedasticity in Multiple Time Series*, manuscript, Department of Economics, University of California at San Diego.

K.F. Kroner and S. Claesens (1991), *Optimal dynamic hedging portfolios and the currency composition of external debt*, J. Int. Money Financ., 10, pp. 131–148.

K.F. Kroner and J. Sultan (1993), *Time varying distribution and dynamic hedging with foreign currency futures*, J. Financ. Quant. Anal., 28, pp. 535–551.

D. Lien and X. Luo (1994), *Multi-period hedging in the presence of conditional heteroskedasticity*, J. Futures Markets, 14, pp. 927–955.

B. Mandelbrot (1963), *The variation of certain speculative prices*, J. Bus., 36, pp. 394–419.

V. Ng, R.F. Engle, and M. Rothschild (1991), *A multi dynamic factor model for stock returns*, J. Economet., 52, pp. 245–266.

T.H. Park and L.N. Switzer (1995), *Bivariate GARCH estimation of the optimal hedge ratios for stock index futures: A note*, J. Futures Markets, 15, pp. 61–67.

R. Schoenberg (1998), *FANPAC User Manual*, Aptech Systems, Maple Valley, WA.

Y.K. Tse (2000), *A test for constant correlations in a multivariate GARCH model*, J. Economet., 98, pp. 107–127.

R. Vanderbei and H.Y. Benson (1999), *On Formulating Semidefinite Programming Problems as Smooth Convex Nonlinear Optimization Problems*, Technical Report ORFE 99-01, Princeton University, Princeton, NJ.