Constrained Nonlinear Programming for Volatility Estimation with GARCH Models

Aslıhan Altay-Salih
Faculty of Management
Bilkent University
06533 Ankara, Turkey

Mustafa Ç. Pınar*
Department of Industrial Engineering
Bilkent University
06533 Ankara, Turkey

Sven Leyffer
Department of Mathematics
University of Dundee, Dundee, Scotland.

July 24, 2001

Abstract

This paper proposes 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 multivariate case as they offer a substantial improvement over the Diagonal VECH and the BEKK models that are popular in the literature.

Keywords. Financial Econometrics, Constrained Nonlinear programming, GARCH, Volatility estimation, Maximum likelihood estimation.

*Corresponding author, e-mail: mustafap@bilkent.edu.tr. Part of this research was conducted while this author was visiting University of Dundee, Department of Mathematics in July 2000 supported by a grant from the Royal Society. The hospitality and encouragement of Professor Roger Fletcher are gratefully acknowledged.

1 Introduction

Volatility plays an important role in several areas of current finance literature. It is central to portfolio selection models in that efficient portfolios are formed by computing the maximum return for a given level of volatility. General equilibrium models like 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 correlation estimation is an important research area for both academia 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), and Fama (1965)), is the driving force for GARCH family of models. The success of these models in the univariate case for volatility estimation has inspired an interest in correlation estimation which is a harder problem, and led to the development and application of the multivariate extensions. [1] The major difficulty in the multivariate case stems form the highly nonlinear and nonconvex nature of the resulting optimization problem.

The first extension in the multivariate direction was the diagonal VECH model of Bollerslev, Engle and Wooldridge (1988) where they assumed constant correlations and extended the univariate case to the vectorized conditional variance matrix. This first attempt can be thought of as a trade-off between estimation intractability and practical applicability since the condition of positive definiteness is difficult to impose. Later, statistical tests have been developed to check the validity of the assumption of constant correlations; see Bera and Kim (1996), Tse (2000). Their results for national stock markets show that the correlations are in fact time varying. Therefore, other methods that can deal with the complexity of the multivariate estimation problem need to be developed.

The Factor ARCH model of Engle, Ng, and Rotschild (1990) and The BEKK model of Baba, Engle, Kraft and Kroner (1989), were attempts to solve this problem by ensuring positive definiteness of the variance-covariance matrices in the process of optimization. All of these models impose very different restrictions on the covariance matrix for computational tractability.

The purpose of the present paper is to solve the optimization problem by proposing a more formal approach by taking a constrained nonlinear programming view of GARCH volatility estimation models both in the univariate and multivariate cases without imposing artificial restrictions for tractability. This is made possible by advances in the numerical optimization algorithms and software literature. ARCH and GARCH models are usually presented to the reader as unconstrained optimization models in econometrics, and finance texts (see e.g., Hamilton (1987), Gourieroux (1992)) with recursive terms 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 interface at the Argonne National Laboratory. We believe this research effort is a worthwhile undertaking as the current financial econometrics literature does not currently use these valuable sources of optimization software, to the best of our knowledge. Second, we establish through our com-

[1] See Engle (1987), Bollerslev, Engle and Wooldridge (1988), Giovannini and Jorion (1989), Engle, Ng, and Rothschild (1990), Bollerslev (1990), Ng, Engle, 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).

putational results that the bivariate GARCH volatility estimation models for which relative few software systems exist in the market, are solved very effectively by our approach, thus contributing a methodology to the econometric finance literature. Furthermore, our results for FTSE and S & P 500 indices demonstrate that our approach tracks realized volatility better than both the diagonal VECH and the BEKK representations.

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 of the bivariate GARCH model on which we concentrate. In section 4, we illustrate our approach by applying it to daily returns of S & P 500 and FTSE 100 indices, report our results, and compare them with the diagonal VECH and BEKK representations. The paper is concluded in section 5.

2 Univariate Model

In this section we briefly review the univariate GARCH volatility estimation models. Excellent references are available on this important topic. The interested reader is referred to [8, 19, 20] for details.

We consider the following autoregressive process

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

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

$$E(\varepsilon_t/\varepsilon_{t-1}) = 0.$$

Instead of assuming that the conditional variance of the noise, i.e., $E(\varepsilon_t^2/\varepsilon_{t-1})$ is time independent, we allow for time dependence through an autoregressive equation for the squared error terms (innovations) as follows

$$E(\varepsilon_t^2/\varepsilon_{t-1}) \equiv h_t = c + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}, \ \alpha_q \neq 0, \beta_p \neq 0. \tag{2.1}$$

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

$$E(\varepsilon_t^2/\varepsilon_{t-1}) \equiv h_t = c + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2, \ a_q \neq 0. \tag{2.2}$$

In the above models, $\phi \in \Re^m$, $\alpha \in \Re^q_{++}$, $\beta \in \Re^p_{++}$, and c is a positive scalar, to ensure asymptotic second order stationarity; see Property 3.19 of [19].

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:

$$\max_{\theta} \log L_T(\theta) \tag{2.3}$$

where

$$\log L_T(\theta) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log h_t(\theta) - \frac{1}{2} \sum_{t=1}^{T} \frac{\varepsilon_t^2}{h_t(\theta)} \tag{2.4}$$

where $\theta = (\phi, \alpha, \beta, c)$, subject to the stationarity condition

$$\sum_{i=1}^{\max\{p,q\}} \alpha_i + \beta_i < 1, \tag{2.5}$$

the specification of conditional variances given by (2.1), and the non-negativity condition on c, α, β.

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

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\sum_{t=1}^{T}\log h_t - \frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2}{h_t} \\
\text{s.t.} \quad & c + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j h_{t-j} = h_t, \forall t = 1,\ldots,T \\
& \sum_{i=1}^{m}\phi_i Y_{t-i} + \varepsilon_t = Y_t, \forall t = 1,\ldots,T \\
& \sum_{i=1}^{\max\{p,q\}} \alpha_i + \beta_i \le 1 \\
& c \ge 0 \\
& \alpha_i \ge 0, \forall\, i = 1,\ldots,q \\
& \beta_i \ge 0, \forall\, i = 1,\ldots,p
\end{aligned}
$$

Regarding issuses of convexity in the above model we can offer the following remarks. 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 non-convex. These observations imply that any attempts 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 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. Let us denote the error terms by $\varepsilon_{lt}, l = 1,\ldots,n$, and the components of $H_t = V(\varepsilon_t/\underline{\varepsilon_{t-1}})$ by h_{klt}. Therefore, we have the following representation:

$$h_{klt} = C_{kl} + \sum_{i=1}^{q}[\sum_{k'l'} a_{klk'l'i}\varepsilon_{k',t-i}\varepsilon_{l',t-i}] + \sum_{j=1}^{p}(\sum_{k'l'} b_{klk'l'}h_{k',l',t-i}) \tag{3.1}$$

Here, one has to make sure that the matrices H_t are symmetric. Therefore, one has to add the following conditions:

$$C_{kl} = C_{lk},$$

$$a_{klk'l'i} = a_{lkk'l'i},\ a_{klk'l'i} = a_{kll'k'i},$$

$$b_{klk'l'i} = b_{lkk'l'i},\ b_{klk'l'i} = b_{kll'k'i}.$$

The log-likelihood function to be maximized in the multivariate case is given as

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

where Θ represents the vector of parameters to be estimated.

Following Kraft and Engle (1982) and Bollerslev, Engle and Wooldridge (1988) an equivalent and natural multivariate extension of univariate GARCH (3.1), which is easier to view than the above representation, is as follows

$$\text{vech}(H_t) = \text{vech}(C) + \sum_{i=1}^{q} A_i \text{vech}(\varepsilon_{t-i}\varepsilon_{t-i}^T) + \sum_{j=1}^{p} B_j \text{vech}(H_{t-j}). \tag{3.2}$$

where vech is the operator which 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 (1993).

Now, we consider the following estimation problem that we refer to as the Constrained NLP formulation:

$$\begin{aligned}
\max \quad & -\frac{1}{2}\sum_{t=1}^{T}(\log\det H_t + \varepsilon_t^T H_t^{-1}\varepsilon_t) \\
\text{s.t.} \quad & \text{vech}(H_t) = \text{vech}(C) + \sum_{i=1}^{q} A_i \text{vech}(\varepsilon_{t-i}\varepsilon_{t-i}^T) + \sum_{j=1}^{p} B_j \text{vech}(H_{t-j}), \forall\, t = 1,\ldots,T \\
& \sum_{i=1}^{m} \phi_{li} Y_{l,t-i} + \varepsilon_{lt} = Y_{lt}, \forall t = 1,\ldots,T, l = 1,\ldots,n \\
& H_t \succeq 0, \forall\, t = 1,\ldots,T
\end{aligned}$$

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 semi-definiteness 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 one. However, this condition considerably simplifies into an implementable constraint in the bivariate case. It is easy to verify that for $n = 2$, the stationarity condition is equivalent to

$$I - A - B \succeq 0$$

which can be incorporated as nonlinear constraint(s) into the model, where we take A and B to be symmetric for tractability. 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

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

where ε_t is a multivariate process of dimension n.

$$\begin{aligned} \max \quad & L(\Theta) \\ \text{s.t.} \quad & H_t = \Omega + A \odot \varepsilon_{t-1}\varepsilon_{t-1}^T + B \odot H_{t-1}, \forall\, t = 1,\dots,T \\ & \sum_{i=1}^{m} \phi_{li} Y_{l,t-i} + \varepsilon_{lt} = Y_{lt}, \forall t = 1,\dots,T, l = 1,\dots,n \\ & H_t \succeq 0, \forall\, t = 1,\dots,T \end{aligned}$$

where the notation $\odot$ is used to represent the componentwise product (Hadamard product) of two matrices of conformable dimensions.

3.2 The BEKK Model

As the positive semi-definiteness conditions of the general VECH model were found hard to handle, Engle and Kroner (1993) proposed to model the variance and covariance function with quadratic forms, which is called the BEKK representation. Now, the conditional variance/covariance matrices are represented in the form

$$H_t = C^T C + B^T H_{t-1} B + A^T \varepsilon_{t-1} \varepsilon_{t-1}' A \tag{3.4}$$

where A, B and C are $n \times n$ matrices. Clearly, this model ensures positive semi-definiteness 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 (3.2).

3.3 The Bivariate Case

The bivariate case is of special interest since we can give an explicit nonlinear programming formulation in this case using a simple formula for the determinant or a Cholesky-type decomposition. For ease of exposition let us consider an ARCH(1) process. We have three distinct conditional variance-covariance components

$$\begin{aligned} h_{11,t} &= E(\varepsilon_{1t}^2 / \underline{\varepsilon_{t-1}}) \\ h_{12,t} &= E(\varepsilon_{1t}\varepsilon_{2t} / \underline{\varepsilon_{t-1}}) \\ h_{22,t} &= E(\varepsilon_{2t}^2 / \underline{\varepsilon_{t-1}}) \end{aligned}$$

The recurrence relation (3.2) becomes

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

Hence, we have the following optimization problem:

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\sum_{t=1}^{T}\left(\log(h_{11,t}h_{22,t}-h_{12,t}^2)+\frac{\varepsilon_{1t}h_{22,t}+\varepsilon_{2t}h_{11,t}+\varepsilon_{1t}\varepsilon_{2t}h_{12,t}}{h_{11,t}h_{22,t}-h_{12,t}^2}\right) \\
\text{s.t.} \quad & h_{11,t} \;=\; c_{11}+a_{11}\varepsilon_{1,t-1}^2+a_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1}+a_{13}\varepsilon_{2,t-1}^2 \,\forall\, t=1,\ldots,T \\
& h_{12,t} \;=\; c_{12}+a_{21}\varepsilon_{1,t-1}^2+a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1}+a_{23}\varepsilon_{2,t-1}^2 \,\forall\, t=1,\ldots,T \\
& h_{22,t} \;=\; c_{22}+a_{31}\varepsilon_{1,t-1}^2+a_{32}\varepsilon_{1,t-1}\varepsilon_{2,t-1}+a_{33}\varepsilon_{2,t-1}^2 \,\forall\, t=1,\ldots,T \\
& \sum_{i=1}^{m}\phi_{1i}Y_{1,t-i}+\varepsilon_{1t} \;=\; Y_{1t}, \forall t=1,\ldots,T \\
& \sum_{i=1}^{m}\phi_{2i}Y_{2,t-i}+\varepsilon_{2t} \;=\; Y_{2t}, \forall t=1,\ldots,T \\
& h_{11,t}h_{22,t}-h_{12,t}^2 \;\geq\; 0, \forall\, t=1,\ldots,T
\end{aligned}
$$

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 semi-definiteness of H_t.

In the present paper, we chose to work with the bivariate Constrained NLP formulation above as it was the easiest to code in AMPL. 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 (non-negative). We utilize both the LDL^T model and the determinantal model in our tests, wherever computationally appropriate.

4 Estimation and Empirical Results

To validate our approach first we applied the constrained NLP formulation to the univariate case. In the univariate case our data consists of daily returns of S & P 500 index with 2000 data points.[2] The data covers 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 model proposed in the present paper. The traditional GARCH estimation is carried out using S-PLUS GARCH module, and the NLP model is solved using the FILTER software [14] for constrained nonlinear programming. 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 model. The value of this exercise is that it validates our approach prior to an application to the multivariate setting.

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 1: Results with the Univariate Model on SP500 Data

For the multivariate application we choose to concentrate on the bivariate case. Our data consist of daily returns of two stock indices: S & P 500 and FTSE 100 with 1500 data points covering from 18.5.1990 to 12.3.1996. We compare the constrained NLP model with the most popular bivariate models available in S-PLUS GARCH module, namely Diagonal VECH and the BEKK specifications. To solve the constrained NLP models for the bivariate case we used the SNOPT software [16]. The nonlinear programs resulting from this exercise have approximately 4500 constraints and 4500 variables. Table 2 reports the coefficients, standard errors, and the log-likelihood values for these three models. We would like to note here that the coefficients are not very easy to interpret intuitively for both the constrained NLP and the BEKK models, compared to the Diagonal VECH model. However, log-likelihood values show that constrained NLP brings a substantial improvement over the Diagonal VECH and BEKK representations. 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. Our constrained NLP approach uses a more general representation compared to its competitors, 2. incorporates the stationarity condition as a side constraint, and 3. employs state-of-the-art optimization software.

Further evidence to the improvement due to the use of the constrained NLP approach can be observed in Figures 1, 2 below where we plot the annualized realized volatility[3] and the conditional annualized volatility obtained from GARCH specifications[4] 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 annualized realized volatility. In Figure 3 we plot realized covariances[5] and the conditional covariances obtained from the three different models. We observe from the figures that the Diagonal VECH and BEKK results exhibit a 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 be given as follows. It is highly likely that the numerical optimization algorithm used in

[2]For GARCH diagnosis, autocorrelation functions and Ljung-Box statistics have been checked. The data can be supplied upon request.

[3]volatility is defined as $\sqrt{dailyreturns^2 \times 252}$.

[4]defined as $\sqrt{\text{conditional variances obtained from the estimations} \times 252}$

[5]realized covariance = daily return S & P 500 $\times$ daily return FTSE 100.

S-PLUS Diagonal VECH and BEKK implementations 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 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 models.

5 Conclusions

The paper proposed 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 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 nonlinear programming. Our results demonstrated that constrained nonlinear programming is a worthwhile option for GARCH estimation problems, especially for the multivariate case as it is a significant competitor to the Diagonal VECH and the BEKK models popular in the literature.

A trivariate application of the

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

Table 2: Results with the Bivariate Model on SP500 and FTSE 100 Data (Numbers in parentheses are standard errors).

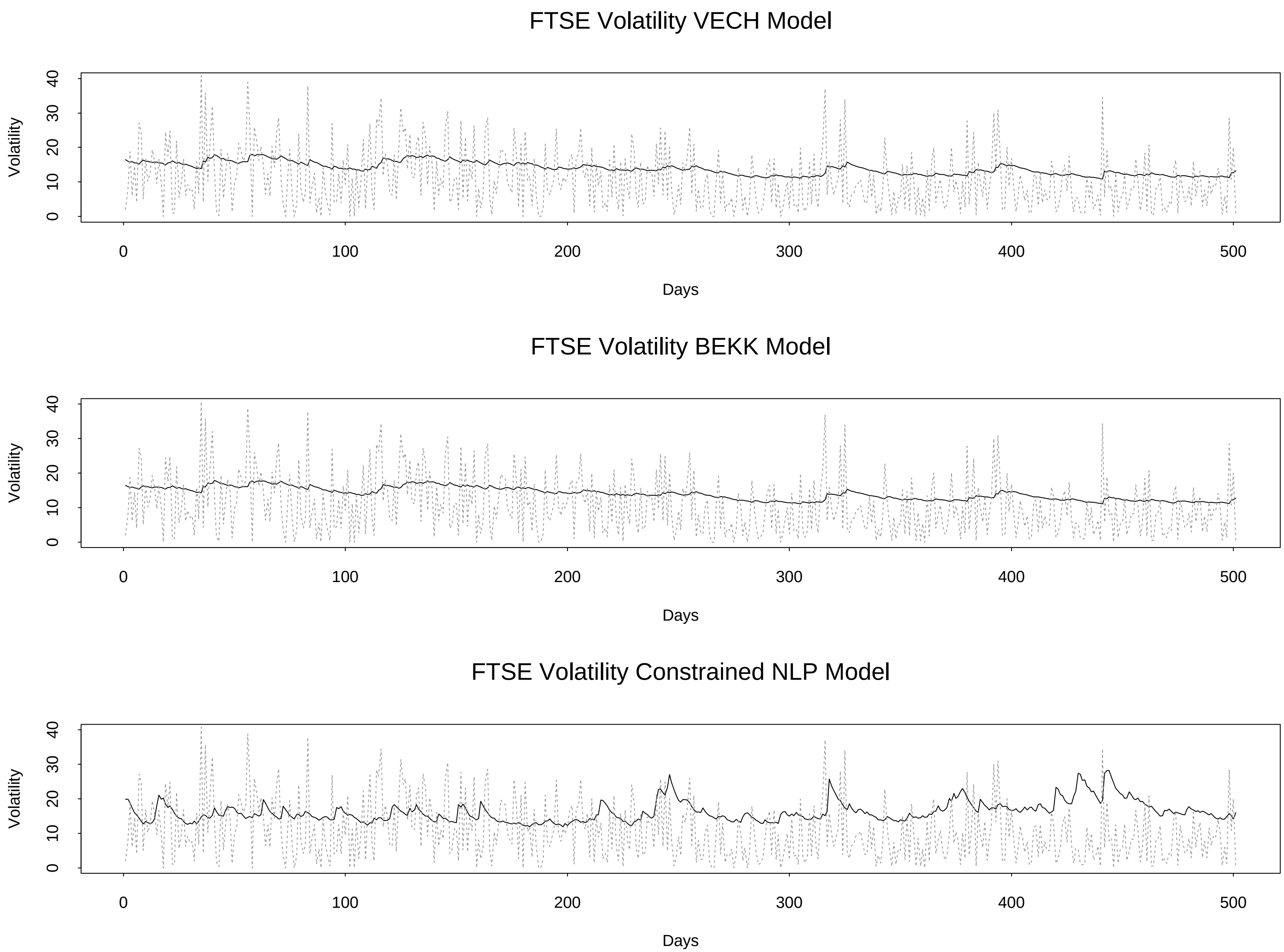

Figure 1: Volatility for FTSE

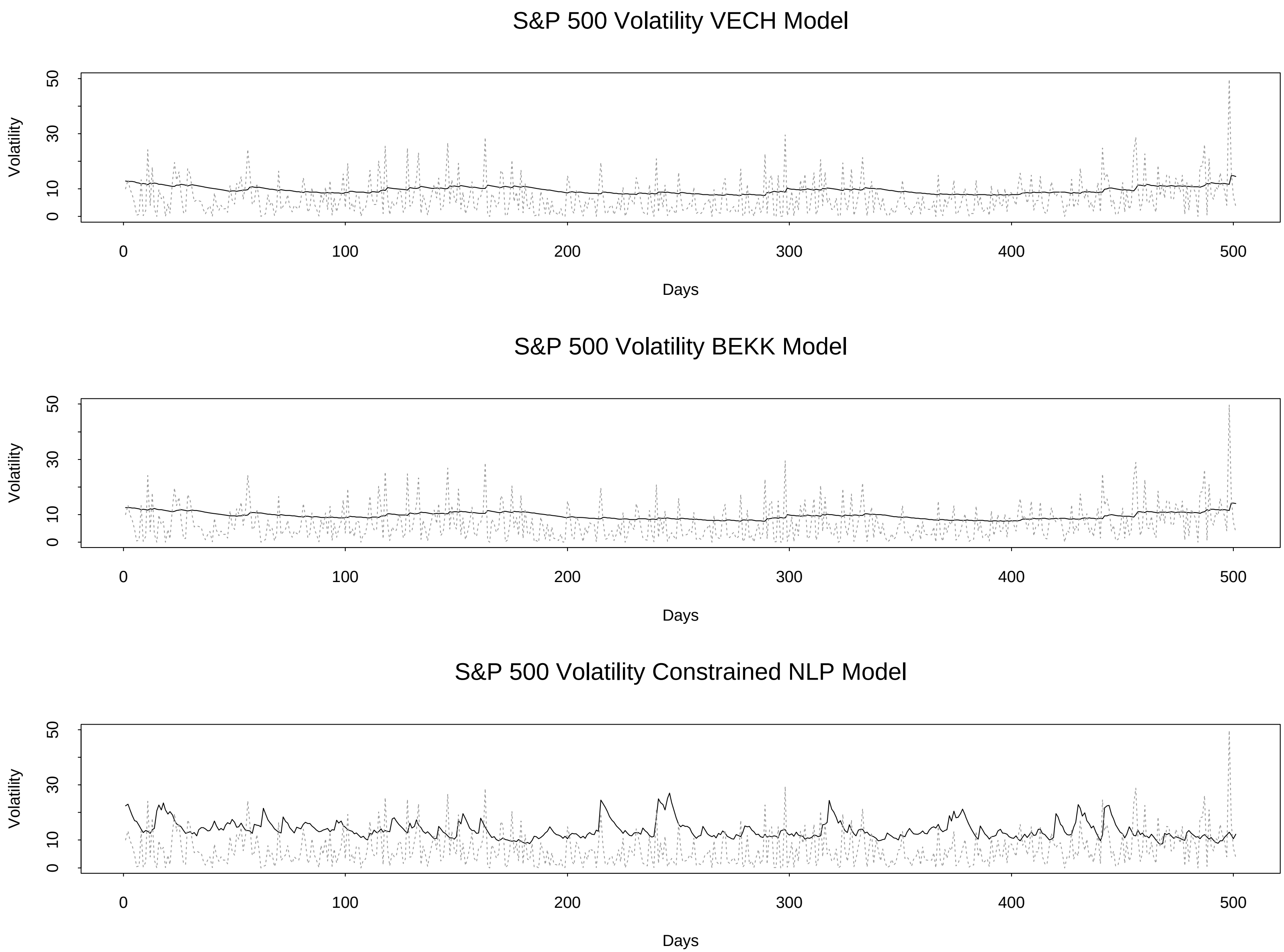

Figure 2: Volatility for S & P 500

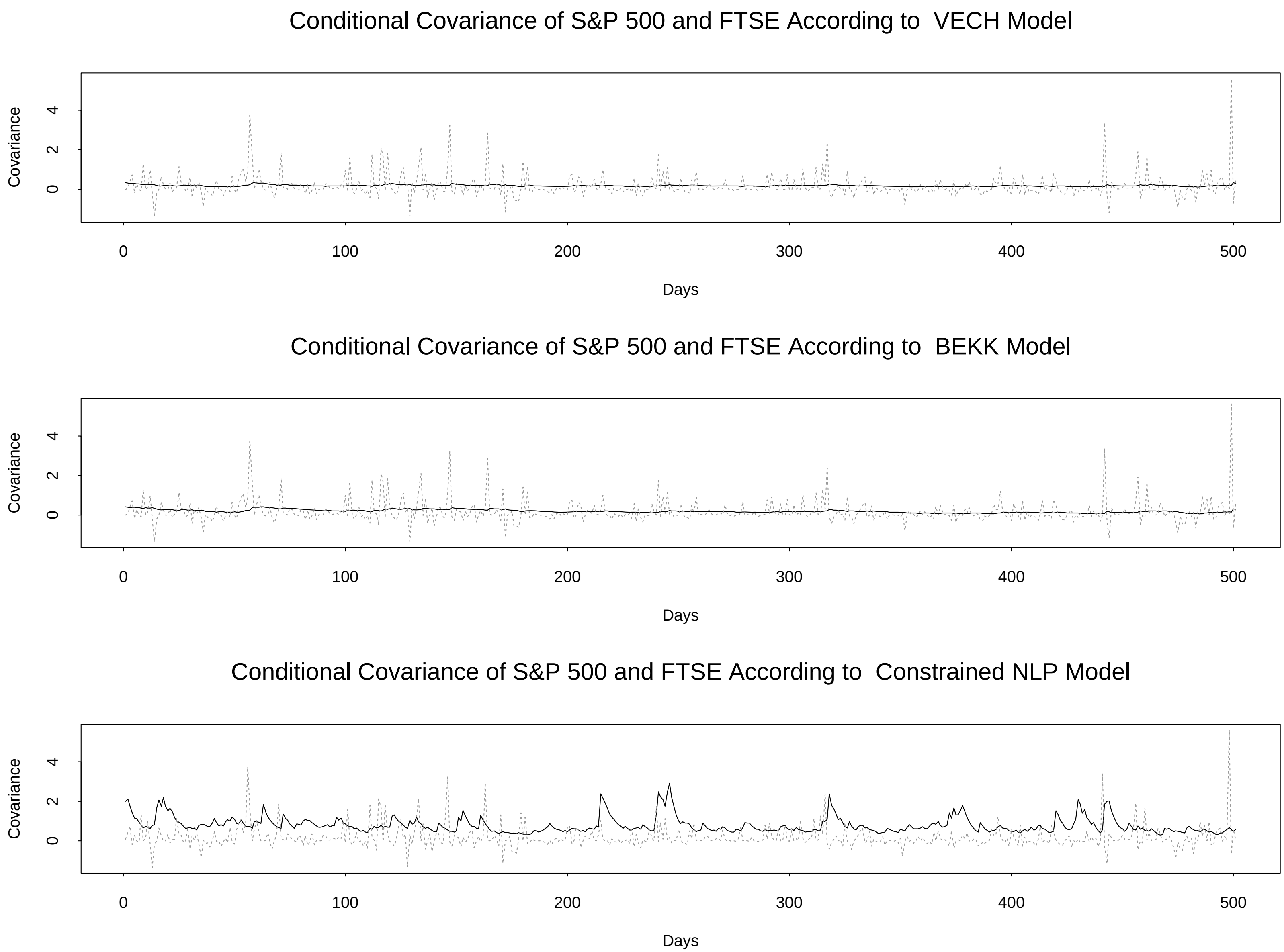

Figure 3: Conditional Covariances of S & P 500 and FTSE

References

[1] Y. Baba, R.F. Engle, D. Kraft, and K.F. Kroner. 1989. Multivariate Simultaneous Generalized ARCH. UCSD, Department of Economics, Manuscript.

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

[3] T. Bollerslev. 1986. Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307–327.

[4] T. Bollerslev, R.F. Engle, and J Wooldridge. 1988. A Capital Asset Pricing Model with Time Varying Covariances, *Journal of Political Economy*, 96, 116–131.

[5] T. Bollerslev. 1990. Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach. *Review of Economics and Statistics*, 72, 498–505.

[6] T. Bollerslev, R.F. Engle and D.B. Nelson. 1994. ARCH Models, in *Handbook of Econometrics, R.F. Engle and D.L. McFadden, Eds.* Elsevier Science, 2961–3038.

[7] J. Conrad, M. Gültekin and G. Kaul. 1991. Asymmetric Predictability of Conditional Variances, *Review of Financial Studies*, 4, 597–622.

[8] J.-J. Droesbeke, B. Fichet and Ph. Tassi, eds. *Modélisation ARCH: Théorie Statistique et Applications dans le Domaine de la Finance*, Editions de L'Université de Bruxelles, 1994.

[9] R.F. Engle. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica*, 50, 987–1008.

[10] R.F. Engle. 1987. Multivariate GARCH with Factor Structures–Cointegration in Variance. unpublished manuscript, Department of Economics, UCSD.

[11] R.F. Engle, V. Ng. and M. Rotschild. 1990. Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills. *Journal of Econometrics*, 45, 213–238.

[12] E.F. Fama. 1965. The Behavior of Stock Market Prices. *Journal of Business*, 38, 34–105.

[13] R. Fletcher and S. Leyffer. 1998. User Manual for FILTER/SQP. University of Dundee Numerical Analysis Report NA-181.

[14] R. Fourer, D. Gay and B. Kernighan. 1993. *AMPL A Modeling Language for Mathematical Programming*, Duxbury Press.

[15] P. Gill, W. Murray, and M. Saunders. User's Guide for SNOPT, `ftp://sdna3.ucsd.edu/pub/peg/reports/sndoc.ps.Z`.

[16] A. Giovannini and P. Jorion. 1989. The Time Variation of Risk and Return in the Foreign Exchange and Stock Market, *Journal of Finance*, 44, 307–325.

[17] C. Gourieroux and A. Monfort. 1996. *Statistique et Modèles Econométriques, Vol. 1: Notions Générales, Estimation, Prévisions, Algorithmes*, Economica, Paris.

[18] C. Gourieroux. 1997. *ARCH Models and Financial Applications.* Springer Series in Statistics. Springer-Verlag, New York.

[19] J.D. Hamilton. 1987. *Time Series Analysis*, Princeton University Press, Princeton, N.J.

[20] G.A. Karolyi. 1995. A Multivariate GARCH Model of International Transmissions of Stock Returns and Volatility: The Case of United States and Canada. *Journal of Business and Economic Statistics*, 13, 11–25.

[21] D.F. Kraft and R.F. Engle. 1982. Autoregressive Conditional Heteroskedasticity in Multiple Time Series. unpublished manuscript, Department of Economics, UCSD.

[22] K.F. Kroner and S. Claesens. 1991. Optimal Dynamic Hedging Portfolios and the Currency Composition of External Debt. *Journal of International Money and Finance*, 10, 131–148.

[23] K.F. Kroner and J. Sultan. 1993. Time Varying Distribution and Dynamic Hedging with Foreign Currency Futures. *Journal of Financial and Quantitative Analysis*, 28, 535–551.

[24] D. Lien and X. Luo. 1994. Multi-period Hedging in the Presence of Conditional Heteroskedasticity. *Journal of Futures Markets*, 14, 927–955.

[25] B. Mandelbrot. 1963. The Variation of Certain Speculative Prices. *Journal of Business*, 36, 394–419.

[26] V. Ng, R.F. Engle, and M. Rotschild. 1991. A Multi Dynamic Factor Model for Stock Returns. *Journal of Econometrics*, 52, 245–266.

[27] T.H. Park and L.N. Switzer. 1995. Bivariate GARCH Estimation of the Optimal Hedge Ratios for Stock Index Futures: A Note. *Journal of Futures Markets*, 15, 61–67.

[28] Y.K. Tse. 2000. A Test for Constant Correlations in a Multivariate GARCH Model. *Journal of Econometrics*, 98, 107–127.

[29] 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.