

A multivariate heavy-tailed distribution for ARCH/GARCH residuals

Dimitris N. Politis*
Department of Mathematics
University of California—San Diego
La Jolla, CA 92093-0112, USA
email: dpolit@ucsd.edu

Abstract

A new multivariate heavy-tailed distribution is proposed as an extension of the univariate distribution of Politis (2004). The properties of the new distribution are discussed, as well as its effectiveness in modeling ARCH/GARCH residuals. A practical procedure for multi-parameter numerical maximum likelihood is also given, and a real data example is worked out.

JEL codes: C3; C5.

Keywords: Heteroscedasticity, maximum likelihood, time series.

*Research partially supported by NSF grant SES-04-18136 funded jointly by the Economics and Statistics Programs of the National Science Foundation.

1 INTRODUCTION

Consider univariate data X_1, \dots, X_n arising as an observed stretch from a financial returns time series $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ such as the percentage returns of a stock price, stock index or foreign exchange rate. The returns series $\{X_t\}$ are assumed strictly stationary with mean zero which—from a practical point of view—implies that trends and other nonstationarities have been successfully removed. Furthermore, for simplicity the returns series $\{X_t\}$ will be assumed to have a symmetric¹ distribution around zero.

The celebrated ARCH models of Engle (1982) were designed to capture the phenomenon of volatility clustering in the returns series. An ARCH(p) model can be described by the following equation:

$$X_t = Z_t \sqrt{a + \sum_{i=1}^p a_i X_{t-i}^2} \quad (1)$$

where a, a_1, \dots, a_p are nonnegative parameters. Since the residuals from a fitted ARCH(p) model do not appear to be normally distributed, practitioners have typically been employing ARCH models with heavy-tailed errors. A popular assumption for the distribution of the Z_t s is the t -distribution with degrees of freedom empirically chosen to match the apparent degree of heavy tails as measured by higher-order moments such as the kurtosis; see e.g. Bollerslev et al. (1992) or Shephard (1996) and the references therein.

Nevertheless, the choice of a t -distribution is quite arbitrary, and the same is true for other popular heavy-tailed distributions, e.g. the double exponential. For this reason, Politis (2004) developed an implicit ARCH model that helps suggest a more natural heavy-tailed distribution for ARCH and/or GARCH residuals; see Bollerslev (1986) for the definition of GARCH models. The motivation for this new distribution is given in Section 2, together with a review of some of its properties. In Section 3, the issue of multivariate data is brought up, and a multivariate version of the new heavy-tailed distribution is proposed. A practical algorithm for multi-parameter numerical maximum likelihood is given in Section 4 based on an approximate steepest-ascent idea. Section 5 deals in detail with the bivariate paradigm. Finally, a real data example is worked out in Section 6.

¹Some authors have raised the question of existence of skewness in financial returns; see e.g. Patton (2002) and the references therein. Nevertheless, at least as a first approximation, the assumption of symmetry is very useful for model building.

2 THE IMPLICIT ARCH MODEL IN THE UNIVARIATE CASE: A BRIEF REVIEW

Engle's (1982) original assumption was that the errors Z_t in model (1) are i.i.d. $N(0, 1)$. Under that assumption, the residuals

$$\hat{Z}_t = \frac{X_t}{\sqrt{\hat{a} + \sum_{i=1}^p \hat{a}_i X_{t-i}^2}} \quad (2)$$

ought to appear normally distributed; here, $\hat{a}, \hat{a}_1, \hat{a}_2, \dots$ are estimates of the nonnegative parameters a, a_1, a_2, \dots . Nevertheless, as previously mentioned, the residuals typically exhibit a degree of heavy tails. In other words, typically there is not a specification for $\hat{a}, \hat{a}_1, \hat{a}_2, \dots$ that will render the residuals \hat{Z}_t normal-looking.

A way to achieve ARCH residuals that appear normal was given in the implicit ARCH model of Politis (2004) that is defined as:

$$X_t = W_t \sqrt{a + a_0 X_t^2 + \sum_{i=1}^p a_i X_{t-i}^2}. \quad (3)$$

When the implicit ARCH model (3) is fitted to the data, the following residuals ensue:

$$\hat{W}_t = \frac{X_t}{\sqrt{\hat{a} + \hat{a}_0 X_t^2 + \sum_{i=1}^p \hat{a}_i X_{t-i}^2}}. \quad (4)$$

As argued in Politis (2004), by contrast to the standard ARCH model (1), there typically always exists here a specification for $\hat{a}, \hat{a}_0, \hat{a}_1, \dots$ that will render the residuals \hat{W}_t normal-looking.² Hence, it is not unreasonable to assume that the W_t errors in model (3) are i.i.d. $N(0, 1)$.

Technically speaking, however, this is not entirely correct since it is easy to see that the W_t s are actually bounded (in absolute value) by $1/\sqrt{a_0}$. A natural way to model the situation where the W_t are thought to be close to

²To see why, note that the specification $\hat{a} = 0, \hat{a}_0 = 1$, and $\hat{a}_i = 0$ for $i > 0$, corresponds to $\hat{W}_t = \text{sign}(X_t)$ that has kurtosis equal to one. As previously mentioned, specifications with $\hat{a}_0 = 0$ typically yield residuals \hat{W}_t with heavier tails than the normal, i.e., kurtosis bigger than 3. Thus, by the intermediate value theorem, a specification for $\hat{a}, \hat{a}_0, \hat{a}_1, \dots$ should exist that would render the kurtosis of \hat{W}_t attaining the intermediate value of 3.

$N(0, 1)$ but happen to be bounded is to use a truncated standard normal distribution, i.e., to assume that the W_t are i.i.d. with probability density given by

$$\frac{\phi(x)\mathbf{1}\{|x| \leq C_0\}}{\int_{-C_0}^{C_0} \phi(y)dy} \quad \text{for all } x \in \mathbf{R} \quad (5)$$

where ϕ denotes the standard normal density, $\mathbf{1}\{S\}$ the indicator of set S , and $C_0 = 1/\sqrt{a_0}$. If/when a_0 is small enough, the boundedness of W_t is effectively not noticeable.

Note that the only difference between the implicit ARCH model (3) and the standard ARCH model (1) is the introduction of the term X_t paired with a nonzero coefficient a_0 in the right-hand-side of equation (3). It is precisely this X_t term appearing in both sides of (3) that gives the characterization ‘implicit’ to model (3).

Nevertheless, we can solve equation (3) for X_t , to obtain:

$$X_t = U_t \sqrt{a + \sum_{i=1}^p a_i X_{t-i}^2} \quad (6)$$

where

$$U_t = \frac{W_t}{\sqrt{1 - a_0 W_t^2}}. \quad (7)$$

Thus, the implicit ARCH model (3) is seen to be tantamount to the regular ARCH(p) model (6) associated with the new innovation term U_t .

However, if W_t is assumed to follow the truncated standard normal distribution (5), then the change of variable (7) implies that the innovation term U_t appearing in the ARCH model (6) has the density $f(u; a_0, 1)$ defined as:

$$f(u; a_0, 1) = \frac{(1 + a_0 u^2)^{-3/2} \exp(-\frac{u^2}{2(1+a_0 u^2)})}{\sqrt{2\pi} (\Phi(1/\sqrt{a_0}) - \Phi(-1/\sqrt{a_0}))} \quad \text{for all } u \in \mathbf{R} \quad (8)$$

where Φ denotes the standard normal distribution function. Equation (8) gives the univariate heavy-tailed density for ARCH residuals proposed in Politis (2004).

Note that the density $f(u; a_0, 1)$ can be scaled to create a two-parameter family of densities with typical member given by

$$f(x; a_0, c) = \frac{1}{c} f\left(\frac{x}{c}; a_0, 1\right) \quad \text{for all } x \in \mathbf{R}. \quad (9)$$

The nonnegative parameter a_0 is a shape parameter having to do with the degree of heavy tails, while the positive parameter c represents scale; note that $f(u; a_0, 1) \rightarrow \phi(u)$ as $a_0 \rightarrow 0$.

It is apparent that $f(u; a_0, c)$ has heavy tails. Except for the extreme case when $a_0 = 0$ and all moments are finite, in general moments are finite only up to (almost) order two. In other words, if a random variable U follows the density $f(u; a_0, c)$ with $a_0, c > 0$, then it is easy to see that

$$E|U|^d < \infty \quad \text{for all } d \in [0, 2) \quad \text{but} \quad E|U|^d = \infty \quad \text{for all } d \in [2, \infty).$$

Thus, working with the heavy-tailed density $f(u; a_0, c)$ is associated with the conviction that financial returns have higher-order moments that are infinite; Politis (2003, 2004) gives some empirical evidence pointing to that fact.³

3 A MULTIVARIATE HEAVY-TAILED DISTRIBUTION FOR ARCH/GARCH

Here, and throughout the remainder of the paper, we will consider the case where the data $\mathbf{X}_1, \dots, \mathbf{X}_n$ are an observed stretch from a *multivariate*, strictly stationary and zero mean, series of returns $\{\mathbf{X}_t, t = 0, \pm 1, \pm 2, \dots\}$, where $\mathbf{X}_t = (X_t^{(1)}, X_t^{(2)}, \dots, X_t^{(m)})'$.

By the intuitive arguments of Section 2, each of the coordinates of \mathbf{X}_t may be thought to satisfy an implicit ARCH model such as (3) with residuals following the truncated standard normal distribution (5). Nevertheless, it is possible in general that the coordinates of the vector \mathbf{X}_t are dependent, and this can be due to their respective residuals being correlated.

Focusing on the j th coordinate, we can write the implicit ARCH model:

$$X_t^{(j)} = W_t^{(j)} \sqrt{a^{(j)} + \sum_{i=0}^p a_i^{(j)} (X_{t-i}^{(j)})^2} \quad \text{for } j = 1, \dots, m \quad (10)$$

³Although strictly speaking the 2nd moment is also infinite, since the moment of order $2-\epsilon$ is finite for any $\epsilon > 0$, one can proceed in practical work as if the 2nd moment were indeed finite; for example, no finite-sample test can ever reject the hypothesis that data following the density $f(u; a_0, 1)$ with $a_0 > 0$ have a finite 2nd moment.

and the corresponding regular ARCH model:

$$X_t^{(j)} = U_t^{(j)} \sqrt{a^{(j)} + \sum_{i=1}^p a_i^{(j)} (X_{t-i}^{(j)})^2} \quad \text{for } j = 1, \dots, m \quad (11)$$

with residuals given by:

$$U_t^{(j)} = \frac{W_t^{(j)}}{\sqrt{1 - a_0^{(j)} (W_t^{(j)})^2}} \quad \text{for } j = 1, \dots, m. \quad (12)$$

As usual, we may define the j th volatility by $s_t^{(j)} = \sqrt{a^{(j)} + \sum_{i=1}^p a_i^{(j)} (X_{t-i}^{(j)})^2}$.

With sufficiently large p , an ARCH model such as (11) can approximate an arbitrary stationary univariate GARCH model; see Bollerslev (1986) or Gouriéroux (1997). For example, consider the multivariate GARCH (1,1) model given by the equation:

$$X_t^{(j)} = s_t^{(j)} U_t^{(j)} \quad (13)$$

with $s_t^{(j)} = \sqrt{C^{(j)} + A^{(j)} (X_{t-1}^{(j)})^2 + B^{(j)} (s_{t-1}^{(j)})^2}$ for $j = 1, \dots, m$.

It is easy to see that the multivariate GARCH model (13) is tantamount to the multivariate ARCH model (11) with $p = \infty$ and the following identifications:

$$a^{(j)} = \frac{C^{(j)}}{1 - B^{(j)}}, \quad \text{and } a_k^{(j)} = A^{(j)} (B^{(j)})^{k-1} \quad \text{for } k = 1, 2, \dots \quad (14)$$

The above multivariate ARCH/GARCH models (11) and (13) are in the spirit of the constant (with respect to time) conditional correlation multivariate models of Bollerslev (1990). For instance, note that the volatility equation for coordinate j involves past values and volatilities from the same j th coordinate only. More general ARCH/GARCH models could also be entertained where $s_t^{(j)}$ is allowed to depend on past values and volatilities of other coordinates as well; see e.g. Vrontos et al. (2003) and the references therein. For reasons of concreteness and parsimony, however, we focus here on the simple models (11) and (13).

To complete the specification of either the multivariate ARCH model (11) or the multivariate GARCH model (13), we need to specify a distribution

for the vector residuals $(U_t^{(1)}, U_t^{(2)}, \dots, U_t^{(m)})'$. For fixed t , those residuals are expected to be heavy-tailed and generally dependent; this assumption will lead to a constant (with respect to time) conditional *dependence* model since correlation alone can not capture the dependence in non-normal situations.

As before, each $W_t^{(j)}$ can be assumed (quasi)normal but bounded in absolute value by $C_0^{(j)} = 1/\sqrt{a_0^{(j)}}$. So, the vector $(W_t^{(1)}, W_t^{(2)}, \dots, W_t^{(m)})'$ may be thought to follow a truncated version of the m -variate normal distribution $N_m(\mathbf{0}, \mathbf{P})$, where \mathbf{P} is a positive definite *correlation* matrix, i.e., its diagonal elements are unities. As well known, the distribution $N_m(\mathbf{0}, \mathbf{P})$ has density

$$\phi_{\mathbf{P}}(\mathbf{y}) = (2\pi)^{-m/2} |\mathbf{P}|^{-1/2} \exp(-\mathbf{y}'\mathbf{P}^{-1}\mathbf{y}/2) \quad \text{for all } \mathbf{y} \in \mathbf{R}^m. \quad (15)$$

Thus, the vector $(W_t^{(1)}, W_t^{(2)}, \dots, W_t^{(m)})'$ would follow the truncated density

$$\frac{\phi_{\mathbf{P}}(\mathbf{w}) \mathbf{1}\{\mathbf{w} \in E_0\}}{\int_{E_0} \phi_{\mathbf{P}}(\mathbf{y}) d\mathbf{y}} \quad \text{for all } \mathbf{w} \in \mathbf{R}^m \quad (16)$$

where E_0 is the rectangle $[-C_0^{(1)}, C_0^{(1)}] \times [-C_0^{(2)}, C_0^{(2)}] \times \dots \times [-C_0^{(m)}, C_0^{(m)}]$.

Consequently, by the change of variables (12) it is implied that the vector innovation term $(U_t^{(1)}, U_t^{(2)}, \dots, U_t^{(m)})'$ appearing in the multivariate ARCH model (11) or the GARCH model (13) has the multivariate density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ defined as:

$$f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1}) = \frac{\phi_{\mathbf{P}}(\mathbf{g}(\mathbf{u})) \prod_{j=1}^m (1 + a_0^{(j)} u_j^2)^{-3/2}}{\int_{E_0} \phi_{\mathbf{P}}(\mathbf{y}) d\mathbf{y}} \quad \text{for all } \mathbf{u} \in \mathbf{R}^m \quad (17)$$

where $\mathbf{u} = (u_1, \dots, u_m)'$, $\mathbf{a}_0 = (a_0^{(1)}, \dots, a_0^{(m)})'$, and $\mathbf{g}(\mathbf{u}) = (g_1(\mathbf{u}), \dots, g_m(\mathbf{u}))'$ with $g_j(\mathbf{u}) = u_j / \sqrt{1 + a_0^{(j)} u_j^2}$ for $j = 1, \dots, m$; note that, by construction, $\mathbf{g}(\mathbf{u})$ is in E_0 for all $\mathbf{u} \in \mathbf{R}^m$.

Therefore, our constant (with respect to time) conditional dependence ARCH/GARCH model is given by (11) or (13) respectively, together with the assumption that the vector residuals $(U_t^{(1)}, U_t^{(2)}, \dots, U_t^{(m)})'$ follow the heavy-tailed multivariate density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ defined in (17). Because the dependence in the new density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ is driven by the underlying (truncated) normal distribution (16), the number of parameters involved in $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ is $m + m(m-1)/2$, i.e., the length of the \mathbf{a}_0 vector plus the size of the lower triangular part of \mathbf{P} ; this number is comparable to the number of parameters found in Bollerslev's (1990) constant conditional correlation models.

The vector \mathbf{a}_0 is a shape parameter capturing the degree of heavy tails in each direction, while the entries of correlation matrix \mathbf{P} capture the multivariate dependence. Again note that the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ can be scaled to create a more general family of densities with typical member given by

$$f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{c}) = \frac{1}{\prod_{j=1}^m c_j} f_{\mathbf{P}}\left(\left(\frac{u_1}{c_1}, \dots, \frac{u_m}{c_m}\right)'; \mathbf{a}_0, \mathbf{1}\right) \quad \text{for all } \mathbf{u} \in \mathbf{R}^m, \quad (18)$$

where $\mathbf{c} = (c_1, \dots, c_m)'$ is the scaling vector.

4 NUMERICAL MAXIMUM LIKELIHOOD

The general density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{c})$ will be useful for the purpose of Maximum Likelihood Estimation (MLE) of the unknown parameters. Recall that the (pseudo)likelihood, i.e., conditional likelihood, in the case of the multivariate ARCH model (11) with residuals following the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ is given by the expression

$$\mathbf{L} = \prod_{t=p+1}^n f_{\mathbf{P}}(\mathbf{X}_t; \mathbf{a}_0, \mathbf{s}_t) \quad (19)$$

where $\mathbf{s}_t = (s_t^{(1)}, \dots, s_t^{(m)})'$ is the vector of volatilities. The (pseudo) likelihood in the GARCH model (13) can be approximated by that of an ARCH model with order p that is sufficiently large; see also Politis (2004) for more details. As Hall and Yao (2003) recently showed, maximum (pseudo)likelihood estimation will be consistent even in the presence of heavy-tailed errors albeit possibly at a slower rate.

As usual, to find the MLEs one has to resort to numerical optimization.⁴ Note that in the case of fitting the GARCH(1,1) model (13), one has to optimize over $4m + m(m-1)/2$ free parameters; the $4m$ factor comes from the parameters $A^{(j)}, B^{(j)}, C^{(j)}$ and $a_0^{(j)}$ for $j = 1, \dots, m$, while the $m + m(m-1)/2$ factor is the number of parameters in \mathbf{P} . As in all numerical search procedures, it is crucial to get good starting values.⁵ In this case, it is

⁴An alternative avenue was recently suggested by Bose and Mukherjee (2003) but it is unclear if/how their method applies to a multivariate ARCH model without finite fourth moments.

⁵As with all numerical optimization procedures, the algorithm must run with a few different starting values to avoid getting stuck at a local—but not global—optimum.

recommended to get starting values for $A^{(j)}, B^{(j)}, C^{(j)}$ and $a_0^{(j)}$ by fitting a univariate GARCH(1,1) model to the j th coordinate; see Politis (2004) for details. A reasonable starting value for the correlation matrix \mathbf{P} is the identity matrix. An alternative—and more informative—starting value for the j, k element of matrix \mathbf{P} is the value of the cross-correlation at lag 0 of series $\{X_t^{(j)}\}$ to series $\{X_t^{(k)}\}$.

Nevertheless, even for m as low as 2, the number of parameters is high enough to make a straightforward numerical search slow and impractical. Furthermore, the gradient (and Hessian) of $\log \mathbf{L}$ are very cumbersome to compute, inhibiting the implementation of a ‘steepest ascent’ algorithm.

For this reason, we now propose an numerical MLE algorithm based on an approximate ‘steepest ascent’ together with a ‘divide-and-conquer’ idea. To easily describe it, let

$$\theta^{(j)} = (A^{(j)}, B^{(j)}, C^{(j)}, a_0^{(j)})' \quad \text{for } j = 1, \dots, m.$$

In addition, divide the parameters in the (lower triangular part of) matrix \mathbf{P} in M groups labeled $\theta^{(m+1)}, \theta^{(m+2)}, \dots, \theta^{(m+M)}$, where M is such that each of the above groups does not contain more than (say) three or four parameters; for example, as long as $m \leq 3$, M can be taken to equal 1.

Our basic assumption is that we have at our disposal a numerical optimization routine that will maximize $\log \mathbf{L}$ over $\theta^{(j)}$, for every j , when all other $\theta^{(i)}$ with $i \neq j$ are held fixed. The proposed DC–profile MLE algorithm is outlined below.⁶

DC–PROFILE MLE ALGORITHM

Step 0. Let $\theta_0^{(j)}$ for $j = 1, \dots, m + M$ denote starting values for $\theta^{(j)}$.

Step $k+1$. At this step, we assume we have at our disposal the k th approximate MLE solution $\theta_k^{(j)}$ for $j = 1, \dots, m + M$, and want to update it to obtain the $(k + 1)$ th approximation $\theta_{k+1}^{(j)}$. To do this, maximize $\log \mathbf{L}$ over $\theta^{(j)}$, for every j , when all other $\theta^{(i)}$ with $i \neq j$ are held fixed at $\theta_k^{(i)}$; denote by $\tilde{\theta}_k^{(j)}$ for $j = 1, \dots, m + M$ the results of the $m + M$ ‘profile’ maximizations. We then define the updated value $\theta_{k+1}^{(j)}$ for $j = 1, \dots, m + M$ as a ‘move in

⁶The initials DC stand for divide-and-conquer; some simple S+ functions for GARCH(1,1) numerical MLE in the univariate case $m = 1$ and the bivariate case $m = 2$ are available at: www.math.ucsd.edu/~politis.

the right direction’, i.e.,

$$\theta_{k+1}^{(j)} = \lambda \tilde{\theta}_k^{(j)} + (1 - \lambda) \theta_k^{(j)} \quad (20)$$

i.e., a convex combination of $\tilde{\theta}_k^{(j)}$ and $\theta_k^{(j)}$. The parameter $\lambda \in (0, 1)$ is chosen by the practitioner; it controls the speed of convergence—as well as the risk of nonconvergence—of the algorithm. A value λ close to one intuitively corresponds to maximum potential speed of convergence but also maximum risk of nonconvergence.

Stop. The algorithm stops when relative convergence has been achieved, i.e., when $\|\theta_{k+1} - \theta_k\| / \|\theta_k\|$ is smaller than some prespecified small number; here $\|\cdot\|$ can be the Euclidean or other norm.

The aforementioned maximizations involve numerical searches over a given range of parameter values; those ranges must also be carefully chosen. A concrete recommendation that has been found to speed up convergence of the profile MLE algorithm is to let the range in searching for $\theta_{k+1}^{(j)}$ be centered around the previous value $\theta_k^{(j)}$, i.e., to be of the type $[\theta_k^{(j)} - b_k, \theta_k^{(j)} + B_k]$ for two positive sequences b_k, B_k that tend to zero for large k reflecting the ‘narrowing-down’ of the searches. For positive parameters, e.g., for $\theta_k^{(j)}$ with $J \leq m$ in our context, the range can be taken as $[d_k \theta_k^{(j)}, D_k \theta_k^{(j)}]$ for two sequences d_k, D_k . The sequence d_k is increasing, tending to 1 from below, while D_k is a decreasing sequence tending to 1 from above; the values $d_1 = 0.7, D_1 = 1.5$ are reasonable starting points.

Remark. The DC–profile optimization algorithm may be applicable to general multiparameter optimization problems characterized by a difficulty in computing the gradient and Hessian of a general objective function \mathbf{L} . The idea is to ‘divide’ the arguments of \mathbf{L} into (say) the m groups $\theta^{(j)}$ for $j = 1, \dots, m$. The number of parameters within each group must be small enough such that a direct numerical search can be performed to accomplish (‘conquer’) the optimization of \mathbf{L} over $\theta^{(j)}$, for every j , when all other $\theta^{(i)}$ with $i \neq j$ are held fixed. The m individual (‘profile’) optimizations are then combined as in eq. (20) to point to a direction of steep (if not steepest) ascent.

5 THE BIVARIATE PARADIGM

For illustration, we now focus on a random vector $\mathbf{U} = (U^{(1)}, U^{(2)})'$ that is distributed according to the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ in the bivariate case $m = 2$. In this case, the correlation matrix $\mathbf{P} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where ρ is a correlation coefficient. Interestingly, as in the case of normal data, the strength of dependence between $U^{(1)}$ and $U^{(2)}$ is measured by this single number ρ . In particular, $U^{(1)}$ and $U^{(2)}$ are independent if and only if $\rho = 0$.

As mentioned before, if $\mathbf{a}_0 \rightarrow \mathbf{0}$, then $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1}) \rightarrow \phi_{\mathbf{P}}(\mathbf{u})$. In addition, $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ is unimodal, and has even symmetry around the origin. Nevertheless, when $\mathbf{a}_0 \neq \mathbf{0}$, the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ is heavy-tailed and non-normal. To see this, note that the contours of the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ with $\mathbf{a}_0 = (0.1, 0.1)'$ and $\rho = 0$ are not circular as in the normal, uncorrelated case; see Figure 1 (i).

Parts (ii), (iii) and (iv) of Figure 1 deal with the case $\rho = 0.5$, and investigate the effect of the shape/tail parameter \mathbf{a}_0 . Figure 1 (iv) where $\mathbf{a}_0 = (0, 0)'$ is the normal case with elliptical contours; in Figure 1 (ii) we have $\mathbf{a}_0 = (0.1, 0.1)'$, and the contours are ‘boxy’-looking, and far from elliptical. The case of Figure 1 (iii) is somewhere in-between since $a_0^{(1)} = 0$ but $a_0^{(2)} \neq 0$.

Comparing the contours of $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ to that of the normal $\phi_{\mathbf{P}}(\mathbf{u})$ we come to an interesting conclusion: the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ looks very much like the $\phi_{\mathbf{P}}(\mathbf{u})$ for small values of the argument \mathbf{u} . As expected, the similarity breaks down in the tails; in some sense, $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ is like a normal density deformed in a way that its tails are drawn out.

Figure 2 further investigates the role of the parameters \mathbf{a}_0 and ρ . Increasing the degree of heavy tails by increasing the value of \mathbf{a}_0 leads to some interesting contour plots, and this is especially true out in the tails.⁷ Similarly, increasing the degree of dependence by increasing the value of ρ leads to unusual contour plots especially in the tails. As mentioned before, the contours of $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ are close to elliptical for small values of \mathbf{u} , but acquire unusual shapes in the tails.

Finally, Figure 2 (iii) shows an interesting interplay between the values of \mathbf{a}_0 and the correlation coefficient ρ . To interpret it, imagine one has n

⁷In both Figures 1 and 2 the same level values were used corresponding to the density taking the values 0.001, 0.005, 0.010, 0.020, 0.040, 0.060, 0.080, 0.100, 0.120, 0.140, 0.160, 0.180, and 0.200.

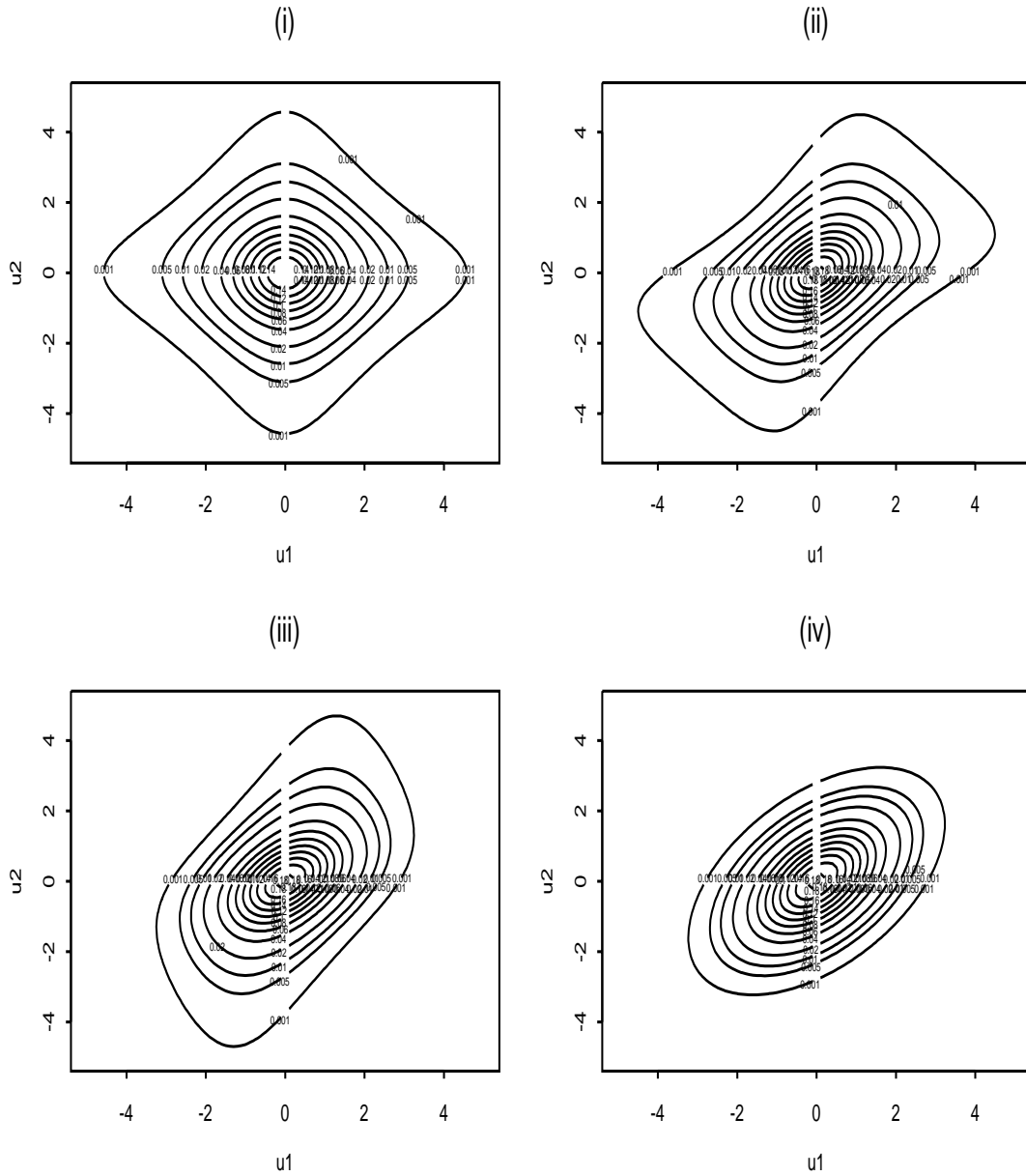


Figure 1: Contour (level) plots of the bivariate density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$. (i) $\mathbf{a}_0 = (0.1, 0.1)'$, $\rho = 0$. (ii) $\mathbf{a}_0 = (0.1, 0.1)'$, $\rho = 0.5$. (iii) $\mathbf{a}_0 = (0, 0.1)'$, $\rho = 0.5$. (iv) $\mathbf{a}_0 = (0, 0)'$, $\rho = 0.5$.

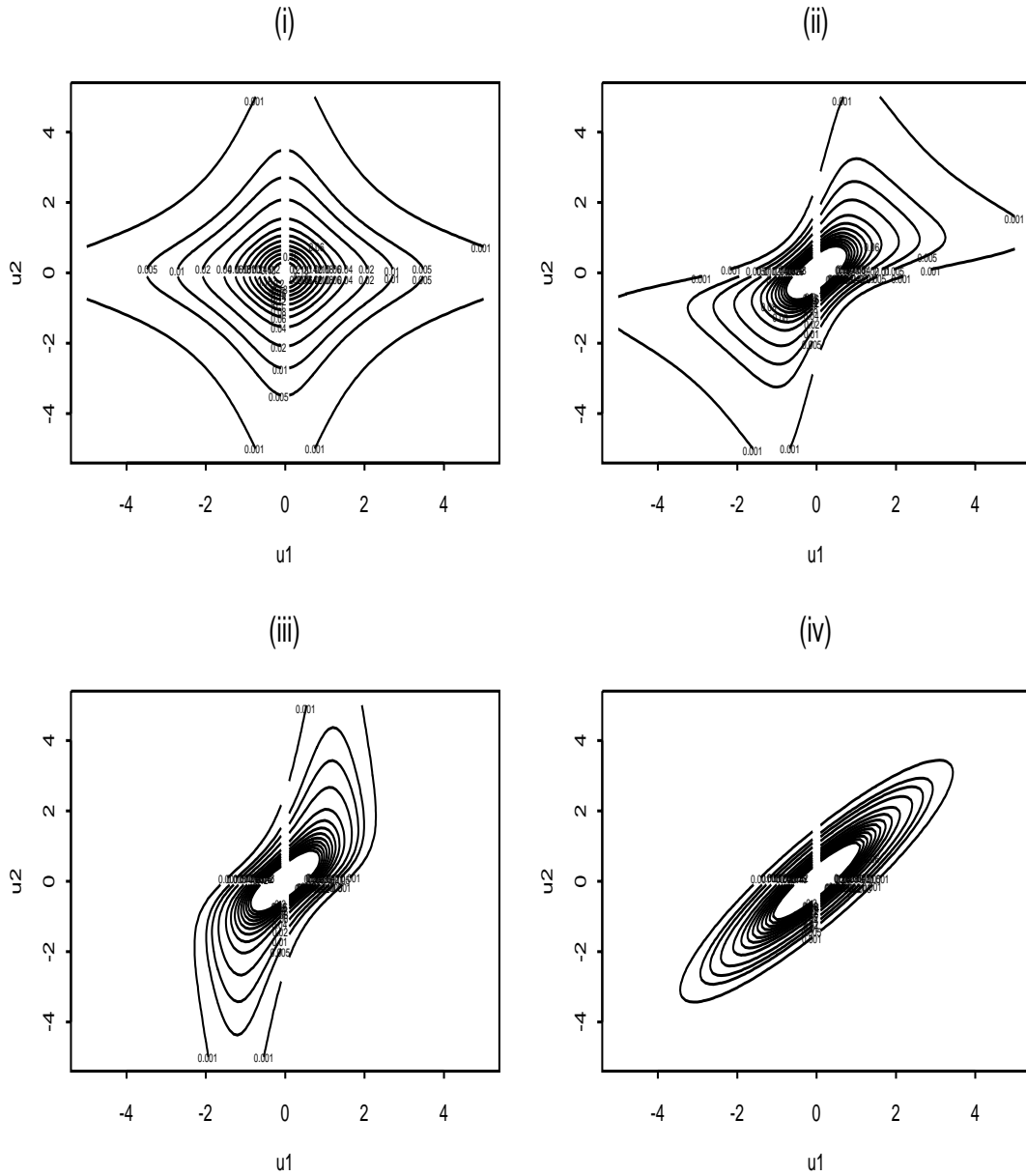


Figure 2: Contour (level) plots of the bivariate density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$. (i) $\mathbf{a}_0 = (0.5, 0.5)'$, $\rho = 0$. (ii) $\mathbf{a}_0 = (0.5, 0.5)'$, $\rho = 0.9$. (iii) $\mathbf{a}_0 = (0, 0.5)'$, $\rho = 0.9$. (iv) $\mathbf{a}_0 = (0, 0)'$, $\rho = 0.9$.

i.i.d. samples from density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$. If n is small, in which case the heavy tails of $U^{(2)}$ have not had a chance to fully manifest themselves in the sample, a fitted regression line of $U^{(2)}$ on $U^{(1)}$ will have a small slope dictated by the elliptical contours around the origin. By contrast, when n is large, the slope of the fitted line will be much larger due to the influence of the tails of $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, \mathbf{1})$ in the direction of $U^{(2)}$.

6 A REAL DATA EXAMPLE

We now focus on a real data example. The data consist of daily returns of the IBM and Hewlett-Packard stock prices from February 1, 1984 to December 31, 1991; the sample size is $n = 2000$. This bivariate dataset is available as part of the `garch` module of the statistical language S+.

A full plot of the two time series is given in Figure 3; the crash of October 1987 is prominent around the middle of those plots. Figure 4 focuses on the behavior of the two time series over a narrow time window spanning two months before to two months after the crash of 1987. In particular, Figure 4 gives some visual evidence that the two time series are not independent; indeed, they seem to be positively dependent, i.e., to have a tendency to ‘move together’. This fact is confirmed by the autocorrelation/cross-correlation plot given in Figure 5; note that the cross-correlation of the two series at lag zero is large and positive (equal to 0.56).

Table 1 shows the MLEs of parameters in univariate GARCH(1,1) modeling of each returns series, while Table 2 contains the approximate MLEs of parameters in bivariate GARCH(1,1) modeling, i.e., entries of Table 2 were calculated using the DC-profile MLE algorithm of Section 4. Although more work on the numerical computability of multivariate MLEs in our context is in order—including gauging convergence and accuracy—it is interesting to note that the entries of Table 2 are markedly different from those of Table 1 indicating that the bivariate modeling is informative. For example, while the sum $\hat{A} + \hat{B}$ was found less than one for both series in the univariate modeling of Table 1, the same sum was estimated to be slightly over one for the IBM series and significantly over one for the H-P series in Table 2; this finding reinforces the viewpoint that financial returns may be lacking a finite second moment—see the discussion in Section 2 and Politis (2004). In addition, the high value of $\hat{\rho}$ in Table 2 confirms the intuition that the two series are

strongly (and positively) dependent.

	\hat{a}_0	\hat{A}	\hat{B}	\hat{C}
IBM	0.066	0.029	0.912	6.32e-06
H-P	0.059	0.052	0.867	2.54e-05

Table 1. MLEs of parameters in *univariate* GARCH(1,1) modeling of each returns series.

	\hat{a}_0	\hat{A}	\hat{B}	\hat{C}	$\hat{\rho}$
IBM	0.105	0.055	0.993	7.20e-05	0.53
H-P	0.062	0.258	0.975	3.85e-04	

Table 2. MLEs of parameters in *bivariate* GARCH(1,1) modeling of the two series; entries of Table 2 were calculated using the DC-profile MLE algorithm of Section 4.

Table 1 shows the MLEs of parameters in univariate GARCH(1,1) modeling of each returns series, while Table 2 contains the approximate MLEs of parameters in bivariate GARCH(1,1) modeling, i.e., entries of Table 2 were calculated using the DC-profile MLE algorithm of Section 4. Although more work on the numerical computability of estimators in this multivariate context is in order—including gauging convergence and accuracy—it is interesting to note that the entries of Table 2 are markedly different from those of Table 1 indicating that the bivariate modeling is informative. For example, while the sum $\hat{A} + \hat{B}$ was found less than one for both series in the univariate modeling of Table 1, the same sum was estimated to be slightly over one for the IBM series and significantly over one for the H-P series in Table 2; this finding reinforces the viewpoint that financial returns may be lacking a finite second moment—see the discussion in Section 2 and Politis (2004). In addition, the high value of $\hat{\rho}$ in Table 2 confirms the intuition that the two series are strongly (and positively) dependent.

Figure 6 shows diagnostic plots associated with the DC-profile MLE algorithm. The first 9 plots show the evolution of the 9 parameter estimates over iterations; the 9 parameters are: a_0, A, B, C for the IBM series; a_0, A, B, C

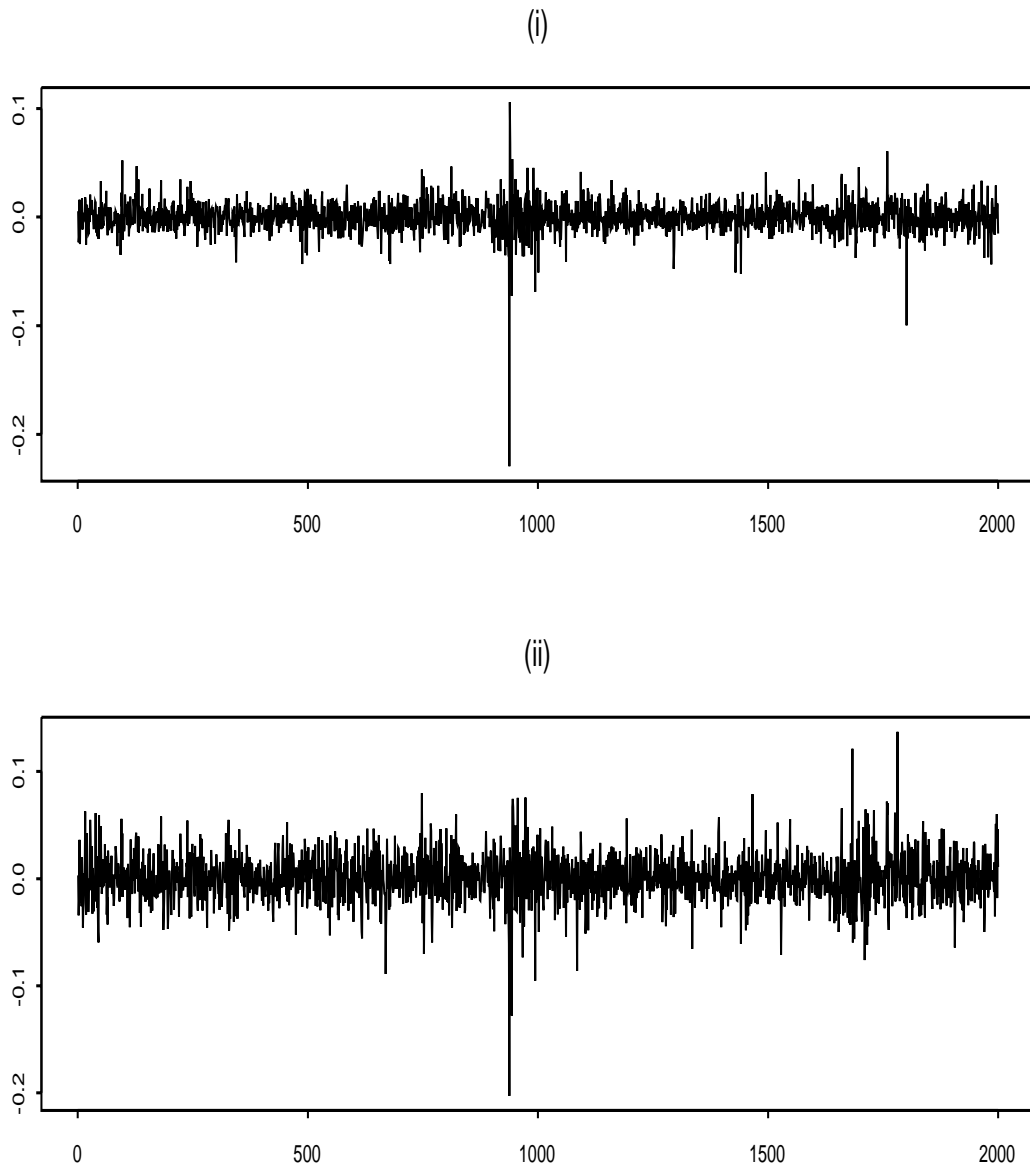


Figure 3: Daily returns from February 1, 1984 to December 31, 1991.
(i) IBM stock returns. (ii) Hewlett-Packard stock returns.

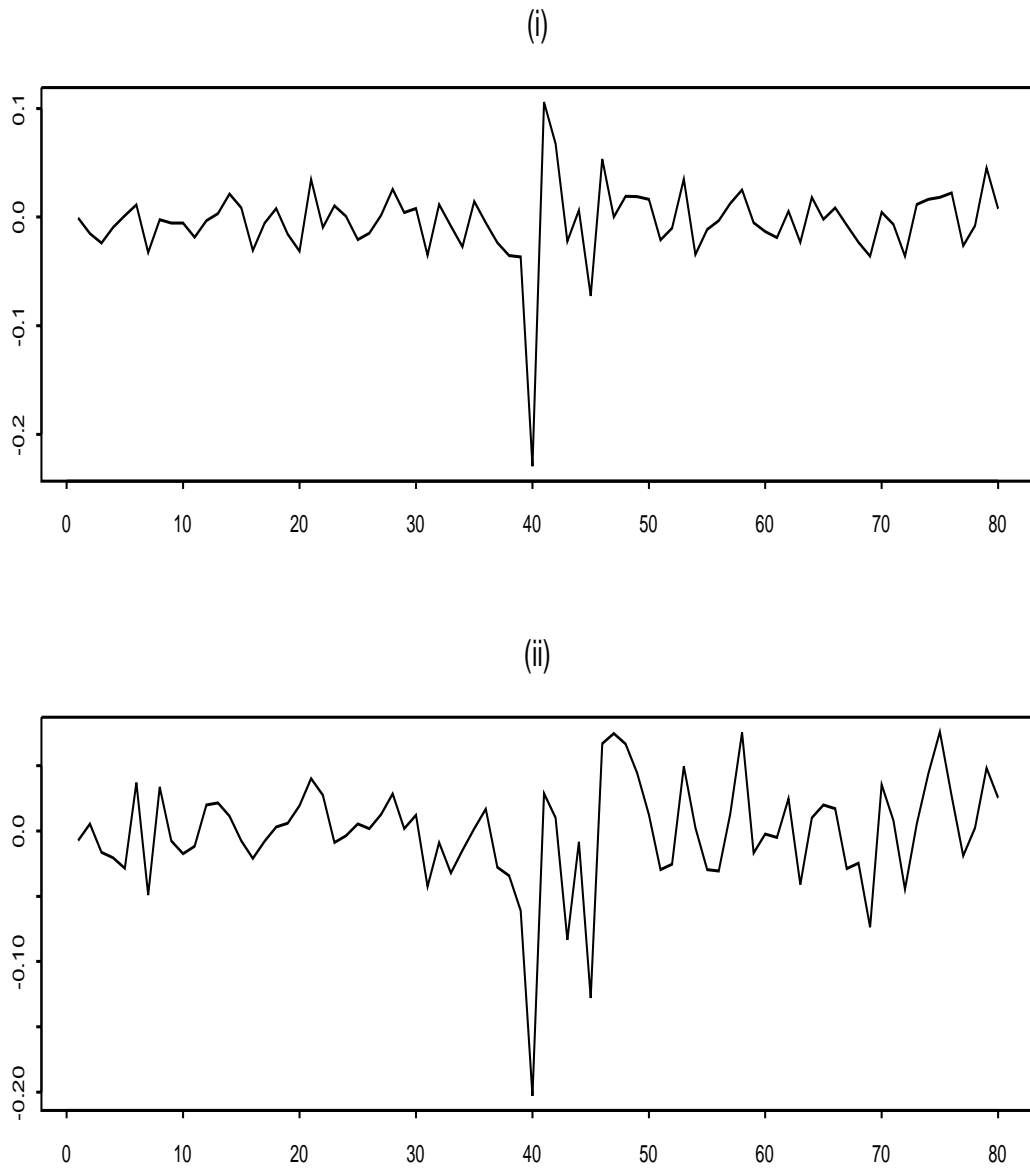


Figure 4: Daily returns from two months before to two months after the crash of 1987. (i) IBM stock returns. (ii) Hewlett-Packard stock returns.

Multivariate Series : IBMvs.HP

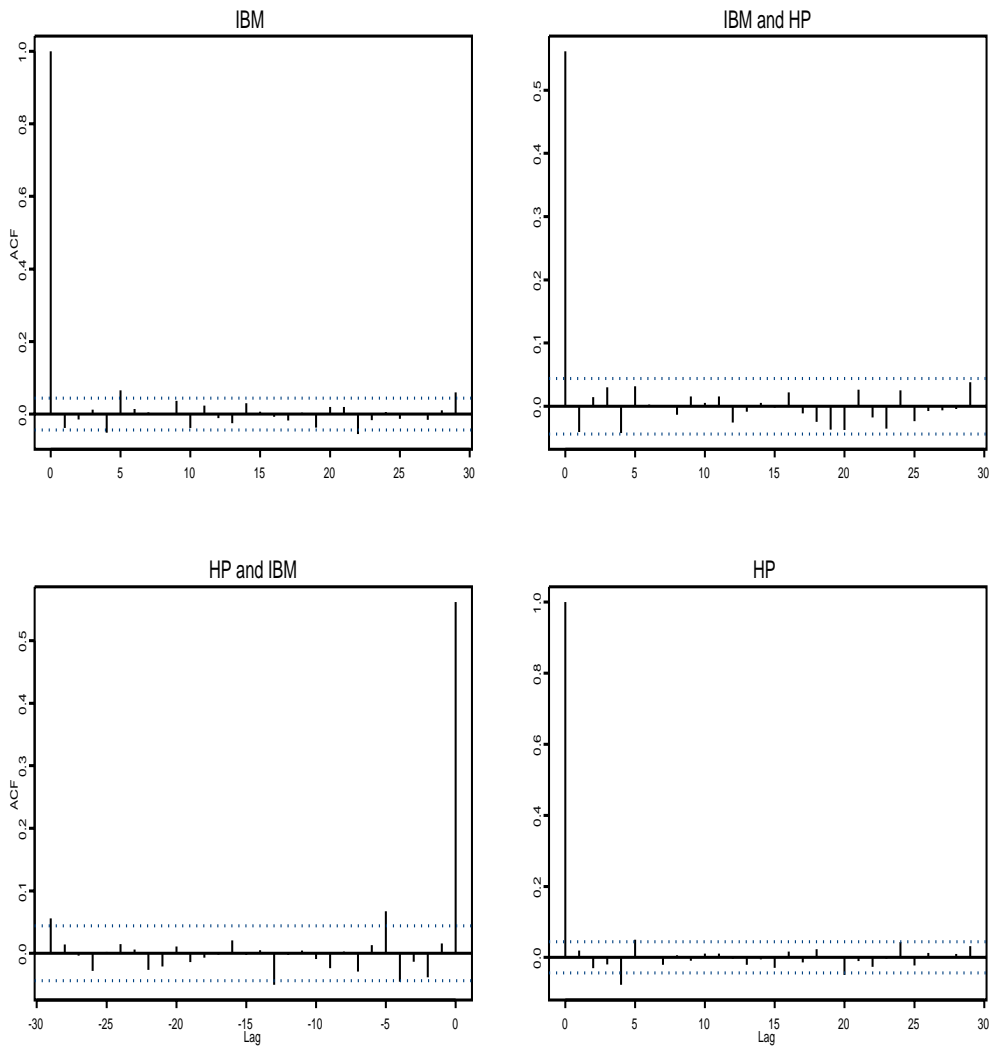


Figure 5: Autocorrelation/cross-correlation plot of IBM vs. Hewlett-Packard daily returns from February 1, 1984 to December 31, 1991.

for the H-P series, and ρ . Those 9 plots show convergence of the DC-profile MLE procedure; this is confirmed by the 10th plot that shows the evolution of the value of the Log-Likelihood corresponding to those 9 parameters. As expected, the 10th plot shows an increasing graph that tapers off towards the right end, showing the maximization is working; only about 30-40 iteration steps were required for convergence of the algorithm.

Finally, note that few people would doubt that a $\hat{\rho}$ value of 0.53, arising from a dataset with sample size 2000, might not be significantly different from zero; nonetheless, it is important to have a hypothesis testing procedure for use in other, less clear, cases. A natural way to address this is via the subsampling methodology—see Politis et al. (1999) for details. There are two complications, however:

- (i) Getting $\hat{\rho}$ is already quite computer-intensive, involving a numerical optimization algorithm; subsampling entails re-calculating $\hat{\rho}$ over subsets (sub-series) of the data, and can thus be too computationally expensive.
- (ii) The estimators figuring in Tables 1 and 2 are not \sqrt{n} -consistent as shown in Hall and Yao (2003); subsampling with estimated rate of convergence could then be employed as in Ch. 8 of Politis et al. (1999).

Future work may focus on the above two important issues as well as on the possibility of deriving likelihood-based standard errors and confidence intervals in this non-standard setting.

References

- [1] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.
- [2] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Review of Economics and Statistics*, 72, 498-505.
- [3] Bollerslev, T., Chou, R. and Kroner, K. (1992). ARCH modelling in finance: a review of theory and empirical evidence, *Journal of Econometrics*, 52, 5-60.

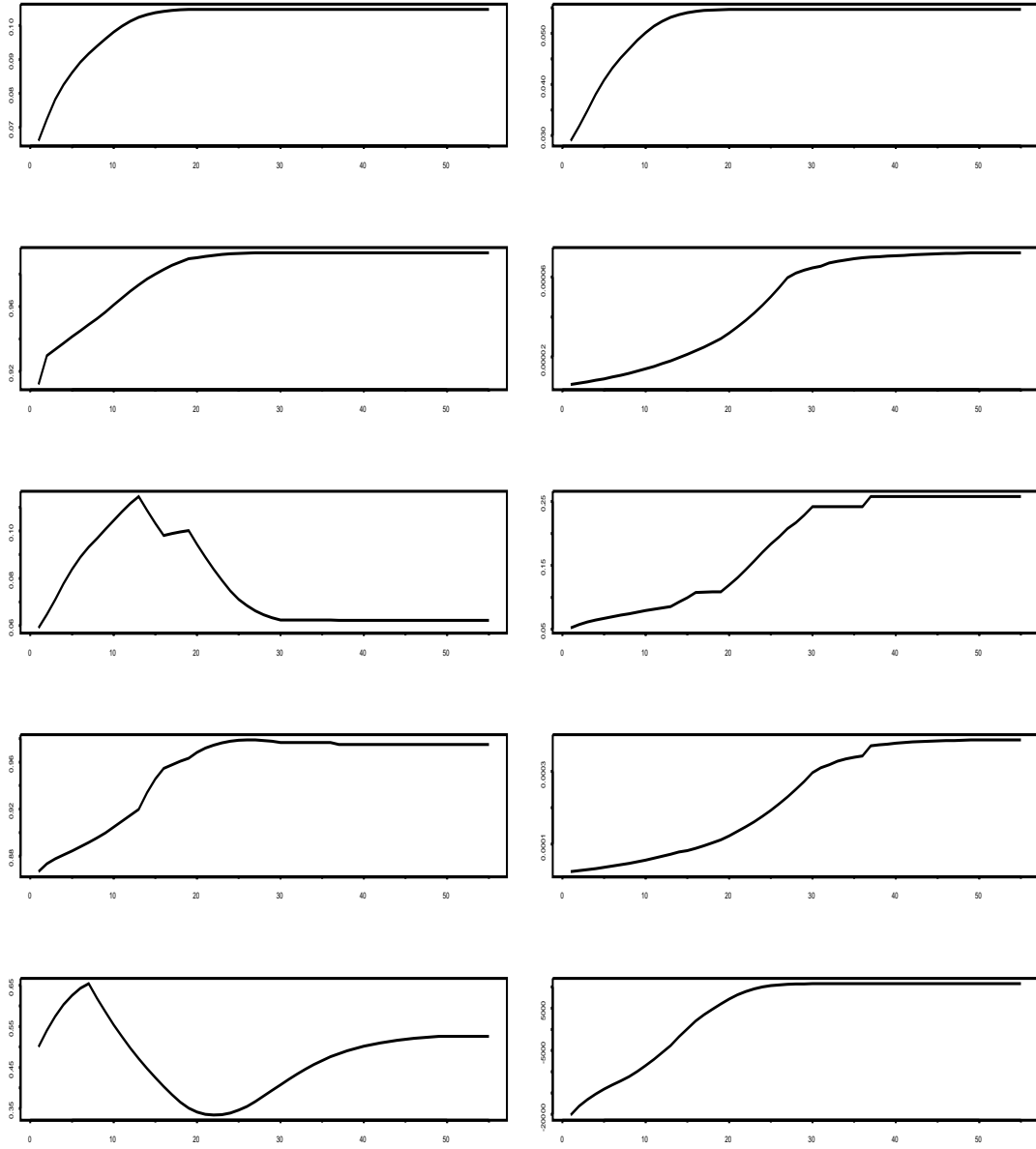


Figure 6: Diagnostic plots associated with the profile MLE algorithm; the first 9 plots show the evolution of the 9 parameter estimates over iterations, while the 10th plot shows the evolution of the value of the Log-Likelihood.

- [4] Bose, A. and Mukherjee, K. (2003), Estimating the ARCH parameters by solving linear equations, *J. Time Ser. Anal.*, vol. 24, no. 2, pp. 127–136.
- [5] Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation, *Econometrica*, 50, 987-1008.
- [6] Gouriéroux, C. (1997). *ARCH Models and Financial Applications*, Springer Verlag, New York.
- [7] Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors, *Econometrica*, 71, 285-317.
- [8] Patton, A.J. (2002). Skewness, asymmetric dependence and portfolios. Preprint (can be downloaded from: <http://fmg.lse.ac.uk/~patton>).
- [9] Politis, D.N. (2003), Model-free volatility prediction, Discussion Paper 2003-16, Department of Economics, UCSD.
- [10] Politis, D.N. (2004), A heavy-tailed distribution for ARCH residuals with application to volatility prediction, *Annals of Economics and Finance*, 5, 283-298.
- [11] Politis, D.N., Romano, J.P., and Wolf, M. (1999) *Subsampling*. Springer, New York.
- [12] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility, in *Time Series Models in Econometrics, Finance and Other Fields*, D.R. Cox, David V. Hinkley and Ole E. Barndorff-Nielsen (eds.), Chapman & Hall, London, pp. 1-67.
- [13] Vrontos, I.D., Dellaportas, P. and Politis, D.N. (2003). A full-factor multivariate GARCH model, *Econometrics Journal*, 6, 312-334.