Fourth-Moments Structures in Financial Time Series

Robert M. Kunst

Forschungsbericht/
Research Memorandum No. 335

October 1993

ABSTRACT

I consider a class of conditionally heteroskedastic models that comprises most linear "ARCH"-type models found in the literature. This class is especially motivated by the fact that two basic kinds of ARCH processes have been suggested in autocorrelated circumstances: Engle (1982) explains conditional variance by lagged errors, Weiss (1984) also by lagged observations. The general framework permits an evaluation of whether the restrictions evolving from the Engle or the Weiss models are valid in practice. My empirical example is a time series of 7000 observations of the Standard & Poor index including the "lunes negro" crash. Evidence is collected from parametric estimation of the outlined models and from an evaluation of descriptive fourth-moments estimates, for which significance bounds are established by means of algebra and simulation.

ZUSAMMENFASSUNG

Ich betrachte eine Klasse bedingt heteroskedastischer Modelle, welche die meisten linearen ARCH-Modelle der ökonometrischen Literatur umfasst. Diese Klasse wird besonders durch die Tatsache motiviert, dass zwei grundlegende Formen von ARCH-Prozessen vorgeschlagen wurden im Falle autokorrelierter Daten. Engle (1982) erklärt bedingte Varianz durch verzögerte Fehler, Weiss (1984) auch durch verzögerte Beobachtungen. Das allgemeine Rahmenwerk erlaubt eine Auswertung davon, ob die durch die Engle- oder Weiss-Modelle auferlegten Restriktionen in der Praxis gelten. Mein empirisches Beispiel ist eine Zeitreihe von 7000 Beobachtungen des Standard & Poor 500 Index, die den "lunes negro"-Börsensturz enthält. Evidenz wird gesammelt sowohl aus Schätzungen im parametrischen Kontext als auch durch deskriptive empirische vierte Momente, deren Verteilung durch Simulation und Algebra approximiert wird.

1. Introduction

For some time now, scientific interest in serially correlated volatility has been soaring. This interest is concentrating primarily on financial series where prediction of means is notoriously unrewarding and hence structure, if any, is to be found through higher-moments properties only. Let us take for example prices of common stocks or stock market indicators. These series are well known to approximate random walks, hence their first differences are unpredictable while forecasts on the series itself are provided by the latest observation plus an eventual "drift constant". A closer look at such series reveals, however, that they show noteworthy temporal clusters of volatility. Changing conditional second moments tend to invalidate the basic "random-walk model" in favor of the more general "martingale" model.

The best known statistical model for the volatility-clustering phenomenon is Engle's (1982) ARCH model, with ARCH standing for "autoregressive conditional heteroskedasticity". The acronym stresses that the model is designed to parallel the central position that the AR and more general Box-Jenkins models have in linear time series analysis. This paper departs from the observation that this model is primarily a model for "white noise" data without serial correlation structure "in means". Problems arising from the reconciliation of linear structure with conditional volatility, already outlined in the original paper by Engle, have been tackled in a slightly different way by Weiss (1984) and were taken up recently by Bera et al. (1992).

This paper is organized as follows. Section 2 presents a conditionally heteroskedastic model class which encompasses most linear ARCH models known from the literature. Conditions for covariance stationarity are given. Section 3 is concerned with issues related to the problem of estimating the parameters of structures as given in the previous section via maximum likelihood. First analytical derivatives are given with respect to all parameters. Section 4 reports the results from attempts to fit the more general ARCH model class to the Standard & Poor 500 Index series. The findings appear to indicate that real-life ARCH structures are richer than allowed by the restrictive classical ARCH models. Section 5 presents some simulation results related to the question under what conditions the more general ARCH class generates strictly stationary solutions. Section 6 develops the tools for evaluating fourth-cross-moment structures directly and applies them to the Standard & Poor 500 Index series. Section 7 concludes.

2. A more general ARCH model

The model class suggested here has the following form

$$\mathrm{E}(\varepsilon_t^2 | I_{t-1}) = h_t = a_0 + \sum_i \sum_{j=1}^{i} a_{ij} \varepsilon_{t-i} \varepsilon_{t-j} \qquad (2.1)$$

with the index i running up to some finite bound R or to ∞ and I_t denoting an information set containing all ε_s for $s \le t$. In contrast to linear time series analysis, this I_t is non-linear in the sense that it contains e.g. information on ε_s as well as on ε_s^2. In other words, the expectations operator stands for conditional expectations per se and cannot be conceived in the usual simplified manner as linear projection. Of course, expectation is still linear with respect to the set of all cross-products, a property which will be exploited later. This model class is similar to the AARCH and GAARCH models introduced by Bera et al. (1992) who derive them, however, indirectly from a random coefficient structure.

In order that (2.1) should make sense, it must obey certain restrictions. In particular, the following assumptions warrant that h_t is a well-defined conditional variance process:

ASSUMPTION 1: $a_0 > 0$

ASSUMPTION 2: The array **B**, for convenience formed from the coefficients a_{ij} in such a way that

$$b_{ii} = a_{ii}$$

$$b_{ij} = a_{ij}/2 \text{ for } i > j \tag{2.2}$$

$$b_{ij} = b_{ji} \quad \text{for } i < j$$

is non-negative definite in the sense that all finite-dimensional square symmetric submatrices are non-negative definite.

These assumptions warrant that h_t is strictly positive whatever the values in the sequence of past ε_s, $s \leq t$ are. Weakening assumption 1 to $a_0 \geq 0$ would allow for an eventual degeneration of the errors process.

Whereas Assumptions 1-2 guarantee that the definition makes sense, they do not establish that there is a stationary solution to (2.1). To that aim, we need the following condition:

ASSUMPTION 3a: The roots of the polynomials in the sequence of characteristic polynomials formed by the diagonal elements of **B** are bounded away from the unit circle, i.e. the modulus of any root is greater than $1+\delta$ for some $\delta > 0$.

ASSUMPTION 3b: The sum $\Sigma\, a_{ii} = a$ converges.

It is well known that instead of these two conditions one could require simply that Σa_{ii} converge to a limit smaller than one. Together with the following description of the stochastic process ε_t, this assumption is crucial for the existence of a covariance-stationary solution.

ASSUMPTION 4: The conditional distribution of ε_t is symmetric in the sense that

$$\mathrm{E}\,(\varepsilon_t | I_{t-1}) = 0 \qquad \text{for all } t \tag{2.3}$$

ASSUMPTION 5: The conditional distribution of ε_t given I_{t-1} is normal.

Even though Assumptions 1-4 would generally suffice to establish the following results together with some more general regularity condition, I will assume for the moment that Assumptions 1-5 hold in order to facilitate the presentation.

THEOREM 1: Under the Assumptions 1-5, (2.1) has a covariance-stationary solution

Proof: This is perhaps most easily proved if Engle's (1982,p.1005) idea is adopted. First we assume that there is an upper bound to the indices in (2.1), say R. Then we stack the elements of concern into an $R(R+1)/2$-vector in the following way

$$w_t' = (\varepsilon_t^2, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-R+1}^2, \varepsilon_t\varepsilon_{t-1}, \varepsilon_{t-1}\varepsilon_{t-2}, \ldots, \varepsilon_{t-R+2}\varepsilon_{t-R+1}, \varepsilon_t\varepsilon_{t-2}, \varepsilon_{t-1}\varepsilon_{t-3}, \ldots, \varepsilon_{t-1}\varepsilon_{t-R+1})$$

The vector containing the main diagonal is followed by the elements of the first subdiagonal, then the second subdiagonal etc. Then we can re-write (2.1) as

$$\mathrm{E}(w_t | I_{t-1}) = a_V + \mathbf{M}w_{t-1}$$

with $a_V' = (a_0, 0, \ldots, 0)$ and with $\mathbf{M}$ constructed as follows

a_{11}	a_{22}	...	$a_{R-1,R-1}$	a_{RR}	a_{21}	a_{32}	...	$a_{R-1,R-2}$	$a_{R,R-1}$	a_{31}	...	$a_{R,2}$	$a_{R,1}$
1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	...	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	...	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0

Apart from the first row, $\mathbf{M}$ contains a sub-diagonal $(R\text{-}1)\mathrm{x}(R\text{-}1)$ identity matrix filled to the right margin with zeros, then a row of zeros, then another $(R\text{-}2)\mathrm{x}(R\text{-}2)$ identity matrix, then a row of zeros etc. until a row with a single 1 and then a zero finishes the diagram. The remainder is filled with zeros. Clearly, the asymptotic solution

$$\lim_{k\to\infty} E(w_t | I_{t-k}) = (\mathbf{I}\text{-}\mathbf{M})^{-1} a_V$$

remains valid. Moreover, this solution represents a covariance-stationary solution if and only if the roots of the characteristic polynomial formed just by the a_{ii} lie outside the unit circle. It is easily seen from the above matrix by expanding along the first row that the remaining a_{ij} $(i>j)$ do not affect this property, i.e. the eigenvalues of **M** are entirely determined by the a_{ii} elements. Assumption 4 is crucial as otherwise the rows $R+1,2R,\ldots$ would not necessarily be zero.

Now suppose $R\to\infty$. Then we can approximate the behavior of w_t by that of finitely-structured w_{Rt} arbitrarily well and, using Assumption 3, stationarity is established in analogy to the linear autoregressive model. ■

Note in particular that stationarity is unaffected by off-diagonal elements in **B**. Off-diagonals in **B** are, however, severely restricted by non-negative definiteness and hence the diagonal of **B** typically dominates the array. Whereas Assumptions 1-4 are more or less minimal conditions for covariance stationarity, strict stationarity may hold in more general circumstances. Nelson (1990) has shown that conditional normality allows for strictly stationary solutions of certain models of type (2.1) not only for $a=1$ in Assumption 3b but also for slightly larger values. A similar behavior was found by Kunst (1993) for certain bivariate ARCH structures. In those cases, naturally unconditional variances do not exist. Nelson also shows that for other conditional distributions (2.3) is not even sufficient to guarantee a strictly stationary solution. An extension to higher-order GARCH models is provided by Bougerol and Picard (1992).

The model class (2.1) comprises most linear heteroskedastic models presented in the literature, in detail:

1] The ARCH model by Engle (1982) with only finitely many a_{ii} different from 0. Definiteness is warranted by $a_{ii} \geq 0$ as **B** is a diagonal matrix.

2] The GARCH model by Bollerslev (1986) has infinitely many non-zero a_{ii} which are, however, finitely parameterized by a ratio of lag polynomials. Again, all off-diagonals are 0 and definiteness is easily checked by the diagonal elements. It is worth noting that positiveness of all non-zero polynomial coefficients is not necessary for Assumption 2 to hold but is usually assumed for simplicity or for numerical reasons.

3] The univariate version of the time series ARCH model suggested by Kunst and Polasek (1993) to model interest rates assumes that e_t is the innovations process from a linear time series model for the data process y_t. If then h_t depends on lagged observations instead of squared innovations as in

$$h_t = a_0 + \sum_{i=1}^{R} a_i (y_{t-i} - \mu)^2 \qquad (2.4)$$

with μ denoting the possibly non-zero mean of the data process, substitution of the linear time series model into y_t immediately renders the basic form (2.1). If all $a_i > 0$, then **B** is again n.n.d. because the transformation from the innovations ε_t has been linear.

4] The (slightly simplified) ARMA-ARCH model by Weiss (1984) essentially merges the previous [1] models into

$$h_t = a_0 + \sum_{i=1}^{R} a_i (y_{t-i} - \mu)^2 + \sum_{i=1}^{S} c_i \varepsilon_{t-i}^2 \tag{2.5}$$

Definiteness is ensured by $a_i \geq 0$, $c_i \geq 0$ because it is an amalgam of the previous models. It is maybe more interesting that stationarity conditions in (2.5) depend on the properties of the linear time series process. There is some kind of trade-off in the sense that more linear dependence can be tolerated if $c_1+\ldots+c_S$ is sufficiently smaller than 1. Theoretically, some non-negativity conditions could be relaxed but this does not appear to have any empirical impact.

3. Estimation issues

Here and in the following I will assume that $R<\infty$. Of course, this is not the only way to define a finite parameterization for estimation purposes and the possibility of models with infinite-dimensional arrays **B** which depend on a finite parameter set - such as GARCH models - should be considered. For the time being, however, I will assume that a finite matrix gives a reasonable approximation to the possibly infinite-dimensional **B**.

Although all ARCH likelihoods can be expressed in the simple way outlined by Engle (1982, p.990)

$$\begin{aligned} \ell &= T^{-1} \sum_{t=1}^{T} \ell_t \\ -\ell_t &= \tfrac{1}{2} \log h_t + \tfrac{1}{2}\, \varepsilon_t^2 / h_t \end{aligned} \tag{3.1}$$

straightforward numerical optimization of (3.1) can be time-consuming. If ε_t has to be estimated from some time-series model for an observed process y_t, h_t becomes a complicated function of lags of y_t and all parameters, i.e. the ARCH parameters of (2.1) as well as the parameters of the linear time-series model. Moreover, all stability and admissibility restrictions are non-linear inequality constraints that make estimation even more cumbersome.

An important simplification can be obtained if the information matrix is block-diagonal in the sense that there is no interaction between parameters of the linear time series model Θ_1 and $\Theta_2 = (a_0, a_{11}, a_{21}, a_{22}, \ldots)'$ of (2.1). In that case, solution of the ML problem can be decomposed into iterative steps of solving for either Θ_i ($i=1,2$) separately, conditional on

[1] Historically, of course, Kunst and Polasek (1993) used a simplified version of Weiss' older model and not the other way round.

the most recent parameters of the other Θ_i (i=2,1). Engle (1982, Theorem 4) stated some sufficient conditions for this property which he calls symmetry and regularity. (2.1) is Engle-regular in the sense that h_t is bounded away from zero by $a_0 > 0$ and that certain expectations of h_t derivatives exist. (2.1) is, however, not Engle-symmetric. Nonetheless, Bera et al. (1992) showed that block-diagonality of the information matrix evolves from symmetry in a much wider sense. In short, $h_t(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ is Bera-symmetric if changing of *all* ε_s to $-\varepsilon_s$ yields the same value while $h_t(.)$ is Engle-symmetric if this property holds for *any* change of individual ε_s to $-\varepsilon_s$.

Computer time can further be shortened by analytically evaluating scores i.e. derivatives of l_t with respect to the parameters. Before proceeding to that point, I would like to suggest a re-parameterization of (2.1) in order to replace the complicated admissibility restrictions into non-negativity constraints. Once this has been done, parameters can be replaced by their squares and estimation can be conducted without further constraints.

It is known from linear algebra that any symmetric non-negative definite $R \times R$-matrix **B** can be decomposed into **B**=**LDL'** where **L** is a lower triangular matrix with a unit diagonal and **D** is a diagonal matrix with positive elements and maybe some zeros on that diagonal (Banachiewicz decomposition). Similarly, any **LDL'** obeying to these restrictions defines a n.n.d. matrix **B** and thus instead of being concerned with the $R(R+1)/2$ parameters on and below the diagonal in **B** - which are equivalent to $a_{11}, a_{21}, a_{22}, \ldots, a_{RR}$ in (2.1) - we can look at the sub-diagonal elements of **L**, say $l_{21}, l_{31}, l_{32}, \ldots, l_{R,R-1}$, and at the R elements on the **D** diagonal. The n.n.d. constraint has been transformed into a simple non-negativity constraint on the **D** elements $d_1, \ldots, d_R$ and the whole model can be re-written as

$$h_t = \boldsymbol{\varepsilon}_{t-1}'\mathbf{LDL}'\boldsymbol{\varepsilon}_{t-1} \tag{3.2}$$

with $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \ldots, \varepsilon_{t-R+1})$. Further assuming this $\boldsymbol{\varepsilon}_t$ to be a linear autoregressive transformation of the observed variable vector $\mathbf{y}_t$ of length $R+S$, say, h_t can also be written in observed variables

$$h_t = \mathbf{y}_{t-1}'\mathbf{C}'\mathbf{LDL}'\mathbf{C}\,\mathbf{y}_{t-1} \tag{3.3}$$

with **C** being an Rx$(R+S)$-matrix with its i-th row containing the autoregressive coefficients $(1, \zeta_1, \ldots, \zeta_{S-1})$ flanked by i-1 zeros to the left and filled up with R-i+1 zeros to the right. Hence, the derivative with respect to ζ_i (i=1,...,S-1) is

$$\frac{\partial h_t}{\partial \zeta_i} = 2\mathbf{y}_{t-1}'\mathbf{C}'\mathbf{LDL}'\mathbf{y}_{t-i-1} \tag{3.4}$$

For technical reasons, the r.h.s. R-vector $\mathbf{y}_{t-i-1}$ has been trimmed from the end while the l.h.s vector $\mathbf{y}_{t-1}$ contains $R+S$ elements. Note that for **L**=**I** the original ARCH case (Engle, p.995) is recovered immediately.

(3.4) is an important factor in the derivative of the likelihood with respect to ζ_i. In detail,

$$\frac{\partial l_t}{\partial \zeta_i} = \frac{\varepsilon_t y_{t-i}}{h_t} + \frac{1}{2h_t}\frac{\partial h_t}{\partial \zeta_i}\frac{\varepsilon_t^2 - h_t}{h_t} \tag{3.5}$$

and substitution of (3.4) into (3.5) yields the required scores.

Derivatives with respect to the ARCH parameters are more straightforward as ε_t does not depend on them. In detail,

$$\frac{\partial h_t}{\partial d_i} = (\sum_{j=i}^{R} l_{ji}\varepsilon_{t-j})^2 \qquad \frac{\partial l_t}{\partial d_i} = \frac{\partial h_t}{\partial d_i}\frac{1}{2h_t}\frac{\varepsilon_t^2 - h_t}{h_t} \tag{3.6}$$

which can be adjusted easily if the unrestricted $\pm|d_i|$ is taken as parameter of concern instead of d_i. Similarly,

$$\frac{\partial h_t}{\partial l_{ij}} = 2d_j(\sum_{k=j}^{R} l_{kj}\varepsilon_{t-k})\varepsilon_{t-j} \qquad \frac{\partial l_t}{\partial d_i} = \frac{\partial h_t}{\partial d_i}\frac{1}{2h_t}\frac{\varepsilon_t^2 - h_t}{h_t} \tag{3.7}$$

The analytical derivatives (3.4)-(3.7) certainly save computing time without any loss of precision. A recent study by Calzolari and Fiorentini (1992) sheds some doubt, however, on the relative advantages of iterative optimization as the block-diagonality of the information matrix does not match an exact property of finite-sample Hessians. Nonetheless, in the presence of more lavish parameterization, some precision would be worth trading for lots of computing time.

4. An example: Standard & Poor's Index

A nice long series for an evaluation of conditionally heteroskedastic models is the Standard and Poor's Index S&P. Daily observations during the time period from July 2nd, 1962 to December 31st, 1990 permit a sample of 7168. Figure 1a shows a time series graph of the series after taking logarithms and Figure 1b shows its first differences. [2] In the linear time series framework, S&P comes close to a random walk. In a sample of 7168, however, even an R^2 of 0.05 indicates statistical rejection of the pure random walk model. In particular, the differences show significant first- and fifth-order autocorrelation, the latter order corresponding to a frequency of five trading days or a week. In contrast, conditional heteroskedasticity within the series is strong, with GARCH(1,1) models yielding parameter estimates close to the so-called IGARCH boundary where error variances become infinite even if conditional Gaussianity holds. The rather large sample should allow to give some insight on whether the more general approach suggested in (2.1) is justified relative to the original ARCH model and whether e.g. restrictions as in the Weiss model hold if the standard ARCH appears insufficient.

[2] All subsequent analysis is related to these first log differences which are therefore simply called "the S&P Index series".

First, a linear first-order autoregression was tried on the differenced series y_t which was amalgamated with a conditionally heteroskedastic structure of type (2.1) with an upper matrix bound of $R=2$. The following parameter estimates were obtained by straightforward optimization of the likelihood.

$$\begin{aligned} y_t &= .00014 + .185\, y_{t-1} + \varepsilon_t \\ h_t &= .000019 + .490\varepsilon_{t-1}^2 + .126\varepsilon_{t-1}\varepsilon_{t-2} + .426\varepsilon_{t-2}^2 \end{aligned} \qquad (4.1)$$

According to the estimated Hessian matrix, all 6 parameters are significant at 1 % and, though the standard ARCH structure dominates, .126 represents a noteworthy off-diagonal element.

The next model estimated was a second-order autoregression with a (2.1) structure with $R=3$ superimposed. The same algorithm needed approximately 10 hours on a PC-486 to converge. The estimated structure was

$$\begin{aligned} y_t &= .00016 + .177\, y_{t-1} - .026\, y_{t-2} + \varepsilon_t \\ h_t &= .000019 + .218\varepsilon_{t-1}^2 + .214\varepsilon_{t-1}\varepsilon_{t-2} + .362\varepsilon_{t-2}^2 + \\ &\quad + .496\varepsilon_{t-1}\varepsilon_{t-3} - .056\, \varepsilon_{t-2}\varepsilon_{t-3} + .354\, \varepsilon_{t-3}^2 \end{aligned} \qquad (4.2)$$

These coefficients are all significant. The third entry of the diagonal matrix in the re-parameterized form, however, turned out to be insignificant. Hence, the above model contains 9 parameters. The restriction $d_3=0$ was corroborated by restricted re-estimation.

For $R=4$ and $R=5$, convergence could only be achieved after imposing zero restrictions. However, $R=5$ is an interesting specification as it accommodates for day-of-the-week effects which are particularly notable from the autocorrelation function of squared returns. After some trial and error, the following model was found to have satisfactory properties with respect to numerical convergence. This restricted model explains current volatility by one linear combination of previous errors and some previous squared errors. It can be interpreted as showing five "factors": the linear combination which does *not* correspond to the linear part of the model; and four distinct Engle-ARCH-type lags.

$$\begin{aligned} y_t &= \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} + \varphi_4 y_{t-4} + \varphi_5 y_{t-5} + \varepsilon_t \\ h_t &= a_0 + d_1(\varepsilon_{t-1} + l_{21}\varepsilon_{t-2} + l_{31}\varepsilon_{t-3} + l_{41}\varepsilon_{t-4} + l_{51}\varepsilon_{t-5})^2 + \\ &\quad + d_2\varepsilon_{t-2}^2 + d_3\varepsilon_{t-3}^2 + d_4\varepsilon_{t-4}^2 + d_5\varepsilon_{t-5}^2 \end{aligned} \qquad (4.3)$$

Parameter estimates for (4.3) are given in Table 1 (penultimate column). For unrestricted models, estimation yielded unsatisfactory results. The iteration process became increasingly lengthy and failed to converge or converged to unstable solutions with many insignificant coefficients. A reason for this behavior could be found in identification problems. If d_i is insignificant, all l_{ji} are no more identified and should be set at zero before conducting estimation.

In summary, neither the restriction suggested by the Engle-ARCH model nor the other extreme (2.4) were supported. Off-diagonal elements were significant and their size was considerable. The decay versus the south-west (or north-east) corner of **B** is much slower than would be prescribed by (2.4).

To check on the stability of (4.3) with respect to time-heterogeneity, (4.3) was re-estimated for subsamples. The selection of subsamples was inspired by Hauser and Kunst (1993) who worked with the same data: a first subsample covers the early years until 1968 which year, however, was excluded because of its irregularities; a second subsample lasts from 1969 through 1978, i.e. the following decade; a third subsample starts in 1979 and ends in 1986, i.e. the year before the "Black Monday" crash. Table 1 gives estimates for these subsamples and for the whole sample according to model (4.3). Also the results from estimating the same models under the restriction $l_{21}=l_{31}=l_{41}=l_{51}=0$ are provided, i.e. from estimating a pure Engle-type ARCH(5) model.

Assuming all regularity conditions to hold, the likelihood ratios for testing the restricted Engle-ARCH model against the more general form (4.3) is χ^2-distributed with 4 degrees of freedom. The restriction is rejected for the whole sample and for the early years but not for the other two subsamples. This means that rejection of the Engle-ARCH model is primarily rooted in the early years and in the crash year 1987. Estimation of subsamples around 1987 enhanced this conclusion (but these results are probably not very interesting on their own and therefore not reported). For the years 1969-1978, unrestricted estimation resulted in an unsatisfactory model which did not meet stationarity conditions. Except for the general structure identified from the time range 1979-86 which, however, violated covariance-stationarity boundaries only slightly, all other estimated models are stationary.

Although many coefficient estimates differ quite a lot among subsamples, some features were remarkably stable. Firstly, in the lag pattern of the ARCH coefficients, d_1 typically dominates (exception 1979-1986) and d_4 is the least significant in many specifications. Secondly, $l_{21}<0$ in all cases while $\varphi_1>0$, implying that the time series factor in (4.3) is not the same as the series itself and hence the data cannot be described by a structure such as (2.4). In summary, structure is weak in the Standard & Poor 500 Index - this is not surprising and not very new either - but the little structure that was found can be explained by neither of the two simplified models, i.e. the Engle-ARCH and the (2.4) model, though the Engle-ARCH may be the less detrimental simplification.

Let me return shortly to the fact that the stationarity conditions were not met in two cases. In the next section, the feature of strict stationarity - as opposed to covariance stationarity - is treated and it is shown that it typically requires less stringent conditions. Hence, the model estimated for 1979-1986 is probably still strictly stationary (while the estimated model for 1969-1978 would probably be too far away from the stationarity boundary to meet even more liberal conditions). Simulations of these structures, however, reveal that the behavior implied by these variance-free strictly stationary processes is not reflected in the Standard & Poor 500 Index. Therefore non-stationary structures - in the sense of covariance stationarity - should be seen as implausible models and as indicating a flaw in the specification of (4.3).

5. Strict stationarity in ARCH generalizations

Whereas conditions for covariance stationarity typically come in the shape of eigenvalue restrictions such as Assumption 3a - e.g. compare Theorem 2 of Engle (1982) or Proposition 1 of Bera et al. (1992) - conditions for strict stationarity are rooted in evaluations of the spectral radius and critically depend on distributional assumptions. Except for very simple cases - see Nelson (1990) - such evaluations do not lead to closed-form conditions. In practice, all results for higher-order models, including the condition on top Lyapunov coefficients by Bougerol and Picard (1992) are accessible through numerical simulation only.

In this section we report results from Monte Carlo simulations of the very simple non-Engle ARCH model

$$\begin{aligned} y_t &= \varphi y_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim N(0, 1 + a y_{t-1}^2) \end{aligned} \tag{5.1}$$

For $\varphi=0$, this model is Engle-ARCH and the boundary of the strict stationarity region is given by Nelson's Theorem 6 as $a=3.562...$ This means that the ARCH process

$$\varepsilon_t \sim N(0, 1 + 3.56\varepsilon_{t-1}^2)$$

is strictly stationary even though its marginal distribution certainly does not possess any useful moments. This known boundary was then used to gauge our simulation experiment. Out of 1000 replications of this borderline process of length 1000, 80 replications transgressed the value of 10^{10} in absolute value. This was taken as a criterion for process stability although, of course, such a criterion necessarily admits errors of two types. Firstly, the theoretical process can be stable whereas the high but finite value is transgressed. Secondly, divergence of the theoretical process may be felt only after more than 1000 observations.

The resulting boundary curve is shown in Figure 2. Note that stationarity of the entire system is critically affected by the ARCH part in such a way that stable processes can evolve from apparently explosive autoregressions. A related process - the "trend-stationary random walk" - has been reported by Sampson (1990). Trajectories of processes in this upper lobe of the stationarity area show a curious behavior, phases of rapid expansion being accompanied by an increase in volatility which eventually leads to an abrupt end of the expansionary phase by hitting upon some smaller number by chance, whereupon the expansion starts anew. We feel, however, that these processes are not of genuine empirical significance.

The graph is symmetric around the abscissa. Note that, for $\varphi \uparrow 1$, the a boundary shifts left only slowly and $a=2$ describes a perfectly admissible model for a random walk. The estimates reported in the last section, however, point to the central region of the graph where the boundary is almost vertical. In this region, the simulated model is close to the Engle-ARCH model. The Engle-ARCH model has **B** diagonal in our notation whereas the simulated AR-ARCH has Toeplitz **B** with geometrically decaying off-diagonals. The estimates in Section 3, however, point to only slow decay in **B**. What can Figure 2 tell about this situation?

The case $\varphi=1$ provides a reasonable upper limit to the slowness of decay in fourth-order characteristics. Figure 2 shows that the accumulated process $(1-B)^{-1}y$ is stationary if y is white noise with such fourth-order characteristics. It follows that y itself is also stationary for the empirically obtained ARCH parameters. We do not explicitly claim that the extreme case of non-decaying fourth moments is supported by the data but the implications of such a feature would certainly be interesting. Such a model would imply that the index under investigation is stationary (or trend-stationary) and that this property is warranted by its extreme reaction to volatility shocks.

The strict stationarity properties of the general model (2.1) with an upper bound of $R=2$ was also investigated by simulation experiments. The two most interesting results from these experiments appears to be that, firstly, with increasing ARCH lag order, the area between the strictly-stationary and the covariance-stationary boundaries shrinks, and, secondly, this area grows if off-diagonal a_{ij} are present as compared with the classical Engle-ARCH model.

6. A direct look at fourth-moments structures

The discrimination problem between traditional Engle-ARCH models and other models in our more general ARCH class can be supported by some evaluation of descriptive measures such as empirical fourth-order cross moments. The principal distinction between Engle-type ARCH models and more general ARCH-type models is that correctly specified Engle-ARCH models only allow for correlation among squared errors whereas e.g. Weiss-ARCH models assume non-zero moments of the form

$$q(i,j)=E(\varepsilon_t^2\varepsilon_{t-i}\varepsilon_{t-j}),\ i,j>0 \tag{5.1}$$

One problem with an empirical evaluation of their sample counterparts

$$T^{-1}\sum_{t=\max(i+1,j+1)}^{T}\hat{\varepsilon}_t^2\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-j} \tag{5.2}$$

is that their distribution depends on the ARCH parameters, even if the process is white noise and even if the conditional distribution is correctly specified as normal. Treatment of these empirical moments builds on the higher-order moments results by Engle (1982, Theorem 1 and its constructive proof). In detail, assuming a first-order ARCH process

$$E_{t-1}\varepsilon_t^2=a_0+a_1\varepsilon_{t-1}^2 \tag{5.3}$$

Engle (p.1004) states that for

$$w_t'=(y_t^{2r},y_t^{2(r-1)},\ldots,y_t^2) \tag{5.4}$$

one has

$$E_{t-k}(w_t) = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \ldots + \mathbf{A}^{k-1})b + \mathbf{A}^k w_{t-k} \tag{5.5}$$

with

$$\mathbf{A} = \begin{bmatrix} p_n a_1^n & p_n \binom{n}{2} a_0 a_1^{n-1} & \cdots & \cdots & p_n \binom{n}{n-1} a_0^{n-1} a_1 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 15a_1^3 & 45a_0 a_1^2 & 45a_0^2 a_1 \\ 0 & 0 & 0 & 3a_1^2 & 6a_0 a_1 \\ 0 & 0 & 0 & 0 & a_1 \end{bmatrix}$$

$$b = \begin{bmatrix} p_n a_0^n \\ \vdots \\ 3a_0^2 \\ a_0 \end{bmatrix} \qquad p_n = \prod_{j=1}^{n} (2j-1) \tag{5.6}$$

These formulae are always valid but the limit only makes sense if unconditional moments exist, which amounts to a rather stringent condition on a_1. Second-order moments exist as long as $a_1 < 1$; for fourth-order moments, this boundary reduces to $3^{-1/2}$, for sixth order to $15^{-1/3}$ and for eighth order (which we shall need further onward) to $105^{-1/4}$ (≈ 0.312). Now, for the moment assuming all moment conditions to be satisfied, the "unconditional" limit amounts to

$$E(w_t) = (\mathbf{I} - \mathbf{A})^{-1} b \tag{5.7}$$

First, we focus on the $q(i,j)$ cross-moments with $i=1$. For conditionally normal distributions, their expectation is clearly 0. Their variance then equals their uncentered second-order moments, hence

$$\begin{aligned} E\varepsilon_t^4 \varepsilon_{t-1}^2 \varepsilon_{t-j}^2 &= E(E_{t-1}\varepsilon_t^4)\varepsilon_{t-1}^2 \varepsilon_{t-j}^2 = \\ &= 3E(a_0 + a_1 \varepsilon_{t-1}^2)^2 \varepsilon_{t-1}^2 \varepsilon_{t-j}^2 = \\ &= 3a_0^2 E\varepsilon_{t-1}^2 \varepsilon_{t-j}^2 + 6a_0 a_1 E\varepsilon_{t-1}^4 \varepsilon_{t-j}^2 + 3a_1^2 E\varepsilon_{t-1}^6 \varepsilon_{t-j}^2 = \\ &= 3a_0^2 E(E_{t-j}\varepsilon_{t-1}^2)\varepsilon_{t-j}^2 + 6a_0 a_1 E(E_{t-j}\varepsilon_{t-1}^4)\varepsilon_{t-j}^2 + \\ &\quad + 3a_1^2 E(E_{t-j}\varepsilon_{t-1}^6)\varepsilon_{t-j}^2 \end{aligned} \tag{5.8}$$

Plugging in first conditional and then unconditional moments from (5.5) and (5.7), the variance of $q(1,j)$ can be calculated. The same track can be used for $q(i,j)$ with $i \neq 1$. Note that the solution in principle requires moments up the order of 12, however, due to the triangular structure of **A**, these are not needed for the evaluation of those moments appearing in (5.8). The variance of the coefficient $q(1,1)$, which is non-zero in population, can be calculated directly.

Once the variance of the theoretical $q(i,j)$ is available, significance bounds for their sample counterparts can be calculated based on their square roots. The martingale central limit theorem by Billingsley (1961) warrants asymptotic convergence of the sample moments to normality. A comparison between the theoretical fractiles evolving from calculations and Monte Carlo fractiles (1000 replications) is given in Table 2. It is seen that the asymptotic approximation is rather accurate and can serve as a reliable indicator for practical purposes.

What happens if the lag order of the ARCH process exceeds 1? Certainly the outlined strategy still works but the involved formulae quickly become unwieldy. One may hope that the sum of coefficients gives some indication about the strength of heteroskedasticity and that higher-order ARCH processes behave "similarly" to first-order processes with a_1 equal to this sum. This conjecture can be corroborated via simulations. Two experiments in this direction are worth reporting.

Firstly, I simulated a fifth-order Engle-type ARCH model with $a_1=a_2=a_3=a_5=0.1$ and $a_4=0$ which provides a reasonable approximation to the volatility of some daily financial series (not too much unlike some ARCH structures found for time segments of the S&P returns) and entails a coefficient sum of 0.4. Empirical fractiles for trajectories of 10000 observations from this process were compared with those from a *first*-order ARCH with $a_1=0.4$. For $j=1$ and for $j>3$, the correspondence turned out to be extremely satisfactory. For $j=2$ and $j=3$, however, the ARCH(1) model produced much wider confidence intervals. These wider confidence intervals are typical for ARCH models with infinite eighth moments which may provide a possible explanation as higher-moment condition are less binding for higher-order ARCH processes with similar coefficient sums. Hence, the generated ARCH(5) model probably has finite eighth moment, unlike the ARCH(1) process with $a_1=0.4$. It is, however, difficult to corroborate this presumption analytically.

What happens if the eighth-order condition is violated? As long as the condition holds, empirical moments should converge to a normal distribution with mean zero and variance $T^{-1/2}$ times the variance of $q(i,j)$. If it fails, probably convergence at a slower rate still obtains but limit laws typically do not have finite moments of order two or even smaller order. The Monte Carlo simulations reported here shed some light on the behavior of finite-sample distributions of $q(i,j)$ in these cases which, maybe unfortunately, are of practical importance.

The second experiment concerns pseudo-data generated on the basis of the last column of Table 1, i.e. the estimated coefficients for a fifth-order AR-ARCH model for the S&P returns series. The simulated fractiles appear as dashed lines in Figure 3. Due to the complicated coefficients lag pattern, the $q(1,j)$ do not decrease monotonously with increasing j. However, percentiles seem to be symmetrical around the symmetry axis of $j=1$, i.e. the only cross-moment which is not zero for the Engle-ARCH model.

The significance bounds obtained from this simulation experiment were then contrasted with the estimated values for $q(1,j)$ from the S&P returns themselves. If the Engle-ARCH model were true, all of these values would have to be zero in population and insignificant in sample. Figure 3 shows the more interesting estimates for $j > 1$ ($q(1,1)$ typically distorts the picture). Only 5 out of 20 empirical moments fall into the 90% significance bounds. Violation of bounds, however, is not so conspicuous that it would render the classical ARCH model useless. Violation of bounds is far more noticeable for $j < 0$ which, however, is not the subject of this paper. The apparition of significantly non-zero $q(1,j)$ with $j < 0$ could, of course, also be due to a misspecification of the structure at positive j but it appears to be too strong to be explained away that easily and is a potential topic for future research. These moments may be closer related to ARCH-in-mean effects.

A caveat of the analysis is that the pseudo-data have been generated without specifically taking into account that they are a residual series. The pattern shown in Figure 3, however, is robust against any endeavors of respecifying the residual series and was narrowly replicated by using residuals from simple linear AR models etc. Moreover, the effect of simulating a residual series appears to make the test procedure shown even more conservative. Anyway, the frequent violation of confidence bounds in Figure 3 cannot be explained away along these lines.

Experiments with $q(i,j)$ for $i=2,3,...$ have also been conducted and, in summary, corroborate the findings from $i=1$. Simulated significance bounds are typically violated and the ARCH specification must be rejected on these grounds but the evidence against the ARCH null is not overwhelming, which corresponds well to the likelihood-ratio test for the whole sample provided by Table 1.

7. Summary and conclusion

Both descriptive moments statistics and parametric models have indicated more or less convincingly that fourth-moments structures in financial series may be more complicated than would be prescribed by the traditional ARCH model. Within the limits of this paper, I restricted attention to cross-moments structures of the form $E(\varepsilon_t^2 \varepsilon_{t-i} \varepsilon_{t-j})$, i.e. to the explanation of volatility by preceding cross-terms. Two points have been neglected intentionally which may deserve further investigation.

Firstly, the analysis is strictly limited to Gaussian assumptions. It is well known that innovations in financial time series are typically not conditionally Gaussian but slightly conditionally leptokurtic (compare Baillie and Bollerslev, 1989). Non-normality could affect some of the parametric model results, probably the validity of the chi-square approximation to the likelihood-ratio test, and most certainly the percentiles of the moments estimates.

Secondly, higher-order moments of different form were neglected. Engle's ARCH model can be seen as the first important attempt to parameterize fourth-order cross structures such as $E(\varepsilon_t^2 \varepsilon_{t-i}^2)$. The GARCH model by Bollerslev (1986) does the same but uses the rational function approximation in place of the previously used polynomial approximation. The later-developed models by Weiss (1984) and Bera et al. (1992) aim at modeling mixed moments such as $E(\varepsilon_t^2 \varepsilon_{t-i} \varepsilon_{t-j})$. Other generalizations such as "ARCH in mean" are concerned with cross-moments of order three and this track is very much at the center of research at the moment (compare Engle and Lee, 1993). There is an ample field for parameterizations of all kinds of higher-order cross moments and, maybe even more important, empirical findings of these higher-order structures may not be independent of one another. In other words, allowing for non-zero third-order moments may change some of the properties of the models treated in this paper, and these effects have also been neglected.

Finally, it should be pointed out that these two caveats are interrelated in the sense that additional structure can be searched for by deviating from the assumption of Gaussianity or by sticking to that assumption and parameterizing higher-moments structures. The two paths are probably alternatives and the second one has recently proved to be more fruitful.

Acknowledgments

The author wishes to thank Michael Hauser for extensive help with all computer work.

Literature

Bera, A.K., M.L. Higgins, S. Lee (1992): "Interaction Between Autocorrelation and Conditional Heteroskedasticity: A Random Coefficient Approach," Journal of Business and Economic Statistics 10, 133-142.

Baillie, R.T. and T. Bollerslev (1989): "The Message in Daily Exchange Rates. A Conditional-Variance Tale," Journal of Business and Economic Statistics 7, 297-305.

Billingsley, P. (1961): "Statistical Inference for Markov Processes" Holt, New York.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," Journal of Econometrics 31, 307-327.

Bougerol, P. and N. Picard (1992) "Stationarity of GARCH processes and of some nonnegative time series," Journal of Econometrics 52, 115-127.

Calzolari, G. and G. Fiorentini (1992): "Alternative Methods for GARCH Estimation," Working paper, University of Florence, presented at the ESEM Meeting, Brussels.

Engle, R.F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation," Econometrica 50, 987-1007.

Engle, R.F. and G.G.J. Lee (1993): "A Permanent and Transitory Component Model of Stock Return Volatility," Paper presented at the Econometric Society European Meeting, Uppsala.

Hauser, M.A. and R.M. Kunst (1993): "Fractionally Integrated Models With ARCH Errors," Research Memorandum No. 322, Institute for Advanced Studies, Vienna, Austria.

Kunst, R.M. (1993): "Stability Conditions for a Bivariate ARCH System Which Is Cointegrated in Mean," Communications in Statistics, Theory and Methods (forthcoming).

Kunst, R.M. and W. Polasek (1993): "Structuring Volatile Swiss Interest Rates: Some Evidence on the Present Value Model and a VAR-VARCH Approach," in: J. Kaehler (ed.) Financial Markets Econometrics (forthcoming).

Nelson, D.B. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," Econometric Theory 6, 318-334.

Sampson, M. (1990): "Trend Stationary Random Walks," Paper presented at the World Conference of the Econometric Society, Barcelona.

Weiss, A.A. (1984): "ARMA models with ARCH errors," Journal of Time Series Analysis 5, 129-143.

TABLE 1: Coefficient estimates from general ARCH model (4.3) and from Engle-ARCH model for Standard & Poor 500 Index series based on subsamples

	1962-1967		1969-1978		1979-1986		1962-1990	
$\sqrt{d_1}$	.601	.627	.500	.504	.175	.175	.529	.533
$\sqrt{d_2}$	.218	.103	1.200	.211	.424	.120	.342	.087
$\sqrt{d_3}$	.183	.008	.473	.286	.731	.860	.254	.176
$\sqrt{d_4}$	.002	.224	.007	.286	.581	.027	.355	.541
$\sqrt{d_5}$	.172	.370	.617	.521	.075	.098	.246	.294
l_{21}	-.220		-.514		-.805		-.568	
l_{31}	.383		-.263		.352		.361	
l_{41}	.197		.447		-.013		-.110	
l_{51}	.218		.746		.117		.252	
$\sqrt{a_0}$	.00311	.00023	.00000	.00522	.00066	.00002	.00392	.00163
μ	.00023	.00035	.00001	.00010	.00043	.00043	.00021	.00031
φ_1	.197	.170	.295	.302	.094	.094	.191	.194
φ_2	.045	.063	-.098	-.103	.002	.002	-.062	-.060
φ_3	-.003	.012	.039	.047	-.016	-.016	-.022	-.016
φ_4	-.041	-.026	.022	.023	-.020	-.021	.027	.026
φ_5	.028	.041	-.037	-.037	.008	.007	.015	.014
Σa_{ii}	.573	.591	2.567	.734	1.112	.795	.796	.702
-2ℓ	13202	13183	21751	21746	17102	17102	61573	61559

TABLE 2: Simulated fourth-moments distribution and theoretical values. 1000 replications for each experiment.

(a): ARCH(1) with $a_1 = 0.2$

		T=1000				T=10000			
		simulated fractiles		asymptotic fractiles		simulated fractiles		asymptotic fractiles	
moment	true value	5%	95%	5%	95%	5%	95%	5%	95%
$q(1,1)$	1.455	1.061	1.948	1.080	1.830	1.315	1.599	1.336	1.573
$q(1,2)$	0.000	-0.221	0.210	-0.221	0.221	-0.065	0.066	-0.070	0.070
$q(1,3)$	0.000	-0.162	0.146	-0.162	0.162	-0.052	0.046	-0.051	0.051
$q(1,4)$	0.000	-0.137	0.128	-0.147	0.147	-0.049	0.045	-0.046	0.046
$q(1,5)$	0.000	-0.143	0.131	-0.144	0.144	-0.047	0.042	-0.045	0.045

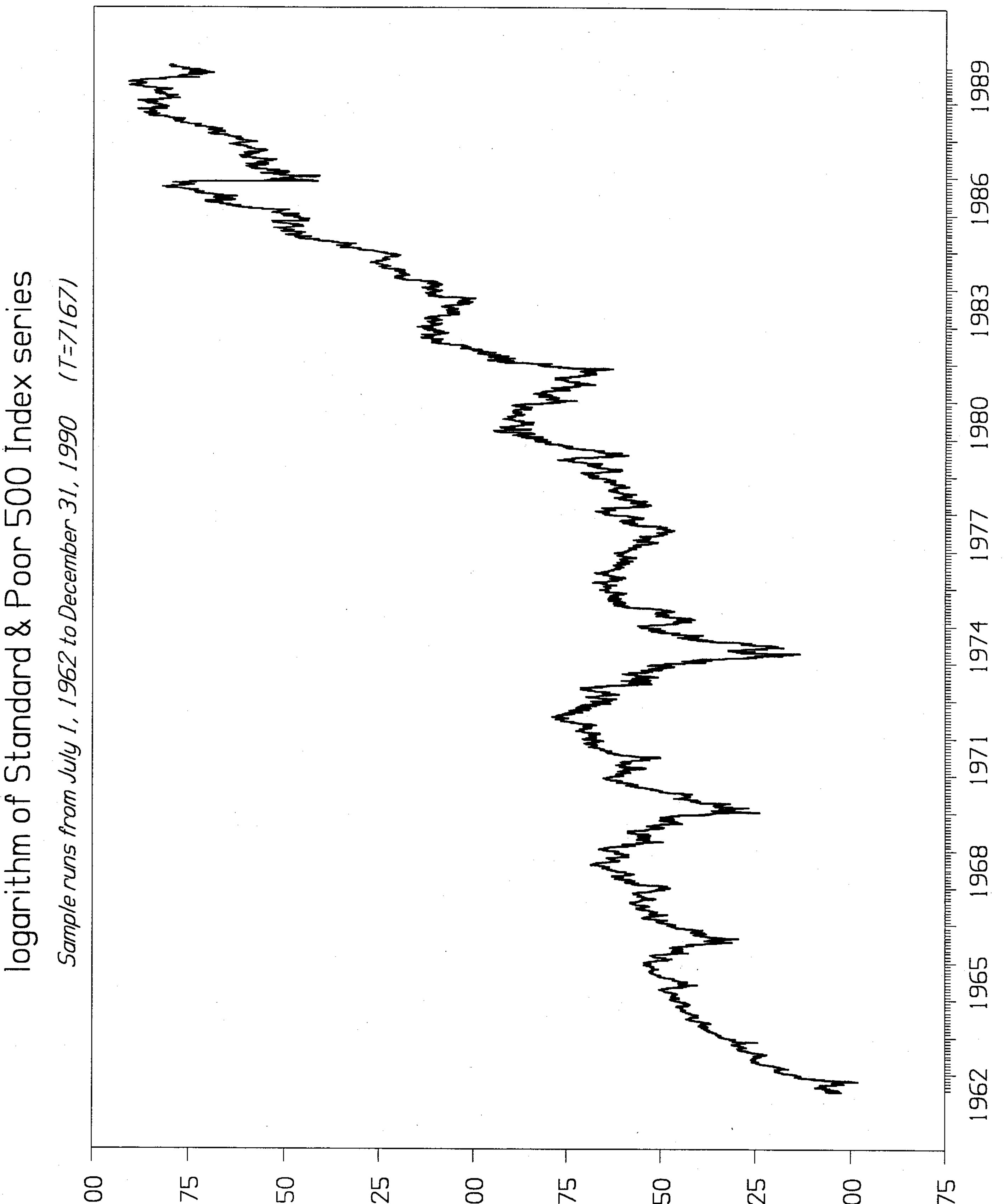
(b): independent observations

		T=1000				T=10000			
		simulated fractiles		asymptotic fractiles		simulated fractiles		asymptotic fractiles	
moment	true value	5 %	95 %	5 %	95 %	5 %	95 %	5 %	95 %
$q(1,1)$	0.000	-0.1687	0.1816	-0.1561	0.1561	-0.0514	0.0464	-0.0494	0.0494
$q(1,2)$	0.000	-0.0970	0.0942	-0.0901	0.0901	-0.0287	0.0293	-0.0285	0.0285
$q(1,3)$	0.000	-0.0903	0.0900	-0.0901	0.0901	-0.0284	0.0290	-0.0285	0.0285
$q(1,4)$	0.000	-0.0854	0.0899	-0.0901	0.0901	-0.0275	0.0270	-0.0285	0.0285
$q(1,5)$	0.000	-0.0976	0.0826	-0.0901	0.0901	-0.0294	0.0276	-0.0285	0.0285

(c): ARCH(1) with $a_1 = 0.5$ (asymptotic distribution unknown; simulated fractiles)

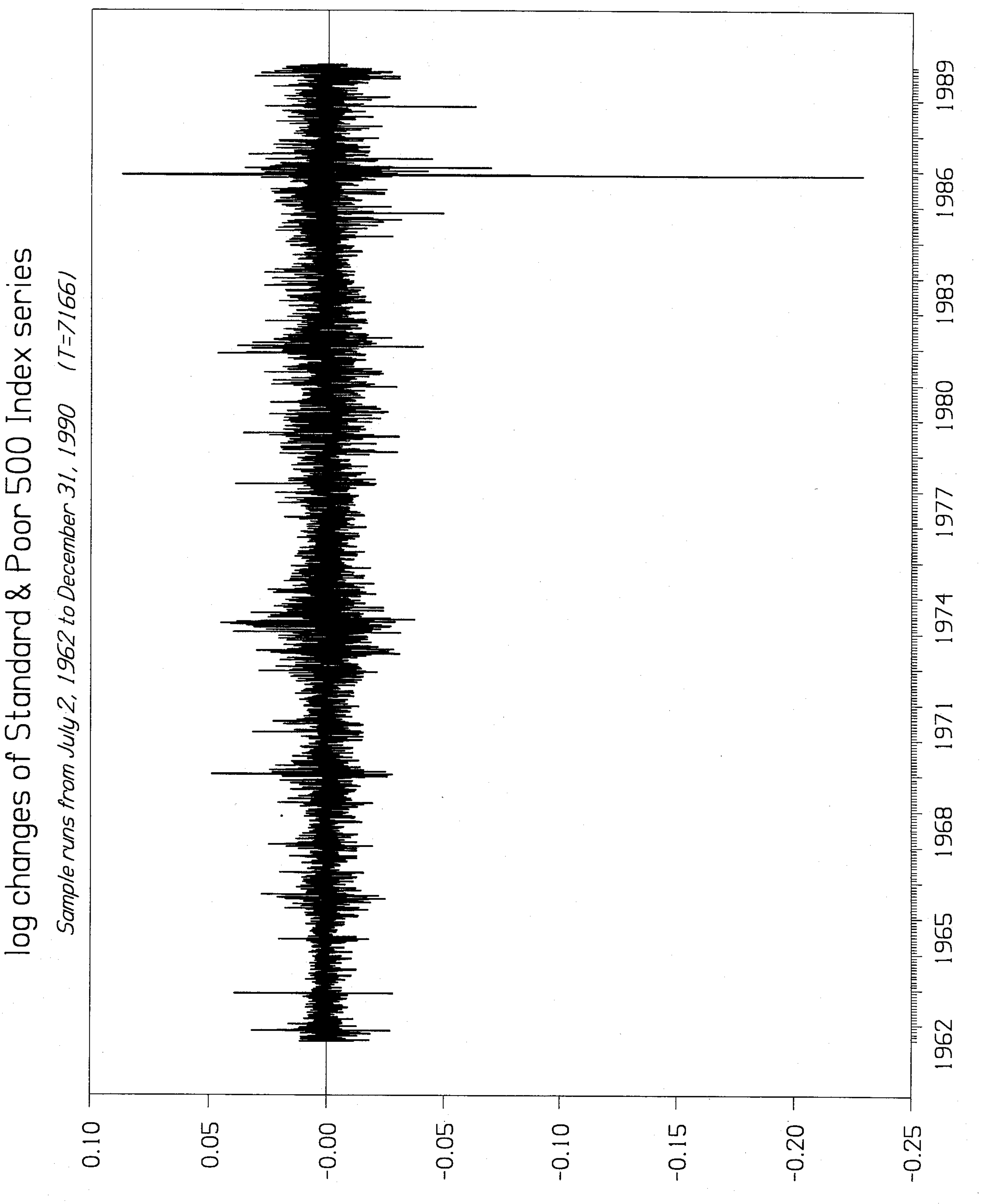
		T=1000		T=10000	
moment	true value	5 %	95 %	5 %	95 %
$q(1,1)$	5.000	1.33	9.05	2.230	8.012
$q(1,2)$	0.000	-1.34	1.34	-0.883	0.979
$q(1,3)$	0.000	-0.87	0.88	-0.723	0.541
$q(1,4)$	0.000	-0.64	0.60	-0.550	0.405
$q(1,5)$	0.000	-0.454	0.532	-0.347	0.313

FIGURE 1a



logarithm of Standard & Poor 500 Index series
Sample runs from July 1, 1962 to December 31, 1990 (T=7167)
6.00
5.75
5.50
5.25
5.00
4.75
4.50
4.25
4.00
3.75
1962
1965
1968
1971
1974
1977
1980
1983
1986
1989

FIGURE 1b



log changes of Standard & Poor 500 Index series
Sample runs from July 2, 1962 to December 31, 1990 (T=7166)
0.10
0.05
-0.00
-0.05
-0.10
-0.15
-0.20
-0.25
1962
1965
1968
1971
1974
1977
1980
1983
1986
1989

FIGURE 2.

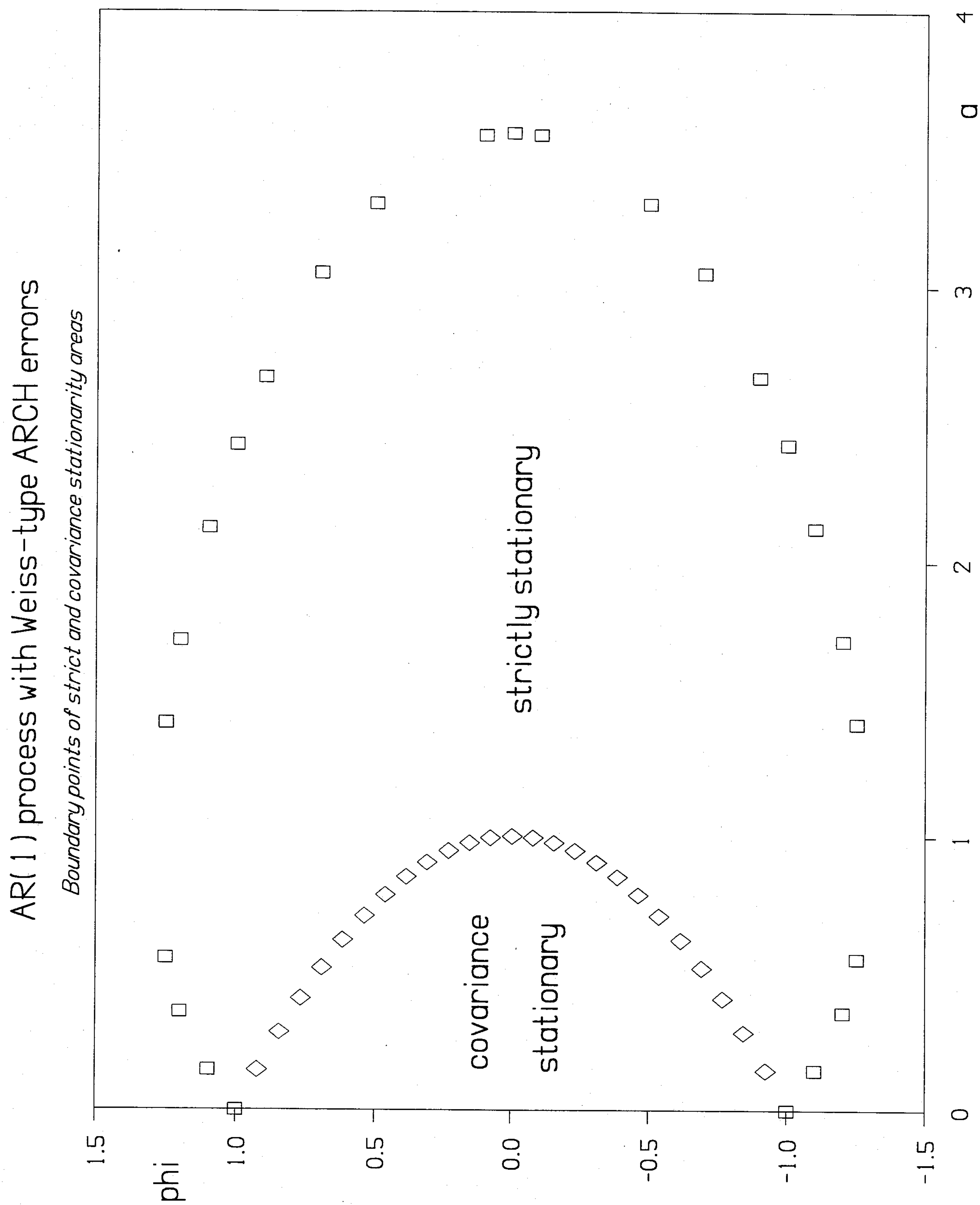
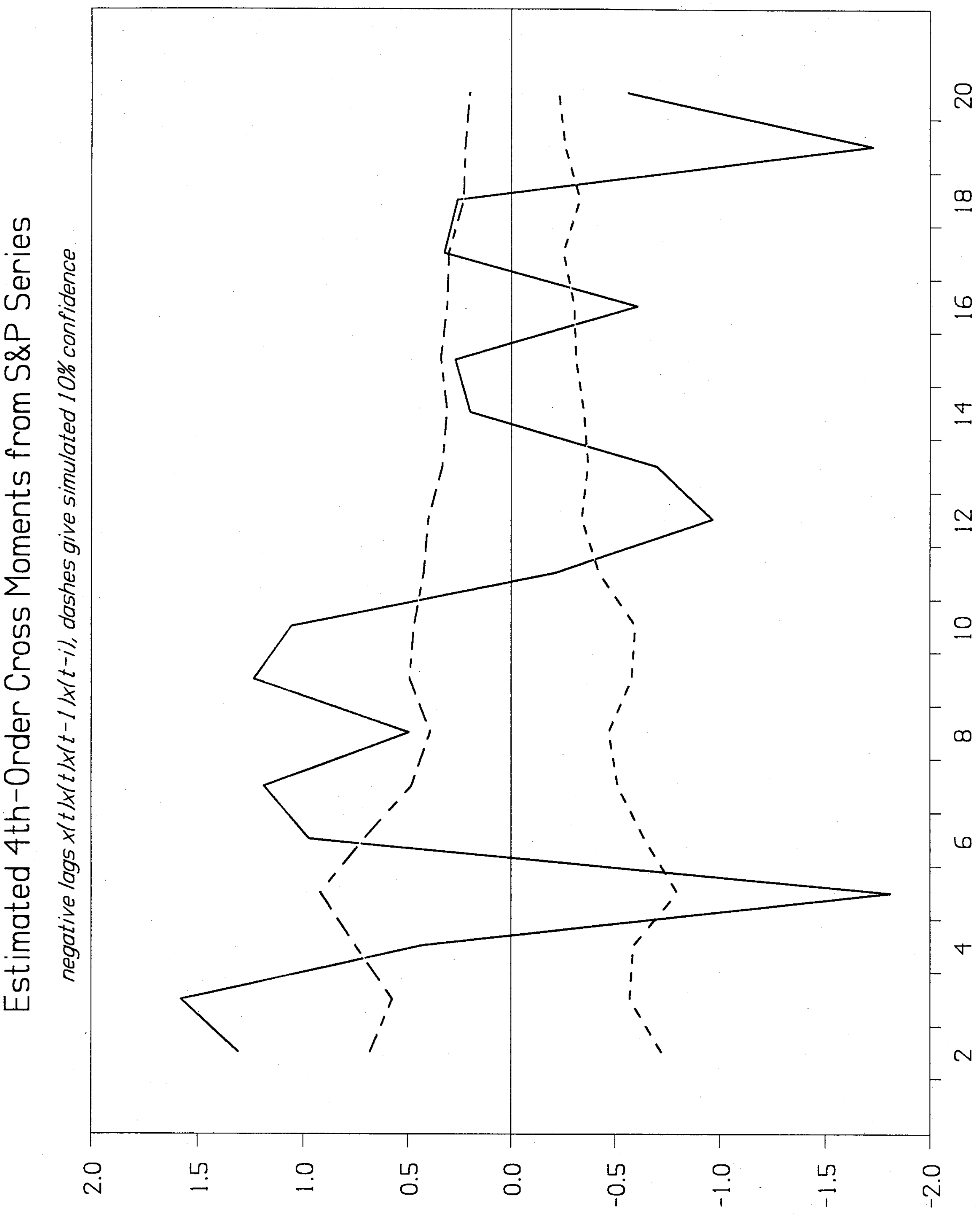

AR(1) process with Weiss-type ARCH errors
Boundary points of strict and covariance stationarity areas
phi
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
0
1
2
3
4
a
covariance
stationary
strictly stationary

FIGURE 3.



Estimated 4th-Order Cross Moments from S&P Series
negative lags x(t)x(t)x(t-1)x(t-i), dashes give simulated 10% confidence
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
2
4
6
8
10
12
14
16
18
20