

Augmented ARCH Models for Financial Time Series: Stability conditions and empirical evidence

Robert M. Kunst

Johannes Kepler University
A-4040 Linz-Auhof, Austria
and

Institute for Advanced Studies
Stumpergasse 56, A-1060 Vienna, Austria

Applied Financial Economics, 1997, **7**, 575–586

Abstract

The class of conditionally heteroskedastic models known as ‘augmented ARCH’ encompasses most linear ‘ARCH’-type models found in the literature and, in particular, two basic ARCH variants for auto-correlated series: Engle (1982) explains conditional variance by lagged errors, Weiss (1984) also by lagged observations. The framework permits an evaluation of whether the restrictions evolving from the Engle or the Weiss models are valid in practice. Time series of stock market indexes for some major stock exchanges yield empirical examples. In most cases, the statistical approximation to actual dynamic behavior is improved substantially by considering augmented ARCH structures.

I INTRODUCTION

For some time now, scientific interest in serially correlated volatility has been soaring. This interest is concentrating primarily on financial time series where prediction of means is notoriously unrewarding and hence structure, if any, is to be found through higher-moments properties only. For many financial price variables, including common stocks

and stock market indicators, the theory of efficient markets implies that (logarithms of) time series follow random walks. Hence, their first differences are unpredictable while forecasts on the series itself are provided by the latest observation plus a possible ‘drift constant’. Empirical evidence typically tends to reject the theory, although the amount of serial correlation in returns is too small and too unstable in time to permit systematic gains by risk-averse investors. A closer look at such series reveals, however, that they show noteworthy temporal clusters of volatility.

The best known statistical model for the volatility-clustering phenomenon is the ARCH model (see Engle, 1982). The acronym denotes ‘autoregressive conditional heteroskedasticity’, stressing 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. The basic 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, were tackled in a slightly different way by Weiss (1984). Tsay (1987) and Bera *et al.* (1992) viewed the problem in the framework of random coefficient models.

Just as the Box-Jenkins ARMA models are built around the archetypal case with Gaussian innovations and this assumption can be substantially relaxed but statistical properties be retained, the typically counterfactual assumption of conditional normality in ARCH models (see Engle, 1982) can be relaxed and ARCH modeling then becomes a powerful descriptive device for parameterizing the fourth-moments structure of time series. This quasi-maximum likelihood or weak ARCH interpretation was previously proposed, explicitly or implicitly, by Weiss (1984), Tsay (1987), and others. In contrast, Engle (1982) and Bollerslev (1986) apparently view the statistical models as approximations to the exact distribution of the data-generating process. However, the latter view implies that one either must restrict attention to the empirically rare Gaussian case or one must develop complicated adaptive estimation techniques. Here, ARCH modeling will be seen as a direct extension of linear ARMA modeling and the first interpretation is adopted.

Even if one restricts attention to cross-moments among squares, such as $E(X_t^2 X_{t-j}^2)$, modeling conveniently goes along with some linear time series specification as a first step to capture the second-order moments of type $E(X_t X_{t-j})$. ARCH models can then be fitted to the linear innovations. However, it is not clear whether to focus on the moments’ structure of the process proper or of its innovations. Engle’s (1982) seminal paper appears to have imposed a slight bias

towards the latter route whereas, for example, Tsay (1987) adopts the former view. In the parametric models used here, modeling is based primarily on innovations moments. However, the influence of higher moments among observed variables and their squares is accounted for, which brings in mixed moments such as $E(X_t^2 X_{t-i} X_{t-j})$.

This paper is organized as follows. Section II reviews the class of augmented ARCH models which encompasses most linear ARCH models known from the literature. Conditions for covariance stationarity are given. Section III presents some simulation results related to strict stationarity of the model class. Section IV is concerned with issues related to the problem of maximum-likelihood estimation of the parameters of the structures. Analytical derivatives are given in the appendix. Section V reports the results from fitting the more general ARCH model class to the Standard & Poor 500 Index series and to other selected stock index time series. The findings indicate that some real-life ARCH structures are richer than allowed by the restrictive classical ARCH models. Section VI concludes.

II THE MODEL CLASS CONSIDERED

We conceive the data y_t as realizations of a stochastic process. We assume we can approximate linear behavior reasonably well by a time series model with innovations ε_t . This includes white noise $y_t = \varepsilon_t$ and the random walk $y_t = y_{t-1} + \varepsilon_t$ as special cases.

The model class considered for ε_t has the following form

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

with possibly $R = \infty$. I_t denotes an information set containing all ε_s for $s \leq t$. In contrast to linear time series analysis, this I_t is non-linear in the sense that it contains information on ε_s as well as on the cross-products $\varepsilon_s \varepsilon_{s-i}$. Expectation is still linear with respect to these cross-products. This model class corresponds closely to the AARCH and GAARCH models of Bera *et al.* (1992) who derived them from a random coefficient structure, and to the CHARMA models by Tsay (1987) who adopted the converse view to (2.1) and let conditional heteroskedasticity depend exclusively on observed variables. Consequently, the stationarity conditions for CHARMA models (and also for GAARCH) are similar to those given in the following.

In order that Equation (2.1) should make sense, it must obey certain restrictions. In particular, the following assumptions, 1 and 2, ensure that h_t is a well-defined conditional variance process with $h_t > 0$

with probability one. Assumption 3 is needed to establish the existence of a stationary solution to Equation 2.1. It could be replaced by requiring simply that Σa_{ii} converge to a limit smaller than one. Assumptions 4 and 5 describe the stochastic process ε_t and may potentially be replaced by weaker conditions.

Assumption 1. $a_0 > 0$.

Assumption 2. The array \mathbf{B} , for convenience, formed from the coefficients a_{ij} in such a way that

$$\begin{aligned} b_{ii} &= a_{ii} \\ b_{ij} &= a_{ij}/2 \text{ for } i > j \\ b_{ij} &= b_{ji} \text{ for } i < j \end{aligned} \tag{2.2}$$

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

Assumption 3a. The roots of the polynomials in the sequence of characteristic polynomials formed by the diagonal elements of \mathbf{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.

Assumption 4. ε_t has an MDS (martingale difference series) property in the sense that

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

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 conditions, I will assume for the moment that Assumption 5 holds in order to facilitate the presentation.

Proposition 1. Under Assumptions 1-5, Equation 2.1 has a covariance-stationary solution

A sketch of the proof of Proposition 1 is given in the appendix. Note in particular that stationarity is unaffected by off-diagonal elements in \mathbf{B} . Off-diagonals in \mathbf{B} are, however, severely restricted by non-negative definiteness and hence the diagonal of \mathbf{B} typically dominates the array. Note that a covariance-stationary solution to Equation 2.1 does not necessarily imply a covariance-stationary solution for ε_t^2 , as fourth-order moments may not be finite. For ARCH processes, conditions for the existence of higher-order moments were developed, Milhøj (1985). Kunst (1997) provides conditions for finite fourth moments in augmented ARCH processes. Violation of these conditions

is encountered frequently in empirical applications and may adversely affect the properties of estimators.

The model class (Equation 2.1) comprises most linear heteroscedastic models presented in the literature. The original ARCH model by Engle (1982) has only finitely many a_{ii} different from 0 and $a_{ij} = 0$ for $i \neq j$. Definiteness holds if $a_{ii} \geq 0$ as \mathbf{B} is a diagonal matrix. 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. 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. In spite of its widespread popularity in empirical finance, the model has some widely ignored drawbacks. For instance, the GARCH model is unable to capture MA-type second-order structures with $\text{corr}(\varepsilon_t^2, \varepsilon_{t-i}^2) = 0, i > i_0$. Product processes as suggested by Taylor (1986) may be the more natural model class to consider.

In relation to serially correlated time series, a more interesting special case of Equation 2.1 is the univariate version of the time series ARCH model suggested by Kunst and Polasek (1994), who assume that ε_t is the innovations process from a linear time series model for the data y_t and that h_t depends on lagged squared mean-adjusted observations

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

If all coefficients are positive, then \mathbf{B} is again n.n.d. In the ARMA-ARCH model by Weiss (1984), h_t depends on both lagged squared observations and lagged squared innovations:

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

In Equation 2.5, definiteness is ensured by $a_i \geq 0, c_i \geq 0$. Stationarity conditions in Equations 2.4 and 2.5 depend on the properties of the linear time series process. There is a trade-off in the sense that more linear dependence can be tolerated if the coefficients a_i and c_i are sufficiently small. Theoretically, some non-negativity conditions could also be relaxed.

III STRICT STATIONARITY IN ARCH GENERALIZATIONS

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 considerably larger values. A similar behavior was found by Kunst (1993) for certain bivariate ARCH structures. In those cases, unconditional variances do not exist. Bougerol and Picard (1992a) provide an extension to higher-order GARCH models.

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 depend on distributional assumptions. Only in very simple cases—see Nelson (1990)—do such evaluations admit closed-form conditions. Results for higher-order models, including the condition on top Lyapunov coefficients by Bougerol and Picard (1992a,b), are accessible through numerical simulation only.

Some ARCH processes can be viewed as random coefficient autoregressions (RCA, see Nicholls and Quinn, 1982) and the stability properties of those processes can be applied. For example, consider the simple non-Engle ARCH model of type (2.4)

$$\begin{aligned} y_t &= \varphi y_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim N(0, 1 + ay_{t-1}^2) \end{aligned} \quad (3.1)$$

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

$$\varepsilon_t \sim N(0, 1 + 3.56y_{t-1}^2) \quad (3.2)$$

is strictly stationary even though its marginal distribution does not possess any useful moments. For arbitrary φ , the process (3.1) is equivalent to the RCA process ¹

$$\begin{aligned} y_t &= \varphi_t y_{t-1} + \varepsilon_t \\ \varphi_t &\text{ i.i.d. } N(\varphi, a) \\ \varepsilon_t &\text{ i.i.d. } N(0, 1) \end{aligned} \quad (3.3)$$

¹Generalizing scales by allowing ε_t to have variance a_0 does not affect the following arguments.

Stability conditions for such RCA processes have been developed by Bougerol and Picard (1992b) based on a theorem by Brandt (1986). The ensuing conditions yield

$$E_{\varphi,a} \log |\varphi_t| < 0 \quad (3.4)$$

This boundary is replicated in Fig. 1. It was corroborated by a simulation experiment. Out of 1000 replications of the borderline processes of length 10,000, around 80 replications transgressed the value of 10^{10} in absolute value. This value was therefore taken as a criterion for process stability in other situations where theoretical conditions were intractable.²

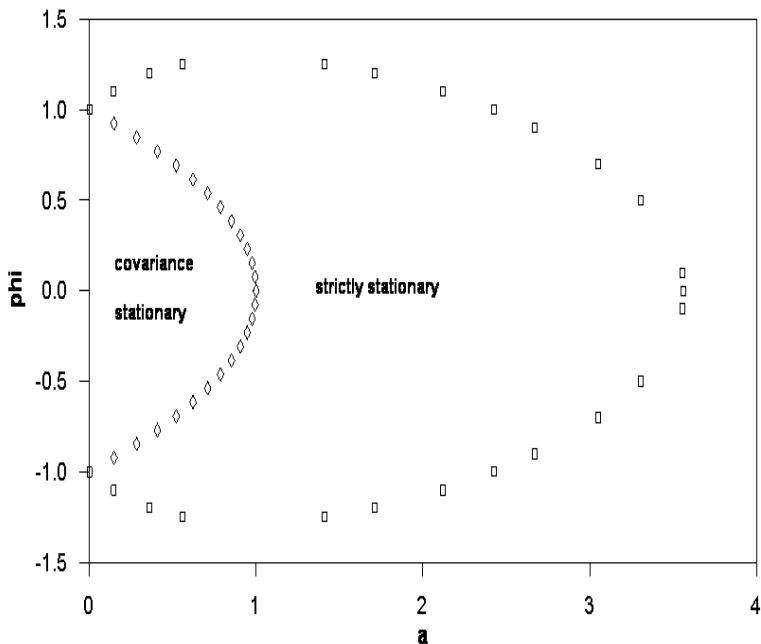


Figure 1: *Boundary points of strict and covariance stationarity areas for the model $y_t = \varphi y_{t-1} + \varepsilon_t$, $E(\varepsilon_t^2 | I_{t-1}) = 1 + a y_{t-1}^2$.*

From Fig. 1, note that stationarity of Equation 3.1 is critically affected by the ARCH part in such a way that stable processes can evolve from apparently explosive autoregressions. 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

²A different and more refined simulation technique was applied by Kleibergen and VanDijk (1993) who replicated Nelson's theoretical boundary for the GARCH(1,1) model.

by hitting upon some smaller number by chance, whereupon the expansion starts anew. However, these processes are of little empirical significance.

The graph is symmetric around the abscissa. For $\varphi \uparrow 1$, the a boundary shifts left only slowly and $a = 2$ yields an admissible model for a random walk. Most estimates reported from actual financial returns series point to the central region of the graph where the boundary is almost vertical and the simulated model is close to the Engle-ARCH model. In the notation of Equation 2.2, the Engle-ARCH model has diagonal \mathbf{B} while the simulated AR-ARCH has Toeplitz \mathbf{B} with geometrically decaying off-diagonals.

Even if—as in some examples reported in Section V—evidence suggests only slow decay in \mathbf{B} , the case $\varphi = 1$ in Equation 3.1 can be viewed as a limiting case, notwithstanding the true properties of the AR process. Figure 1 shows that the accumulated process $(1 - B)^{-1}y$ is stationary if y is white noise with such fourth-order characteristics. Hence, y itself is also stationary for the empirically obtained ARCH parameters. Though this extreme case of non-decaying fourth moments is not generally supported by the data, the implications of such a feature would certainly be interesting. Such a model would imply that the level 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, Equation 2.1, with $R = 2$ were also investigated by simulation experiments.³ The following parameterization admits a reasonable graphical representation

$$h_t = 1 + a_1 \varepsilon_{t-1}^2 + a_2 \varepsilon_{t-2}^2 + \tau \sqrt{a_1 a_2} \varepsilon_{t-1} \varepsilon_{t-2} \quad (3.5)$$

By non-negative definiteness, τ is restricted to the closed interval $[-1, 1]$. $\tau = 0$ yields the Engle-ARCH(2) model. For $\tau = 0$, $\tau = 0.5$, and $\tau = 1.0$, simulated boundary curves of the strict stationarity area are presented in Fig. 2. For $\tau < 0$, behavior is symmetric to positive τ . Note that, for $\tau = 1$, the boundary appears to be nearly

³In this case, theoretical existence and uniqueness of the stationary distribution obeying the defining equation are not warranted as such. In the covariance-stationary case outlined in section I, these follow from the time-constancy of the generating mechanism and covariance stationarity. In Engle-type ARCH models, they follow from the techniques outlined by Bougerol and Picard (1992a) for positive autoregressions. In Weiss-type models, they follow from the equivalence to random coefficients structures and from the results by Bougerol and Picard (1992b). For the general model, these properties have not been established yet. It is reasonable to assume, however, that ergodic distributions will be well-behaved here and this presumption is also supported by Monte Carlo results.

linear unless a_1 is very small. Also note that the curves for $\tau = 0$ and $\tau = 0.5$ are very close to each other, indicating that significant additional ground for the stationarity region is gained for large deviations from the classical model only. $a_1 = 0$ yields the same boundary value for a_2 as $a_2 = 0$ does for a_1 . Figure 2 shows, however, that behavior with respect to the two parameters is not at all symmetric.

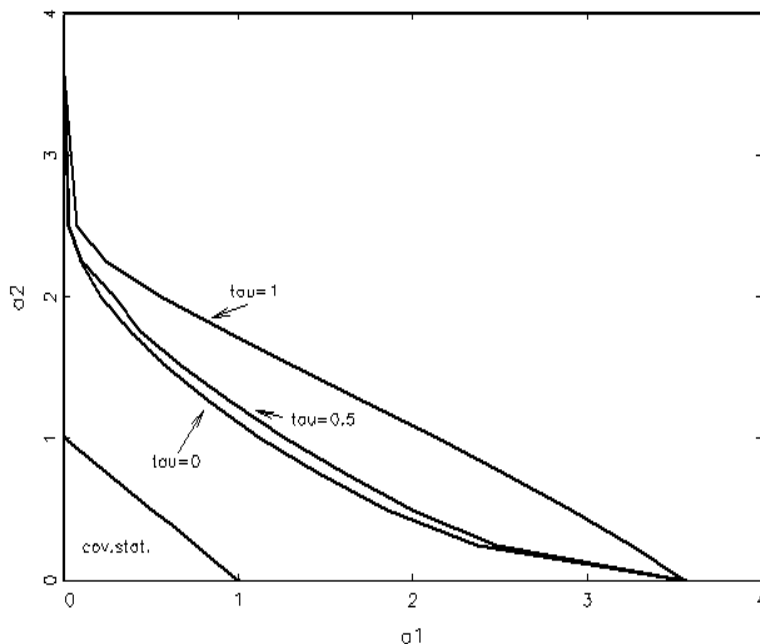


Figure 2: *Simulated boundaries of strict and covariance stationarity areas for the model $y_t = \varepsilon_t$, $E(\varepsilon_t^2 | I_{t-1}) = 1 + a_1 \varepsilon_{t-1}^2 + a_2 \varepsilon_{t-2}^2 + \tau \sqrt{a_1 a_2} \varepsilon_{t-1} \varepsilon_{t-2}$. Points to the right correspond to non-stationary models.*

It evolves from these and some further (unreported) experiments 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.

IV ESTIMATION ISSUES

Here and in the following it will be assumed that $R < \infty$. This is not the only way to define a finite parameterization for estimation purposes, and models with infinite-dimensional arrays \mathbf{B} which depend on

a finite parameter set—such as GARCH models—may be worth considering. To ease presentation, however, I assume that a finite matrix gives a reasonable approximation to the possibly infinite-dimensional \mathbf{B} .

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

$$\begin{aligned}\ell &= T^{-1} \sum_{t=1}^T \ell_t \\ -\ell_t &= \frac{1}{2} \log h_t + \frac{1}{2} \varepsilon_t^2 / h_t\end{aligned}\tag{4.1}$$

but straightforward numerical optimization of Equation 4.1 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 Equation 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 $\theta_2 = (a_0, a_{11}, a_{21}, a_{22}, \dots)'$ of Equation 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 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. Equation 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. Equation 2.1 is, however, not Engle-symmetric but Bera *et al.* (1992) showed that block-diagonality of the information matrix evolves from symmetry in a much wider sense. In short, $h_t(\varepsilon_t, \varepsilon_{t-1}, \dots)$ is Bera-symmetric if changing of all ε_s to $-\varepsilon_s$ yields the same value while $h_t(\cdot)$ is Engle-symmetric if this property holds for *any* change of individual ε_s to $-\varepsilon_s$.

Computation time can further be shortened by analytically evaluating scores, i.e. derivatives of ℓ_t with respect to parameters. These can be based on a convenient re-parameterization of Equation 2.1 that allows to replace the complicated admissibility restrictions by non-negativity constraints. If parameters are expressed through squares, estimation can proceed without further constraints. Formulae for scores and an outline of their derivation are given in the appendix.

V EMPIRICAL EXAMPLES

Standard & Poor 500

A reliable long series frequently used for evaluations of conditionally heteroscedastic models is the Standard and Poor's Index S&P 500. Daily observations during the time period from 2 July 1962 to 31 December 1990 permit a sample of 7168.⁴ In the linear time series framework, the logarithm of 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 heteroscedasticity 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 some insight on whether the more general approach suggested in Equation 2.1 is justified relative to the original ARCH model and whether, for example, restrictions as in the Weiss model hold if the standard ARCH appears insufficient.

First, a linear first-order autoregression was tried on the differenced series y_t which was amalgamated with a conditionally heteroscedastic 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

$$\begin{aligned}y_t &= 0.00014 + 0.185y_{t-1} + \varepsilon_t \\h_t &= 0.000019 + 0.490\varepsilon_{t-1}^2 + 0.126\varepsilon_{t-1}\varepsilon_{t-2} + 0.426\varepsilon_{t-2}^2\end{aligned}\quad (5.1)$$

According to the estimated Hessian matrix, all six parameters are significant at 1 % and, though the standard ARCH structure dominates, 0.126 represents a noteworthy off-diagonal element.

The next model estimated was a second-order autoregression with an Equation 2.1 structure with $R = 3$ superimposed. The estimated structure was

$$\begin{aligned}y_t &= 0.00016 + 0.177y_{t-1} - 0.026y_{t-2} + \varepsilon_t \\h_t &= 0.000019 + 0.218\varepsilon_{t-1}^2 + 0.214\varepsilon_{t-1}\varepsilon_{t-2} + 0.362\varepsilon_{t-2}^2 \\&\quad + 0.496\varepsilon_{t-1}\varepsilon_{t-3} - 0.056\varepsilon_{t-2}\varepsilon_{t-3} + 0.354\varepsilon_{t-3}^2\end{aligned}\quad (5.2)$$

⁴All subsequent analysis is related to these first log differences which are therefore simply called 'the S&P Index series'.

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 nine parameters. The restriction $d_3 = 0$ was corroborated by restricted re-estimation.

For $R = 4$ and $R = 5$, numerical convergence could only be achieved after imposing zero restrictions. $R = 5$ 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 numerical properties. 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

$$\begin{aligned}
y_t &= \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} \\
&\quad + \varphi_4 y_{t-4} + \varphi_5 y_{t-5} + \varepsilon_t \\
h_t &= a_0 + d_1 (\varepsilon_{t-1} + l_{21} \varepsilon_{t-2} + l_{31} \varepsilon_{t-3} + l_{41} \varepsilon_{t-4} + l_{51} \varepsilon_{t-5})^2 \\
&\quad + d_2 \varepsilon_{t-2}^2 + d_3 \varepsilon_{t-3}^2 + d_4 \varepsilon_{t-4}^2 + d_5 \varepsilon_{t-5}^2
\end{aligned} \tag{5.3}$$

Parameter estimates for Equation 5.3 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. This behavior may be related to the aforementioned identification problems.

In summary, neither the restriction suggested by the Engle-ARCH model nor the other extreme Equation 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 \mathbf{B} is much slower than would be prescribed by Equation 2.4.

Although I use no explicit Bayesian framework, I adopt the view that fitting models to data is an exercise of optimum approximation within classes of models rather than estimation of true structures. Having abandoned the idea that the S&P Index really follows a conditionally Gaussian ARCH model, many specification tests do not make sense any more as the concept of 'misspecification' becomes an empty word. However, one may still be interested in the sensitivity of relative advantages of models with respect to changes of the sampling intervals.

Hence, not to 'test' but to check on the stability of Equation 5.3 with respect to time-heterogeneity, Equation 5.3 was re-estimated for subsamples. The selection of subsamples was inspired by Hauser and

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

	1962–67		1969–78		1979–86		1962–90	
$\sqrt{d_1}$	0.601	0.627	0.500	0.504	0.175	0.175	0.529	0.533
$\sqrt{d_2}$	0.218	0.103	1.200	0.211	0.424	0.120	0.342	0.087
$\sqrt{d_3}$	0.183	0.008	0.473	0.286	0.731	0.860	0.254	0.176
$\sqrt{d_4}$	0.002	0.224	0.007	0.286	0.581	0.027	0.355	0.541
$\sqrt{d_5}$	0.172	0.370	0.617	0.521	0.075	0.098	0.246	0.294
l_{21}	-0.220		-0.514		-0.805		-0.568	
l_{31}	0.383		-.263		0.352		0.361	
l_{41}	0.197		0.447		-0.013		-0.110	
l_{51}	0.218		0.746		0.117		0.252	
$\sqrt{a_0}$	0.00311	0.00023	0.00000	0.00522	0.00066	0.00002	0.00392	0.00163
μ	0.00023	0.00035	0.00001	0.00010	0.00043	0.00043	0.00021	0.00031
φ_1	0.197	0.170	0.295	0.302	0.094	0.094	0.191	0.194
φ_2	0.045	0.063	-.098	-.103	0.002	0.002	-.062	-.060
φ_3	-.003	0.012	0.039	0.047	-.016	-.016	-.022	-.016
φ_4	-.041	-.026	0.022	0.023	-.020	-.021	0.027	0.026
φ_5	0.028	0.041	-.037	-.037	0.008	0.007	0.015	0.014
$\sum a_{ii}$	0.573	0.591	2.567	0.734	1.112	0.795	0.796	0.702
-2ℓ	13 202	13 183	21 751	21 746	17 102	17 102	61 573	61 559

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 (5.3). 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 ratio for testing the restricted Engle-ARCH model against the more general form (Equation 5.3) would be χ^2 -distributed with four degrees of freedom. The restriction would be 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 (these results are probably not very interesting on their own and therefore not reported). For the years 1969–78, unrestricted estimation resulted in an unsatisfactory model which did not meet covariance-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.

Many coefficient estimates differ quite a lot among subsamples, but some features were remarkably stable. Firstly, in the lag pattern of ARCH coefficients, d_1 dominates (except 1979–86) 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 Equation 5.3 is not the same as the series itself and hence the data cannot be described by a structure such as Equation 2.4. In summary, the 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.

In two cases, covariance-stationarity conditions were not met. In Section III, the feature of strict stationarity—as opposed to covariance stationarity—was treated and it was shown that it typically requires less stringent conditions. Hence, the model estimated for 1979–86 is probably still strictly stationary (while the estimated model for 1969–78 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 Equation 5.3.

A German stock market index

The German stock market index provided by Morgan Stanley Capital International (MSCI) used here has 2843 observations starting at 1 August 1984, and ending at 13 November 1995. Logarithmic differences of the original series are analyzed. Linear time dependence is weak but autocorrelations are statistically significant up to a lag of seven. A linear autoregressive model of order five appears to capture most of this low-order serial correlation structure. Just as in the Standard & Poor Index, serial correlation in the squared residuals is more substantial. This second-order correlation is low at lag three but then increases again and is still felt at lag lengths exceeding one trading week.

In order to use the likelihood optimization algorithm, extremely insignificant elements have to be eliminated *a priori* to avoid numerical difficulties. A good exploratory tool is linear least-squares regression of the squared residuals on lagged squares and cross-products. The third-order diagonal element appeared to be very small and the following model structure was used for non-linear optimization:

$$y_t = \mu + \sum_{i=1}^5 \varphi_i y_{t-i} + \varepsilon_t \quad \mathbb{E}(\varepsilon_t^2 | I_{t-1}) = \alpha_0 + \varepsilon'_- \mathbf{L} \mathbf{D} \mathbf{L}' \varepsilon_-$$

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 & 0 \\ l_{41} & l_{42} & 0 & 1 & 0 \\ l_{51} & l_{52} & 0 & l_{54} & 1 \end{bmatrix}$$

$$\mathbf{D} = \text{diag}(d_1, d_2, 0, d_4, d_5)$$

This statistical model contains 19 parameters and the corresponding estimates are given in Table 2. Note that, due to the preliminary search procedure, standard errors are not valid and have been omitted. It should be mentioned, however, that almost all coefficients turn out to be significantly different from zero if the usual guideline is followed, including the l_{ij} for $i \neq j$ that discriminate the augmented ARCH from the standard ARCH model.

This result is confirmed by a comparison with a fifth-order standard ARCH model. The difference in the log likelihood is extremely

Table 2: *Coefficient estimates from general ARCH model (4.3) and from Engle-ARCH model for Stock Price Indexes from several countries.*

	Germany		France		United Kingdom		Japan	
$\sqrt{d_1}$	0.279	0.263	0.346	0.318	0.239	0.243	0.428	0.438
$\sqrt{d_2}$	0.520	0.211	0.061	0.085	0.591	0.475	0	0.513
$\sqrt{d_3}$	0.153	0.125	0.236	0.266	0	0.115	0.348	0.189
$\sqrt{d_4}$	0	0.614	0	0.051	0.085	0.232	0.331	0.159
$\sqrt{d_5}$	0.024	0.895	0.332	0.297	0	0.174	0.477	0.151
l_{21}	0.117		-0.032		0.225		0.599	
l_{31}	-0.394		0.009		0.030		-0.173	
l_{41}	-0.298		0.131		-0.142		-0.103	
l_{51}	0.422		0.027		-0.022		0.171	
l_{32}	0.117		0.098		-0.041			
l_{42}	-0.394		0.196		-0.139			
l_{52}	-0.298		-0.005		0.079			
l_{43}			0.037				-0.042	
l_{53}			0.003				0.118	
l_{54}	0.422				0.024		0.205	
$\sqrt{a_0}$	0.00160	0.00497	0.00218	0.00187	0.00372	0.00992	0.01434	0.00043
μ	0.00023	0.00042	0.00040	0.00034	0.00043	0.00047	-0.00001	0.00046
φ_1	0.044	0.059	0.246	0.223	0.081	0.084	0.198	0.187
φ_2	0.016	0.022	-0.007	0.010	0.015	0.015	-0.081	-0.090
φ_3	-0.015	-0.009	-0.036	-0.033	-0.012	-0.013	0.008	-0.001
φ_4	0.032	0.029	0.051	0.014	-0.006	-0.007	0.035	0.037
φ_5	0.027	0.028	0.012	0.010	-0.001	-0.002	-0.048	-0.045
$\sum a_{ii}$	0.566	1.307	0.292	0.270	0.427	0.387	0.727	0.540
-2ℓ	44 953	44 891	45 837	45 832	48 505	48 503	44 596	44 314

significant if the usual asymptotic chi-square guideline—here with eight degrees of freedom—is adopted. Note, however, that the models are incompletely nested because of the restriction $d_3 = 0$ in the general model. The standard ARCH model is not covariance stationary but it appears to be strictly stationary. It generates peculiar trajectories that do not match the behavior of the observed time series. In contrast, the trajectories generated from the augmented ARCH structure cannot be distinguished visually from the data. Another indicator of the poor approximation by the standard ARCH model is the comparatively large d_3 estimate that appears to pick up some part of the higher-order structure which is more appropriately reflected in the l_{ij} coefficients of the augmented model.

Although such correspondence between generated data from a fitted model and observed data is likely not sufficient to justify the use of the fitted model, the usage of post-sample simulation techniques is to be appreciated. Many reported statistical models do not produce plausible trajectories and sometimes are unable to produce any trajectories at all. Obviously, these models must be discarded.

It also pays to compare sample moments and other features between artificial and observed data. In the present case, the observed skewness and high kurtosis is not reflected fully in most trajectories of comparable length. Most sample kurtosis values in simulations ranged between 4 and 15, whereas observed kurtosis is 16.4. As in most applications of ARCH models, the conditional distribution in the actual data is non-Gaussian and this causes the slight mismatch.

On the other hand, observed negative skewness cannot be represented in the ARCH framework. If this feature is regarded as relevant, bilinear models may be an obvious suggestion. In contrast, the evidence does not support the ARCH-in-mean model which also generates non-normal third moments from its parametric structure and has a stronger basis in economic theory. This observation also holds for our examples, as was found by some tentative estimation.

A French stock market index

The MSCI time series on the Paris stock exchange covers the same time range as the German series, from 1 August 1984 to 13 November 1995. Due to different holidays, the French series is slightly longer with 2846 observations. We again analyze the time series of the log differences or approximate returns.

The linear time series structure is weaker than in the German series, and so is the second-order structure in the residuals. Only the first-order lag and some lags exceeding one trading week were found to

be significant, which may suggest a first-order autoregression instead of the fifth-order autoregression that we have been tentatively using so far. Regarding the ARCH structure, the fourth-order lag turned out to be very small and hence d_4 was set to zero before conducting maximum likelihood estimation.

Two models are considered in Table 2: first, the most general ARCH model with a fifth-order autoregression, $d_4 = 0$ being the only restriction; secondly, a standard fifth-order ARCH model. It was also attempted to fit a more parsimonious augmented model by eliminating all seemingly insignificant parameters based on a linear first-order autoregression and $d_2 = d_4 = 0$. However, it turned out that this model involves a substantial deterioration in the quality of fit relative to the first model. On the other hand, the standard ARCH model comes close to the full model. Generated trajectories from all three models are similar to each other and plausibly close to the actual data. In consequence, evidence on the French index does not seem to support the usage of the augmented ARCH model, as the standard ARCH model achieves a reasonable precision.

A British stock market index

The UK stock market index provided by MSCI covers the time range from 1 August 1984 to 10 November 1995. For the logarithmic differences, there are 2863 observations available.

The British series shows a noteworthy linear autocorrelation around the monthly lag that we are unable to account for, restricting ourselves to low-order models. ARCH effects seem to be somewhat stronger than for France but weaker than in the German or in the Standard & Poor series. Nevertheless, the data is highly leptokurtic. Hence, a large amount of the leptokurtosis observed cannot be modeled by the parametric ARCH process but rather is rooted in the conditional distribution. In consequence, the efficiency of the maximum likelihood estimation procedure may be adversely affected to a much larger degree than in the other series examined.

Fitting a standard fifth-order ARCH model results in a sum of coefficients of almost 0.4. The first, second, and fourth lag turn out to be the most significant. Curiously, this maximum likelihood result is seriously at odds with the naive least-squares fitting procedure which yielded an insignificant second-order lag. Again, this discrepancy may be due to the high leptokurtosis in the conditional distribution. Anyway, an augmented ARCH model was fitted based on non-zero elements at the places one, two, and four and also on non-zero elements in the corresponding columns of \mathbf{L} . A variety of alternative models

was tried but this design happened to be the optimum with regard to the achieved likelihood.

The results are again given in Table 2. It is seen that there is no significant difference between the quality of fit of the standard ARCH model fit and of the augmented model. In this regard, the British series resembles the French series. However, due to the high degree of non-normality, the quasi-maximum likelihood procedure appears less reliable.

A Japanese index

The Japanese series on the MSCI index for the Tokyo stock exchange has 2944 observations from 1 August 1984 to 13 November 1995. We again focus on the logarithmic differences. Also for Japan, the linear correlation structure is weak and the volatility correlation is considerably stronger. However, in this case all dependence is at lags of less than five and evidence on ‘misspecification’ due to omitted longer-range dependence is much weaker than for the European index series. The second-order ARCH lag appears to be insignificant and was omitted for the specification search within the augmented ARCH model class.

The coefficient estimates corresponding to the best model found are given in Table 2. The general model is significantly better than the restricted classical ARCH model. Note that the Engle-ARCH model contains a noteworthy second-order ARCH lag. It was also tried to add this coefficient to the augmented model but this resulted in no further improvement. Simulations from the identified structures confirmed the inadequacy of the Engle-ARCH model. Trajectories showed linear time trends with only small added perturbations. Simulated augmented ARCH processes had shapes similar to the observed series. Hence, the Japanese index constitutes another example of the potential benefits from considering the more general class.

VI SUMMARY AND CONCLUSION

For the important US Standard & Poor 500 Index and for a Japanese and a German stock index, the evidence has indicated more or less convincingly that fourth-moments structures in financial series may be more complicated than the traditional ARCH model implies. A non-parametric comparison of sample moments (unreported) also supported this result. The statistical evidence presented is stronger than the ‘weak evidence on more general structures’ found by Tsay (1987) in an exchange rate series. For two other countries, France and the

United Kingdom, the statistical description achieved by the standard ARCH model appears to be sufficient.

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 may affect the efficiency of quasi maximum likelihood estimation and the validity of the chi-square approximation to the likelihood-ratio test.

Secondly, higher-order moments of different form were neglected. In its 'weak ARCH' interpretation, 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. It should be pointed out that the 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.

The framework presented here necessarily excludes all complicated non-linear models whose popularity is still increasing. We hold the view that, unless convincing theoretical motivations for special non-linear classes can be found, the application of arbitrary functional forms is not natural from a time-series analytical point of view. If dependence in volatility is investigated, then a full understanding of simple stationary models describing cross moments of order up to four is sure an important point to begin with.

ACKNOWLEDGEMENTS

The author wishes to thank Michael Hauser for extensive help with all computer work, the participants of the NBER Time Series Seminar 1993, in particular Ruey Tsay, for helpful comments, and a referee for drawing attention to the comparative empirical application presented here. The usual proviso applies.

References

- [1] Bera, A.K., Higgins, M.L., and Lee, S. (1992) Interaction Between Autocorrelation and Conditional Heteroskedasticity A Random Coefficient Approach, *Journal of Business and Economic Statistics* **10**, 133-142.
- [2] Baillie, R.T. and Bollerslev, T. (1989) The Message in Daily Exchange Rates. A Conditional-Variance Tale, *Journal of Business and Economic Statistics* **7**, 297-305.
- [3] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307-327.
- [4] Bougerol, P. and Picard, N. (1992a) Stationarity of GARCH processes and of some nonnegative time series, *Journal of Econometrics* **52**, 115-127.
- [5] Bougerol, P. and Picard, N. (1992b) Strict Stationarity of Generalized Autoregressive Processes, *The Annals of Probability* **20**, 1714-1730.
- [6] Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, *Advances in Applied Probability* **18**, 211-220.
- [7] Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**, 987-1007.
- [8] Engle, R.F. and Lee, G.G.J. (1993) A Permanent and Transitory Component Model of Stock Return Volatility, Paper presented at the ESEM Meeting, Uppsala.
- [9] Hauser, M.A. and Kunst, R.M. (1993) Fractionally Integrated Models With ARCH Errors, Research Memorandum No. 322, Institute for Advanced Studies, Vienna, Austria.
- [10] Kleibergen, F. and Van Dijk, H.K. (1993) Non-Stationarity in GARCH Models: A Bayesian Analysis, *Journal of Applied Econometrics* **8**, S41-S61.

- [11] Kunst, R.M. (1993) Stability Conditions for a Bivariate ARCH System Which Is Cointegrated in Mean, *Communications in Statistics, Theory and Methods* **22**, 2941-2953.
- [12] Kunst, R.M. (1997) Fourth-order Moments of Augmented ARCH Processes. *Communications in Statistics, Theory and Methods*, **26**, 1425–41.
- [13] Kunst, R.M. and Polasek, W. (1994) Structuring Volatile Swiss Interest Rates: Some Evidence on the Present Value Model and a VAR-VARCH Approach, in *Econometric Analysis of Financial Markets*, Eds J. Kaehler and P. Kugler, Physica-Verlag, Heidelberg.
- [14] Lamoureux, C. and Lastrapes, W. (1990) Persistence in Variance, Structural Change, and the GARCH Model, *Journal of Business & Economic Statistics* **8**, 225-234.
- [15] Milhøj, A. (1985) The Moment Structure of ARCH Processes, *Scandinavian Journal of Statistics* **12**, 281-292.
- [16] Nelson, D.B. (1990) Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory* **6**, 318-334.
- [17] Nicholls, D.F. and Quinn, B.G. (1982) Random coefficient autoregressive models: an introduction. *Lecture Notes in Statistics*, Vol. 11, Springer, New York.
- [18] Taylor, S. (1986) *Modelling Financial Time Series*, Wiley.
- [19] Tsay, R.S. (1987) Conditional Heteroscedastic Time Series Models, *Journal of the American Statistical Association* **82**, 590-604.
- [20] Weiss, A.A. (1984) ARMA models with ARCH errors, *Journal of Time Series Analysis* **5**, 129-143.

Appendix A

Proof of Proposition 1

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

$$w'_t = (\varepsilon_t^2, \varepsilon_{t-1}^2, \dots, \varepsilon_{t-R+1}^2, \varepsilon_t \varepsilon_{t-1}, \varepsilon_{t-1} \varepsilon_{t-2}, \dots, \varepsilon_{t-R+2} \varepsilon_{t-R+1}, \varepsilon_t \varepsilon_{t-2}, \varepsilon_{t-1} \varepsilon_{t-3}, \dots, \varepsilon_t \varepsilon_{t-R+1})$$

(Alternatively, we could use $\text{vech}((\varepsilon_t, \dots, \varepsilon_{t-R+1})(\varepsilon_t, \dots, \varepsilon_{t-R+1})')$ as a 'state vector' but the present derivation appears more convenient.)

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 Equation 2.1 as

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

with $a'_V = (a_0, 0, \dots, 0)$ and with \mathbf{M} constructed as follows

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

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

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

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 \mathbf{M} are entirely determined by the a_{ii} elements. Assumption 4 is crucial as otherwise the rows $R+1, 2R, \dots$ would not necessarily be zero.

Now suppose $R \rightarrow \infty$. 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.

APPENDIX B

Scores for the augmented ARCH model

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

$$h_t = \varepsilon'_{t-1} \mathbf{L}\mathbf{D}\mathbf{L}' \varepsilon_{t-1} \quad (\text{A.1})$$

with $\varepsilon_t = (\varepsilon_t, \dots, \varepsilon_{t-R+1})'$. A disadvantage of the re-parameterization in Equation A.1 is that some elements of \mathbf{L} become unidentified if \mathbf{D} has zero elements. In practice, it is suggested to check on the d_i via some preliminary Engle-ARCH estimation, i.e., imposing $\mathbf{L} = \mathbf{I}$. If, say, d_i is close to 0, $d_i = 0$ and $l_{ji} = 0$ ($j = i, \dots, R$) can be imposed during the following model selection stages. After finishing these, the significance of $d_i = 0$ can be re-checked. The empirical results presented below were based on this selection procedure. The advantages sure outweigh this inconvenience.

Further assuming ε_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

$$h_t = \mathbf{y}'_{t-1} \mathbf{C}' \mathbf{L}\mathbf{D}\mathbf{L}' \mathbf{C} \mathbf{y}_{t-1} \quad (\text{A.2})$$

with \mathbf{C} being an $R \times (R+S)$ -matrix with its i th row containing the autoregressive coefficients $(1, \zeta_1, \dots, \zeta_{S-1})$ 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 ($i = 1, \dots, S-1$) is

$$\frac{\partial h_t}{\partial \zeta_i} = 2\mathbf{y}'_{t-1} \mathbf{C}' \mathbf{L}\mathbf{D}\mathbf{L}' \mathbf{y}_{t-i-1} \quad (\text{A.3})$$

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

Equation A.3 allows to evaluate the scores with respect to ζ_i . In detail,

$$\frac{\partial l_t}{\partial \zeta_i} = \frac{\varepsilon_t y_{t-i}}{h_t} + \frac{1}{2h_t} \frac{\partial h_t}{\partial \zeta_i} \frac{\varepsilon_t^2 - h_t}{h_t} \quad (\text{A.4})$$

and substitution of Equation A.3 into Equation A.4 yields an operable expression.

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

$$\frac{\partial h_t}{\partial d_i} = \left(\sum_{j=1}^R l_{ji} \varepsilon_{t-j} \right)^2 \quad \frac{\partial l_t}{\partial d_i} = \frac{\partial h_t}{\partial d_i} \frac{1}{2h_t} \frac{\varepsilon_t^2 - h_t}{h_t} \quad (\text{A.5})$$

and similarly

$$\frac{\partial h_t}{\partial l_{ij}} = 2d_j \left(\sum_{k=j}^R l_{kj} \varepsilon_{t-k} \right) \varepsilon_{t-j} \quad \frac{\partial l_t}{\partial l_{ij}} = \frac{\partial h_t}{\partial l_{ij}} \frac{1}{2h_t} \frac{\varepsilon_t^2 - h_t}{h_t} \quad (\text{A.6})$$