Conditional Heteroskedasticity in Asset Returns: A New Approach

Daniel B. Nelson

Econometrica, Volume 59, Issue 2 (Mar., 1991), 347-370.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199103%2959%3A2%3C347%3ACHIARA%3E2.0.CO%3B2-V

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Econometrica is published by The Econometric Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/econosoc.html.

Econometrica
©1991 The Econometric Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR

http://www.jstor.org/
Tue Jan 29 10:09:41 2002

Econometrica, Vol. 59, No. 2 (March, 1991), 347–370

CONDITIONAL HETEROSKEDASTICITY IN ASSET RETURNS: A NEW APPROACH

BY DANIEL B. NELSON[1]

GARCH models have been applied in modelling the relation between conditional variance and asset risk premia. These models, however, have at least three major drawbacks in asset pricing applications: (i) Researchers beginning with Black (1976) have found a negative correlation between current returns and future returns volatility. GARCH models rule this out by assumption. (ii) GARCH models impose parameter restrictions that are often violated by estimated coefficients and that may unduly restrict the dynamics of the conditional variance process. (iii) Interpreting whether shocks to conditional variance "persist" or not is difficult in GARCH models, because the usual norms measuring persistence often do not agree. A new form of ARCH is proposed that meets these objections. The method is used to estimate a model of the risk premium on the CRSP Value-Weighted Market Index from 1962 to 1987.

KEYWORDS: Autoregressive conditional heteroskedasticity (ARCH), generalized autoregressive conditional heteroskedasticity (GARCH), exponential ARCH, market volatility, risk premium, nonlinear time series.

1. INTRODUCTION

AFTER THE EVENTS of October 1987, few would argue with the proposition that stock market volatility changes randomly over time. Understanding the way in which it changes is crucial to our understanding of many areas in macroeconomics and finance, for example the term structure of interest rates (e.g., Barsky (1989), Abel (1988)), irreversible investment (e.g., Bernanke (1983), McDonald and Siegel (1986)), options pricing (e.g., Wiggins (1987)), and dynamic capital asset pricing theory (e.g., Merton (1973), Cox, Ingersoll, and Ross (1985)).

Recent years have also seen a surge of interest in econometric models of changing conditional variance. Probably the most widely used, but by no means the only such models,[2] are the family of ARCH (autoregressive conditionally heteroskedastic) models introduced by Engle (1982). ARCH models make the conditional variance of the time t prediction error a function of time, system parameters, exogenous and lagged endogenous variables, and past prediction errors. For each integer t, let ξ_t be a model's (scalar) prediction error, b a vector of parameters, x_t a vector of predetermined variables, and σ_t^2 the variance of ξ_t given information at time t. A univariate ARCH model based on

[1] This paper is a revision of part of Chapter III of my Ph.D. dissertation (Nelson (1988)). The Department of Education and the University of Chicago Graduate School of Business provided financial support. I am indebted to George Constantinides, John Cox, Nancy Hammond, Daniel McFadden, Mark Nelson, Adrian Pagan, James Poterba, G. William Schwert, Stephen Taylor, Jeffrey Wooldridge, Arnold Zellner, two referees, and a co-editor for helpful comments. Seminar participants at the 1987 Econometric Society Winter meetings, at the 1989 A.S.A. meetings, at Berkeley, Chicago, M.I.T., Northwestern, Princeton, the Research Triangle Econometrics Workshop, Rochester, Stanford, Wharton, and Yale also made helpful suggestions. Jiahong Shi provided able research assistance. Remaining errors are mine alone.

[2] See, e.g., Poterba and Summers (1986), French, Schwert, and Stambaugh (1987), and Nelson (1988, Chapter 1).

Engle (1982) equations 1–5 sets

$$\xi_t = \sigma_t z_t, \tag{1.1}$$

$$z_t \sim \text{i.i.d. with} \tag{1.2}$$

$$E(z_t) = 0, \quad \operatorname{Var}(z_t) = 1, \quad \text{and}$$

$$\begin{aligned} \sigma_t^2 &= \sigma^2(\xi_{t-1}, \xi_{t-2}, \ldots, t, x_t, b) \\ &= \sigma^2(\sigma_{t-1} z_{t-1}, \sigma_{t-2} z_{t-2}, \ldots, t, x_t, b). \end{aligned} \tag{1.3}$$

The system (1.1)–(1.3) can easily be given a multivariate interpretation, in which case z_t is an n by one vector and σ_t^2 is an n by n matrix. We refer to any model of the form (1.1)–(1.3), whether univariate or multivariate, as an ARCH model.

The most widely used specifications for $\sigma^2(\cdot, \cdot, \ldots, \cdot)$ are the linear ARCH and GARCH models introduced by Engle (1982) and Bollerslev (1986) respectively, which make σ_t^2 linear in lagged values of $\xi_t^2 = \sigma_t^2 z_t^2$ by defining

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j z_{t-j}^2 \sigma_{t-j}^2, \qquad \text{and} \tag{1.4}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{p} \alpha_j z_{t-j}^2 \sigma_{t-j}^2, \tag{1.5}$$

respectively, where ω, the α_j, and the β_i are nonnegative. Since (1.4) is a special case of (1.5), we refer to both (1.4) and (1.5) as GARCH models, to distinguish them as special cases of (1.3).

The GARCH-M model of Engle and Bollerslev (1986a), adds another equation

$$R_t = a + b\sigma_t^2 + \xi_t, \tag{1.6}$$

in which σ_t^2, the conditional variance of R_t, enters the conditional mean of R_t as well. For example if R_t is the return on a portfolio at time t, its required rate of return may be linear in its risk as measured by σ_t^2.

Researchers have fruitfully applied the new ARCH methodology in asset pricing models: for example, Engle and Bollerslev (1986a) used GARCH(1, 1) to model the risk premium on the foreign exchange market, and Bollerslev, Engle, and Wooldridge (1988) extended GARCH(1, 1) to a multivariate context to test a conditional CAPM with time varying covariances of asset returns.

Substituting recursively for the $\beta_i \sigma_{t-i}^2$ terms lets us rewrite (1.5) as[3]

$$\sigma_t^2 = \omega^* + \sum_{k=1}^{\infty} \phi_k z_{t-k}^2 \sigma_{t-k}^2. \tag{1.7}$$

It is readily verified that if ω, the α_j, and the β_i are nonnegative, ω^* and the ϕ_k are also nonnegative. By setting conditional variance equal to a constant plus a

[3] The representation (1.7) assumes that $\{\sigma_t^2\}$ is strictly stationary, so that the recursion can be carried into the infinite past.

weighted average (with positive weights) of past squared residuals, GARCH models elegantly capture the volatility clustering in asset returns first noted by Mandelbrot (1963): "...large changes tend to be followed by large changes—of either sign—and small changes by small changes..." This feature of GARCH models accounts for both their theoretical appeal and their empirical success.

On the other hand, the simple structure of (1.7) imposes important limitations on GARCH models: For example, researchers beginning with Black (1976) have found evidence that stock returns are negatively correlated with changes in returns volatility—i.e., volatility tends to rise in response to "bad news" (excess returns lower than expected) and to fall in response to "good news" (excess returns higher than expected).[4] GARCH models, however, assume that only the magnitude and not the positivity or negativity of unanticipated excess returns determines feature σ_t^2. If the distribution of z_t is symmetric, the change in variance tomorrow is conditionally uncorrelated with excess returns today.[5] In (1.4)–(1.5), σ_t^2 is a function of lagged σ_t^2 and lagged z_t^2, and so is invariant to changes in the algebraic sign of the z_t's—i.e., only the size, not the sign, of lagged residuals determines conditional variance. This suggests that a model in which σ_t^2 responds asymmetrically to positive and negative residuals might be preferable for asset pricing applications.

Another limitation of GARCH models results from the nonnegativity constraints on ω^* and the ϕ_k in (1.7), which are imposed to ensure that σ_t^2 remains nonnegative for all t with probability one. These constraints imply that increasing z_t^2 in any period increases σ_{t+m}^2 for all $m \geqslant 1$, ruling out random oscillatory behavior in the σ_t^2 process. Furthermore, these nonnegativity constraints can create difficulties in estimating GARCH models. For example, Engle, Lilien, and Robins (1987), had to impose a linearly declining structure on the α_j coefficients in (1.4) to prevent some of them from becoming negative.

A third drawback of GARCH modelling concerns the interpretation of the "persistence" of shocks to conditional variance. In many studies of the time series behavior of asset volatility (e.g., Poterba and Summers (1986), French, Schwert, and Stambaugh (1987), and Engle and Bollerslev (1986a)), the central question has been how long shocks to conditional variance persist. If volatility shocks persist indefinitely, they may move the whole term structure of risk premia, and are therefore likely to have a significant impact on investment in long-lived capital goods (Poterba and Summers (1986)).

There are many different notions of convergence in the probability literature (almost sure, in probability, in L^p), so whether a shock is transitory or

[4] The economic reasons for this are unclear. As Black (1976) and Christie (1982) note, both financial and operating leverage play a role, but are not able to explain the extent of the asymmetric response of volatility to positive and negative returns shocks. Schwert (1989a, b) presents evidence that stock volatility is higher during recessions and financial crises, but finds only weak relations between stock market volatility and measures of macroeconomic uncertainty. See also Nelson (1988), and Pagan and Hong (1988).

[5] Adrian Pagan pointed out, however, that in a GARCH-M model σ_t^2 may rise (fall) on average when returns are negative (positive), even though $\sigma_{t+1}^2 - \sigma_t^2$ and R_t are conditionally uncorrelated, since in (1.6) $E[|z_t| \,|R_t < 0] > E[|z_t| \,|R_t \geqslant 0]$ if a and b are positive.

persistent may depend on our definition of convergence. In linear models it typically makes no difference which of the standard definitions we use, since the definitions usually agree. In GARCH models, the situation is more complicated. For example, the IGARCH(1, 1) model of Engle and Bollerslev (1986a) sets

$$(1.8) \qquad \sigma_t^2 = \omega + \sigma_{t-1}^2\left[(1-\alpha) + \alpha z_{t-1}^2\right], \qquad 0 < \alpha \leqslant 1.$$

When $\omega = 0$, σ_t^2 is a martingale. Based on the nature of persistence in linear models, it seems that IGARCH(1, 1) with $\omega > 0$ and $\omega = 0$ are analogous to random walks with and without drift, respectively, and are therefore natural models of "persistent" shocks. This turns out to be misleading, however: in IGARCH(1, 1) with $\omega = 0$, σ_t^2 collapses to zero almost surely, and in IGARCH(1, 1) with $\omega > 0$, σ_t^2 is strictly stationary and ergodic (Geweke (1986), Nelson (1990a)) and therefore does not behave like a random walk, since random walks diverge almost surely.

The reason for this paradox is that in GARCH(1, 1) models, shocks may persist in one norm and die out in another, so the conditional moments of GARCH(1, 1) may explode even when the process itself is strictly stationary and ergodic (Nelson (1990a)).

The object of this paper is to present an alternative to GARCH that meets these objections, and so may be more suitable for modelling conditional variances in asset returns. In Section 2, we describe this σ_t^2 process, and develop some of its properties. In Section 3 we estimate a simple model of stock market volatility and the risk premium. Section 4 concludes. In Appendix I, we provide formulas for the moments of σ_t^2 and ξ_t in the model presented in Section 2. Proofs are in Appendix II.

2. EXPONENTIAL ARCH

If σ_t^2 is to be the conditional variance of ξ_t given information at time t, it clearly must be nonnegative with probability one. GARCH models ensure this by making σ_t^2 a linear combination (with positive weights) of positive random variables. We adopt another natural device for ensuring that σ_t^2 remains nonnegative, by making $\ln(\sigma_t^2)$ linear in some function of time and lagged z_t's. That is, for some suitable function g:

$$(2.1) \qquad \ln\left(\sigma_t^2\right) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}), \qquad \beta_1 \equiv 1,$$

where $\{\alpha_t\}_{t=-\infty,\infty}$ and $\{\beta_k\}_{k=1,\infty}$ are real, nonstochastic, scalar sequences. Pantula (1986) and Geweke (1986) have previously proposed ARCH models of this form; in their log-GARCH models, $g(z_t) = \ln|z_t|^b$ for some $b > 0$.[6]

[6] See also Engle and Bollerslev (1986b).

To accommodate the asymmetric relation between stock returns and volatility changes noted in Section 1, the value of $g(z_t)$ must be a function of both the magnitude and the sign of z_t. One choice, that in certain important cases turns out to give σ_t^2 well-behaved moments, is to make $g(z_t)$ a linear combination of z_t and $|z_t|$:

$$(2.2) \qquad g(z_t) \equiv \theta z_t + \gamma[|z_t| - E|z_t|].$$

By construction, $\{g(z_t)\}_{t=-\infty,\infty}$ is a zero-mean, i.i.d. random sequence. The two components of $g(z_t)$ are θz_t and $\gamma[|z_t| - E|z_t|]$, each with mean zero. If the distribution of z_t is symmetric, the two components are orthogonal, though of course they are not independent. Over the range $0 < z_t < \infty$, $g(z_t)$ is linear in z_t with slope $\theta + \gamma$, and over the range $-\infty < z_t \leqslant 0$, $g(z_t)$ is linear with slope $\theta - \gamma$. Thus, $g(z_t)$ allows the conditional variance process $\{\sigma_t^2\}$ to respond asymmetrically to rises and falls in stock price.

To see that the term $\gamma[|z_t| - E|z_t|]$ represents a magnitude effect in the spirit of the GARCH models discussed in Section 1, assume for the moment that $\gamma > 0$ and $\theta = 0$. The innovation in $\ln(\sigma_{t+1}^2)$ is then positive (negative) when the magnitude of z_t is larger (smaller) than its expected value. Suppose now that $\gamma = 0$ and $\theta < 0$. The innovation in conditional variance is now positive (negative) when returns innovations are negative (positive). Thus the exponential form of ARCH in (2.1)–(2.2) meets the first objection raised to the GARCH models in Section 1.

In Section 1 we also argued that the dynamics of GARCH models were unduly restrictive (i.e., oscillatory behavior is excluded) and that they impose inequality constraints that were frequently violated by estimated coefficients. But note that in (2.1)–(2.2) there are no inequality constraints whatever, and that cycling is permitted, since the β_k terms can be negative or positive.

Our final criticism of GARCH models was that it is difficult to evaluate whether shocks to variance "persist" or not. In exponential ARCH, however, $\ln(\sigma_t^2)$ is a linear process, and its stationarity (covariance or strict) and ergodicity are easily checked. If the shocks to $\{\ln(\sigma_t^2)\}$ die out quickly enough, and if we remove the deterministic, possibly time-varying component $\{\alpha_t\}$, then $\{\ln(\sigma_t^2)\}$ is strictly stationary and ergodic. Theorem 2.1 below states conditions for the ergodicity and strict stationarity of $\{\exp(-\alpha_t)\sigma_t^2\}$ and $\{\exp(-\alpha_t/2)\xi_t\}$, which are $\{\sigma_t^2\}$ and $\{\xi_t\}$ with the influence of $\{\alpha_t\}$ removed.

THEOREM 2.1: *Define* $\{\sigma_t^2\}$, $\{\xi_t\}$, *and* $\{z_t\}$ *by* (1.1)–(1.2) *and* (2.1)–(2.2), *and assume that* γ *and* θ *do not both equal zero. Then* $\{\exp(-\alpha_t)\sigma_t^2\}$, $\{\exp(-\alpha_t/2)\xi_t\}$, *and* $\{\ln(\sigma_t^2) - \alpha_t\}$ *are strictly stationary and ergodic and* $\{\ln(\sigma_t^2) - \alpha_t\}$ *is covariance stationary if and only if* $\sum_{k=1}^{\infty}\beta_k^2 < \infty$. *If* $\sum_{k=1}^{\infty}\beta_k^2 = \infty$, *then* $|\ln(\sigma_t^2) - \alpha_t| = \infty$ *almost surely. If* $\sum_{k=1}^{\infty}\beta_k^2 < \infty$, *then for* $k > 0$, $\mathrm{Cov}\{z_{t-k}, \ln(\sigma_t^2)\} = \beta_k[\theta + \gamma E(z_t|z_t|)]$, *and* $\mathrm{Cov}[\ln(\sigma_t^2), \ln(\sigma_{t-k}^2)] = \mathrm{Var}[g(z_t)]\sum_{j=1}^{\infty}\beta_j\beta_{j+k}$.

The stationarity and ergodicity criterion in Theorem 2.1 is exactly the same as for a general linear process with finite innovations variance,[7] so if, for example, $\ln(\sigma_t^2)$ follows an AR(1) with AR coefficient $\Delta, \ln(\sigma_t^2)$ is strictly stationary and ergodic if and only if $|\Delta| < 1$.

There is often a simpler expression for $\ln(\sigma_t^2)$ than the infinite moving average representation in (2.1). In many applications, an ARMA process provides a parsimonious parameterization:

$$\ln(\sigma_t^2) = \alpha_t + \frac{(1 + \Psi_1 L + \cdots + \Psi_q L^q)}{(1 - \Delta_1 L - \cdots - \Delta_p L^p)} g(z_{t-1}). \tag{2.3}$$

We assume that $[1 - \Sigma_{i=1,p} \Delta_i y^i]$ and $[1 + \Sigma_{i=1,q} \Psi_i y^i]$ have no common roots. By Theorem 2.1, $\{\exp(-\alpha_t)\sigma_t^2\}$ and $\{\exp(-\alpha_t/2)\xi_t\}$ are then strictly stationary and ergodic if and only if all the roots of $[1 - \Sigma_{i=1,p} \Delta_i y^i]$ lie outside the unit circle. While an ARMA representation may be suitable for many modelling purposes, Theorem 2.1 and the representation (2.1) also allow "long memory" (Hosking (1981)) processes for $\{\ln(\sigma_t^2)\}$.

Next consider the covariance stationarity of $\{\sigma_t^2\}$ and $\{\xi_t\}$. According to Theorem 2.1, $\Sigma\beta_i^2 < \infty$ implies that $\{\exp(-\alpha_t)\sigma_t^2\}$ and $\{\exp(-\alpha_t/2)\xi_t\}$ are strictly stationary and ergodic. This strict stationarity, however, need not imply covariance stationarity, since $\{\exp(-\alpha_t)\sigma_t^2\}$ and $\{\exp(-\alpha_t/2)\xi_t\}$ may fail to have finite unconditional means and variances. For some distributions of $\{z_t\}$ (e.g., the Student t with finite degrees of freedom), $\{\exp(-\alpha_t)\sigma_t^2\}$ and $\{\exp(-\alpha_t/2)\xi_t\}$ typically have no finite unconditional moments. The results for another commonly used family of distributions, the GED (Generalized Error Distribution (Harvey (1981), Box and Tiao (1973)[8]), are more encouraging. The GED includes the normal as a special case, along with many other distributions, some more fat tailed than the normal (e.g., the double exponential), some more thin tailed (e.g., the uniform). If the distribution of $\{z_t\}$ is a member of this family and is not too fat tailed, and if $\Sigma_{i=1,\infty}\beta_i^2 < \infty$, then $\{\sigma_t^2\}$ and $\{\xi_t\}$ have finite unconditional moments of arbitrary order.

The density of a GED random variable normalized to have a mean of zero and a variance of one is given by

$$f(z) = \frac{v \exp\left[-\left(\frac{1}{2}\right)|z/\lambda|^v\right]}{\lambda 2^{(1+1/v)}\Gamma(1/v)}, \qquad -\infty < z < \infty, \qquad 0 < v \leqslant \infty, \tag{2.4}$$

where $\Gamma(\cdot)$ is the gamma function, and

$$\lambda \equiv \left[2^{(-2/v)}\Gamma(1/v)/\Gamma(3/v)\right]^{1/2}. \tag{2.5}$$

[7] The assumption that z_t has a finite variance can be relaxed. For example, let z_t be i.i.d. Cauchy, and $g(z_t) \equiv \theta z_t + \gamma|z_t|$, with γ and θ not both equal to zero. By the Three Series Theorem (Billingsley (1986, Theorem 22.8)), σ_t^2 is finite almost surely if and only if $|\Sigma_{t \geqslant 1, \beta_t \neq 0}\gamma\beta_i \cdot \ln(1 + \beta_i^{-2})| < \infty$ and $\Sigma_{t \geqslant 1}|\beta_t| < \infty$. Then $\{\exp(-\alpha_t)\sigma_t^2\}$ and $\{\exp(-\alpha_t/2)\xi_t\}$ are strictly stationary and ergodic.

[8] Box and Tiao call the GED the exponential power distribution.

v is a tail-thickness parameter. When $v = 2$, z has a standard normal distribution. For $v < 2$, the distribution of z has thicker tails than the normal (e.g., when $v = 1$, z has a double exponential distribution) and for $v > 2$, the distribution of z has thinner tails than the normal (e.g., for $v = \infty$, z is uniformly distributed on the interval $[-3^{1/2}, 3^{1/2}]$).

THEOREM 2.2: *Define* $\{\sigma_t^2, \xi_t\}_{t=-\infty,\infty}$ *by* (1.1)–(1.2) *and* (2.1)–(2.2), *and assume that* γ *and* θ *do not both equal zero. Let* $\{z_t\}_{t=-\infty,\infty}$ *be i.i.d. GED with mean zero, variance one, and tail thickness parameter* $v > 1$, *and let* $\sum_{i=1}^{\infty}\beta_i^2 < \infty$. *Then* $\{\exp(-\alpha_t)\sigma_t^2\}$ *and* $\{\exp(-\alpha_t/2)\xi_t\}$ *possess finite, time-invariant moments of arbitrary order. Further, if* $0 < p < \infty$, *conditioning information at time* 0 *drops out of the forecast pth moments of* $\exp(-\alpha_t)\sigma_t^2$ *and* $\exp(-\alpha_t/2)\xi_t$ *as* $t \to \infty$:

$$\text{(2.6)} \qquad \operatorname*{plim}_{t\to\infty} E\left[\exp(-p\alpha_t)\sigma_t^{2p} \middle| z_0, z_{-1}, z_{-2}, \ldots\right] - E\left[\exp(-p\alpha_t)\sigma_t^{2p}\right] = 0, \quad \text{and}$$

$$\text{(2.7)} \qquad \operatorname*{plim}_{t\to\infty} E\left[\exp(-p\alpha_t/2)|\xi_t|^p \middle| z_0, z_{-1}, z_{-2}, \ldots\right] - E\left[\exp(-p\alpha_t)|\xi_t|^p\right] = 0,$$

where plim *denotes the limit in probability.*

That is, if the distribution of the z_t is GED and is thinner-tailed than the double exponential, and if $\sum\beta_i^2 < \infty$, then $\exp(-\alpha_t)\sigma_t^2$ and $\exp(\alpha_t/2)\xi_t$ are not only strictly stationary and ergodic, but have arbitrary finite moments, which in turn implies that they are covariance stationary.

Since the moments of $\{\exp(-\alpha_t)\sigma_t^2\}$ and $\{\exp(-\alpha_t/2)\xi_t\}$ are of interest for forecasting, Appendix I derives the conditional and unconditional moments (including covariances) of $\{\exp(-\alpha_t)\sigma_t^2\}$ and $\{\exp(-\alpha_t/2)\xi_t\}$ under a variety of distributional assumptions for $\{z_t\}$, including Normal, GED, and Student t.

3. A SIMPLE MODEL OF MARKET VOLATILITY

In this Section, we estimate and test a simple model of market risk, asset returns, and changing conditional volatility. We use this model to examine several issues previously investigated in the economics and finance literature, namely (i) the relation between the level of market risk and required return, (ii) the asymmetry between positive and negative returns in their effect on conditional variance, (iii) the persistence of shocks to volatility, (iv) "fat tails" in the conditional distribution of returns, and (v) the contribution of nontrading days to volatility.

We use the model developed in Section 2 for the conditional variance process, assuming an ARMA representation for $\ln(\sigma_t^2)$. To allow for the possibility of nonnormality in the conditional distribution of returns, we assume that the $\{z_t\}$ are i.i.d. draws from the GED density (2.4). To account for the

contribution of nontrading periods to market variance, we assume that each nontrading day contributes as much to variance as some fixed fraction of a trading day, so if, for example, this fraction is one tenth, than σ_t^2 on a typical Monday would be 20% higher than on a typical Tuesday. Other researchers (e.g., Fama (1965), and French and Roll (1986)) have found that nontrading periods contribute much less than do trading periods to market variance, so we expect that $0 < \delta \ll 1$.[9]

Specifically, we model the log of conditional variance as

$$(3.1) \qquad \ln(\sigma_t^2) = \alpha_t + \frac{(1 + \Psi_1 L + \cdots + \Psi_q L^q)}{(1 - \Delta_1 L + \cdots - \Delta_p L^p)} g(z_{t-1}),$$

where z_t is i.i.d. GED with mean zero, variance one, and tail thickness parameter $v > 0$, and $\{\alpha_t\}$ is given by

$$(3.2) \qquad \alpha_t = \alpha + \ln(1 + N_t \delta),$$

where N_t is the number of nontrading days between trading days $t-1$ and t, and α and δ are parameters. If the unconditional expectation of $\ln(\sigma_t^2)$ exists, then it equals $\alpha + \ln(1 + N_t\delta)$. Together, (3.1)–(3.2) and Theorem 2.1 imply that $\{(1 + N_t\delta)^{-1/2}\xi_t\}$ is strictly stationary and ergodic if and only if all the roots of $(1 - \Delta_1 Y - \cdots - \Delta_p Y^p)$ lie outside the unit circle.[10]

We model excess returns R_t as

$$(3.3) \qquad R_t = a + bR_{t-1} + c\sigma_t^2 + \xi_t,$$

where the conditional mean and variance of ξ_t at time t are 0 and σ_t^2 respectively, and where a, b, and c are parameters. The bR_{t-1} term allows for the autocorrelation induced by discontinuous trading in the stocks making up an index (Scholes and Williams (1977), Lo and MacKinlay (1988)). The Scholes and Williams model suggests an MA(1) form for index returns, while the Lo and MacKinlay model suggests an AR(1) form, which we adopt. As a practical matter, there is little difference between an AR(1) and an MA(1) when the AR and MA coefficients are small and the autocorrelations at lag one are equal, since the higher-order autocorrelations die out very quickly in the AR model. As Lo and MacKinlay note, however, such simple models do not adequately explain the short-term autocorrelation behavior of the market indices, and no fully satisfactory model yet exists.

The theoretical justification for including the $c\sigma_t^2$ term in (3.3) is meager, since the required excess return on a portfolio is linear in its conditional variance only under very special circumstances. In Merton's (1973) intertemporal CAPM model, for example, the instantaneous expected excess return on the market portfolio is linear in its conditional variance if there is a representative agent with log utility. Merton's conditions (e.g., continuous time, continuous

[9] French and Roll (1986) and Barclay, Litzenberger, and Warner (1990) offer economic interpretations.

[10] Again, we assume that $[1 - \Sigma_{t=1,p}\Delta_t y^t]$ and $[1 + \Sigma_{t=1,q}\Psi_t y^t]$ have no common roots.

trading, and a true "market" portfolio) do not apply in our model even under the log utility assumption. Backus and Gregory (1989), and Glosten, Jagannathan, and Runkle (1989) give examples of equilibrium models in which a regression of returns on σ_t^2 yields a negative coefficient.[11] There is, therefore, no strong theoretical reason to believe that c is positive. Rather, the justification for including $c\sigma_t^2$ is pragmatic: a number of researchers using GARCH models (e.g., French, Schwert, and Stambaugh (1987) and Chou (1987)) have found a statistically significant positive relation between conditional variance and excess returns on stock market indices, and we therefore adopt the form (3.3).

For a given ARMA(p, q) exponential ARCH model, the $\{z_t\}_{t=1,T}$ and $\{\sigma_t^2\}_{t=1,T}$ sequences can be easily derived recursively given the data $\{R_t\}_{t=1,T}$, and the initial values $\sigma_1^2, \ldots, \sigma_{1+\max\{p,q+1\}}^2$. To close the model, $\ln(\sigma_1^2), \ldots, \ln(\sigma_{1+\max\{p,q+1\}}^2)$ were set equal to their unconditional expectations $(\alpha + \ln(1+\delta N_1)), \ldots, (\alpha + \ln(1+\delta N_{1+\max\{p,q+1\}}))$. This allows us to write the log likelihood L_T as

$$(3.4) \qquad L_T = \sum_{t=1}^{T} \ln(v/\lambda) - \left(\tfrac{1}{2}\right)\left|\left(R_t - a - bR_{t-1} - c\sigma_t^2\right)/\sigma_t\lambda\right|^v - (1+v^{-1})\ln(2) - \ln\left[\Gamma(1/v)\right] - \tfrac{1}{2}\ln\left(\sigma_t^2\right),$$

where λ is defined in (2.5). Given the parameters and initial states, we can easily compute the likelihood (3.4) recursively, using (3.1) and setting

$$(3.5) \qquad z_t = \sigma_t^{-1}\left(R_t - a - bR_{t-1} - c\sigma_t^2\right).$$

In light of (3.5), however, it may be that setting the out-of-sample values of $\ln(\sigma_t^2)$ equal to their unconditional expectation is not innocent: using the true parameter values but the wrong σ_t^2 in (3.5) leads to an incorrect fitted value of z_t, in turn leading to an incorrect fitted value for σ_{t+1}^2, and so on. In simulations using parameter estimates similar to those reported below, fitted σ_t^2 generated by (3.4)–(3.5) with an incorrect starting value converge very rapidly to the σ_t^2 generated by (3.4)–(3.5) and a correct starting value. In a continuous time limit, it is possible to prove that this convergence takes place instantaneously (Nelson (1990b)).

Under sufficient regularity conditions, the maximum likelihood estimator is consistent and asymptotically normal. Unfortunately, verifying that these conditions hold in ARCH models has proved extremely difficult in both GARCH models and in the Exponential ARCH model introduced in this paper. Weiss (1986) developed a set of sufficient conditions for consistency and asymptotic normality in a variant of the linear ARCH formulation of Engle (1982). These conditions are quite restrictive, and are not satisfied by the coefficient estimates obtained in most studies using this form of ARCH. In the GARCH-M model, in which conditional variance appears in the conditional mean, the asymptotics are

[11] See also Gennotte and Marsh (1987).

even more uncertain, and no sufficient conditions for consistency and asymptotic normality are yet known. The asymptotics of exponential ARCH models are equally difficult, and as with other ARCH models, a satisfactory asymptotic theory for exponential ARCH is as yet unavailable. In the remainder of this paper we assume (as is the usual practice of researchers using GARCH models) that the maximum likelihood estimator is consistent and asymptotically normal.

For our empirical analysis, we use the daily returns for the value-weighted market index from the CRSP tapes for July 1962–December 1987. An immediate problem in using this data is that we wish to model the excess returns process but do not have access to any adequate daily riskless returns series. As an initial approximation to the riskless rate, we extracted the monthly Treasury bill returns from the CRSP tapes, assumed that this return was constant for each calendar day within a given month, and computed daily excess returns using this riskless rate series and value-weighted CRSP daily market returns.[12] As a check on whether measurement errors in the riskless rate series are likely to bias the results seriously, we also fit the model using the capital gains series, ignoring both dividends and the riskless interest rate. As shown below, it made virtually no difference in either the estimated parameters or the fitted variances.

To select the order of the ARMA process for $\ln(\sigma_t^2)$, we used the Schwarz Criterion (Schwarz (1978)), which provides consistent order-estimation in the context of linear ARMA models (Hannan (1980)). The asymptotic properties of the Schwarz criterion in the context of ARCH models are unknown.

The maximum likelihood parameter estimates were computed on VAX 8650 and 8550 computers using the IMSL subroutine DUMING. Table I lists likelihood values for ARMA models of various orders on the CRSP excess returns series. For both the excess returns and capital gains series, the Schwarz Criterion selected an ARMA(2, 1) model for $\ln(\sigma_t^2)$.[13] Table II gives the parameter estimates and estimated standard errors for both ARMA(2, 1) models. The estimated correlation matrix of the parameter estimates in the excess returns model is in Table III. The asymptotic covariance matrix was computed using the score.

First note that except for the parameter c (the risk premium term in (3.3)), the two sets of coefficient estimates are nearly identical. The fitted values of $\ln(\sigma_t^2)$ from the two models are even more closely related: their means and variances for the 1962–1987 period are nearly equal (-9.9731 and 0.6441, vs. -9.9766 and 0.6415, respectively) and the two series have a sample correlation of 0.9996. In other words, the series are practically identical, so ignoring dividends and interest payments appears likely to introduce no important errors in forecasting the volatility of broad market indices.

[12] Logarithmic returns are used throughout: i.e., if S_t is the level of the value-weighted market index at time t and d_t are the dividends paid at t, then the value-weighted market return, capital gain, and excess return are computed as $\ln[(S_t + d_t)/S_{t-1}]$, $\ln[S_t/S_{t-1}]$, and $\ln[(S_t + d_t)/S_{t-1}] - RR_t$ respectively, where RR_t is our proxy for the riskless rate.

[13] The AIC (Akaike (1973)) chose the highest-order model estimated.

TABLE I

LIKELIHOOD VALUES FOR ARMA MODELS FOR CRSP VALUE-WEIGHTED EXCESS RETURNS

Observations = 6408.

Deterministic Conditional Variance Model (i.e., $\gamma = \theta = \Delta = \Psi = 0$), Likelihood = 22273.313.

ARMA Exponential ARCH Model Likelihoods:

		AR Order				
		0	1	2	3	4
MA Order	0	22384.898	22888.052	22891.937	22894.237	22894.902
	1	22429.942	22893.687	22915.454 (SC)	22916.799	22916.894
	2	22473.728	22894.167	22916.762	22917.035	22918.708
	3	22532.768	22894.385	22916.941	22918.853	22922.439
	4	22590.536	22900.990	22917.670	22918.857	22923.752 (AIC)

(SC = Model selected by the criterion of Schwarz (1978).)
(AIC = Model selected by the information criterion of Akaike (1973).)

TABLE II

PARAMETER ESTIMATES FOR THE CRSP EXCESS RETURNS AND CAPITAL GAINS MODELS
(Standard Errors in Parentheses)

Parameter	CRSP Excess Returns	CRSP Capital Gains
α	−10.0593	−10.0746
	(0.3462)	(0.3361)
δ	0.1831	0.1676
	(0.0277)	(0.0271)
γ	0.1559	0.1575
	(0.0125)	(0.0126)
Δ_1	1.92938	1.92914
	(0.0145)	(0.0146)
Δ_2	−0.92941	−0.92917
	(0.0145)	(0.0146)
Ψ	−0.9782	−0.9781
	(0.0062)	(0.0063)
θ	−0.1178	−0.1161
	(0.0090)	(0.0090)
a	$3.488 \cdot 10^{-4}$	$3.416 \cdot 10^{-4}$
	$(9.850 \cdot 10^{-5})$	$(9.842 \cdot 10^{-5})$
b	0.2053	0.2082
	(0.0123)	(0.0123)
c	−3.3608	−1.9992
	(2.0261)	(2.0347)
v	1.5763	1.5760
	(0.0320)	(0.0320)

TABLE III

ESTIMATED CORRELATION MATRIX FOR PARAMETER ESTIMATES
EXCESS RETURNS MODEL
(Only lower triangle reported)

Parameter					
α	1				
δ	−0.0087	1			
γ	−0.0023	−0.1639	1		
Δ_1	0.0139	0.0715	−0.2996	1	
Δ_2	−0.0138	−0.0719	0.2994	−0.999992	1
Ψ	−0.0094	−0.0392	0.1020	−0.8356	0.8352
	1				
θ	−0.0472	0.1909	−0.0591	0.4342	−0.4346
	−0.2211	1			
a	−0.0674	−0.0342	−0.0222	0.0734	−0.0715
	0.0064	0.0251	1		
b	−0.0144	−0.0395	−0.0109	0.0439	−0.0443
	−0.0256	−0.0772	−0.1979	1	
c	−0.0702	0.1111	0.0815	−0.0891	0.0871
	−0.0151	−0.0064	−0.8287	0.1198	1
v	0.0591	0.1763	0.2853	−0.0182	0.0179
	−0.0414	−0.1554	0.0042	0.0402	0.1240
	1				
	α	δ	γ	Δ_1	Δ_2
	Ψ	θ	a	b	c
	v				

Next we examine the empirical issues raised earlier in the Section:

(i) Market Risk and Expected Return: The estimated risk premium is negatively (though weakly) correlated with conditional variance, with $c \approx -3.361$ with a large standard error of about 2.026. This contrasts with the significant positive relation between returns and conditional variance found by researchers using GARCH-M models (e.g., Chou (1987), and French, Schwert, and Stambaugh (1987)), but agrees with the findings of other researchers not using GARCH models (e.g., Pagan and Hong (1988)). Given the results of Gennotte and Marsh (1987), Glosten, Jagannathan, and Runkle (1989), and Backus and Gregory (1989), our findings of a negative (albeit insignificant) coefficient should not be too surprising.

(ii) The asymmetric relation between returns and changes in volatility, as represented by θ, is highly significant. Recall that a negative value of θ indicates that volatility tends to rise (fall) when returns surprises are negative (positive). The estimated value for θ is about -0.118 (with a standard error of about 0.008) which is significantly below zero at any standard level. Figure 1 plots the estimated $g(z)$ function.

Figures 2 and 3 plot σ_t (the daily conditional standard deviation of returns) and the log value of the CRSP value-weighted market index, respectively. The σ_t series is extremely variable, with lows of less than one-half of a percent and highs over five percent. All the major episodes of high volatility are associated with market drops.

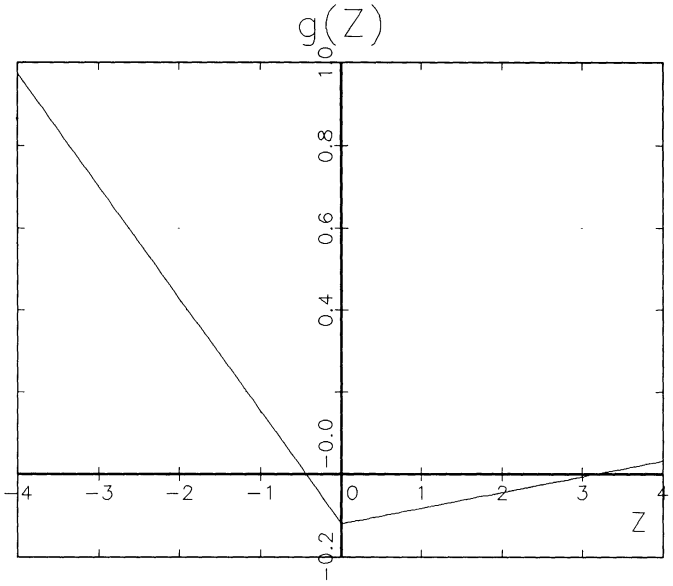

FIGURE 1

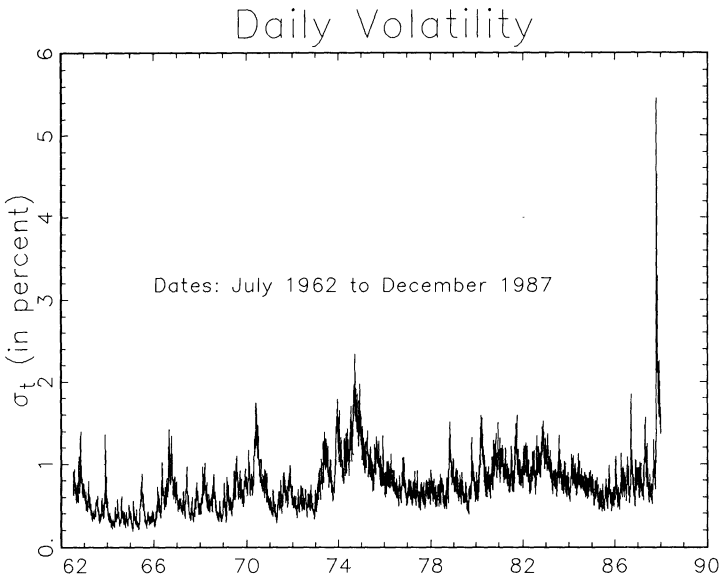

FIGURE 2

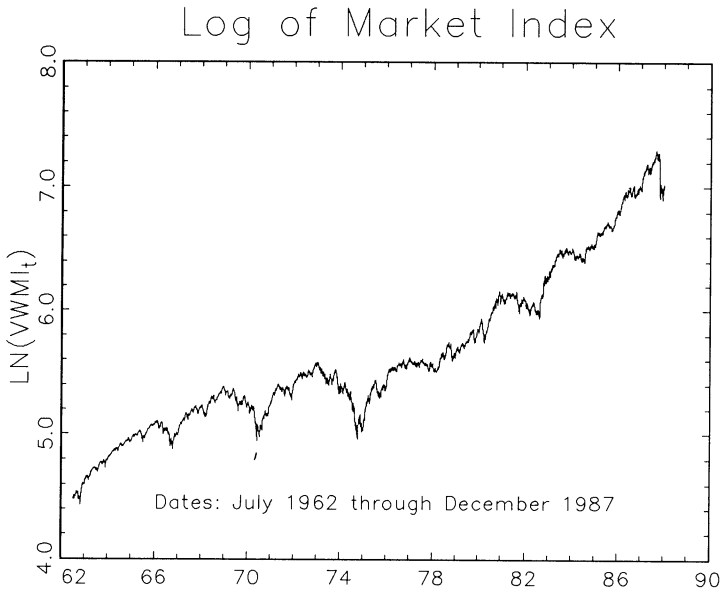

FIGURE 3

(iii) Persistence of Shocks: The largest estimated AR root is approximately 0.99962 with a standard error of about 0.00086, so the t statistic for a unit root is only about -0.448. To gain intuition about the degree of persistence implied by the largest AR root ρ, it is useful to think of the half-life h of a shock associated with this root, i.e., the number h such that

$$\rho^h = 1/2. \tag{3.6}$$

$\rho = 0.99962$ implies a half-life h of over 1820 trading days, about 7.3 years. In contrast, the half-life implied by the smaller AR root is less than two weeks. While this indicates substantial persistence and perhaps nonstationarity, it is hard to know how seriously to take the point estimates, since we have only about 25 years of data, about four times our estimated half-life for the larger AR root. The usual cautions about interpreting an estimated AR root near the unit circle as evidence of truly infinite persistence also apply.[14]

(iv) Fat Tails: It is well known that the distribution of stock returns has more weight in the tails than the normal distribution (e.g., Mandelbrot (1963), Fama (1965)), and that a stochastic process is thick tailed if it is conditionally normal with a randomly changing conditional variance (Clark (1973)). Our estimated model generates thick tails with both a randomly changing conditional variance σ_t^2 and a thick-tailed conditional distribution for ξ_t. Recall from the discussion

[14] See, e.g., Cochrane (1988). It is also unclear what the effect of a unit root in $\ln(\sigma_t^2)$ is on the asymptotic properties of the parameter estimates. The (unverified) regularity conditions for asymptotic normality require that the scoring function and hessian obey a central limit theorem and uniform weak law of large numbers respectively, which may or may not require $\{\sigma_t^2, \xi_t\}$ to have finite moments. It may be that the standard asymptotics are valid even in the presence of a unit root in $\ln(\sigma_t^2)$.

of the GED(v) distribution in Section 2 that if $v < 2$, the distribution of z_t (and therefore the conditional distribution of ξ_t) has thicker tails than the normal distribution. The estimated v is approximately 1.58 with a standard error of about 0.03, so the distribution of the z_t is significantly thicker-tailed than the normal.

(v) The estimated contribution of nontrading days to conditional variance is roughly consistent with the results of French and Roll (1986). The estimated value of δ is about 0.183, with a standard error of about 0.028, so a nontrading day contributes less than a fifth as much to volatility as a trading day.

The general results just discussed are quite robust to which ARMA model is selected, though of course the parameter estimates change somewhat. The results also appear to be quite robust with respect to the sample period: another paper, Nelson (1989), reports strikingly similar parameter estimates in an exponential ARCH model fit to daily capital gains on the Standard 90 stock index from 1928 to 1956.

Specification Tests

To test the fit of the model, several conditional moment tests (Newey (1985)) were fit using orthogonality conditions implied by correct specification. Correct specification of the model has implications for the distribution of $\{z_t\}$. For example, $E[z_t] = 0$, $E[z_t^2] = 1$, and $E[g(z_t)] = 0$. Since the GED distribution is symmetric, we also require that $E[z_t \cdot |z_t|] = 0$. The first four orthogonality conditions test these basic properties. Correct specification also requires that $\{\xi_t^2 - \sigma_t^2\}$ and $\{\xi_t\}$ (or equivalently $\{z_t^2 - 1\}$ and $\{z_t\}$) are serially uncorrelated. Accordingly, we test for serial correlation in z_t and z_t^2 at lags one through five.

Table IV reports first the sample averages for the fourteen selected orthogonality conditions and their associated t statistics, and then chi-square statistics and probability values for two combinations of the orthogonality conditions, the first including only conditions relating to correct specification of the conditional variance process σ_t^2 and the second also testing for correct specification of the conditional mean process.

In the first chi-square test, the CRSP model does extremely well, with a probability value of 0.94. Considered individually, none of the first nine orthogonality conditions are significantly different from zero at any standard significance level. When the last five conditions, which test for serial correlation in $\{z_t\}$, are included, the probability value drops to 0.16, which still does not reject at any standard significance level, although statistically significant serial correlation is found at lag two. Overall, the fit of the CRSP model of the conditional variance process $\{\sigma_t^2\}$ seems remarkably good.[15,16]

[15] Engle, Lilien, and Robbins (1987) and Pagan and Sabau (1987) based conditional moment tests on $\{\xi_t\}$ rather than on $\{z_t\}$. Basing tests on $\{z_t\}$ is analogous to a GLS correction and seems likely to increase the power of the specification tests. As a check, chi-square tests were recomputed using $\{\xi_t\}$ instead of $\{z_t\}$. The test statistics were drastically lower in each instance.

[16] The specification tests for the 1928–1956 Standard 90 capital gains data reported in Nelson (1989) were not as favorable: the tests found evidence of negatively skewed returns and serially correlated residuals, rejecting the model at any standard level.

TABLE IV

SPECIFICATION TEST RESULTS FOR CRSP VALUE-WEIGHTED EXCESS RETURNS, ARMA(2, 1) MODEL[a]

	Orthogonality Conditions	Sample Averages	t Statistics
1	$E(z_t) = 0$	-0.007	-0.349
2	$E(z_t^2) - 1 = 0$	-0.001	-0.037
3	$E\lvert z_t\rvert - \lambda 2^{1/v}\Gamma(2/v)/\Gamma(1/v) = 0$	$2.583 \cdot 10^{-5}$	0.002
4	$E[z_t \cdot (\lvert z_t\rvert - E\lvert z\rvert)] = 0$	-0.027	-1.254
5	$E[(z_t^2 - 1)(z_{t-1}^2 - 1)] = 0$	0.102	1.023
6	$E[(z_t^2 - 1)(z_{t-2}^2 - 1)] = 0$	0.028	0.529
7	$E[(z_t^2 - 1)(z_{t-3}^2 - 1)] = 0$	0.079	0.864
8	$E[(z_t^2 - 1)(z_{t-4}^2 - 1)] = 0$	0.042	0.947
9	$E[(z_t^2 - 1)(z_{t-5}^2 - 1)] = 0$	0.019	0.469
10	$E(z_t \cdot z_{t-1}) = 0$	0.022	0.949
11	$E(z_t \cdot z_{t-2}) = 0$	-0.034	-2.521
12	$E(z_t \cdot z_{t-3}) = 0$	0.018	1.377
13	$E(z_t \cdot z_{t-4}) = 0$	0.015	1.174
14	$E(z_t \cdot z_{t-5}) = 0$	0.020	1.563

[a] χ^2 statistic for conditional moment test using the first 9 orthogonality conditions = 3.46. With 9 degrees of freedom, the probability value = 0.94. χ^2 statistic for conditional moment test using all 14 orthogonality conditions = 19.14. With 14 degrees of freedom, the probability value = 0.16.

Our conditional moment tests leave many potential sources of misspecification unchecked. It therefore seems desirable to check the forecasting performance of the model during periods of rapidly changing volatility, and to check for large outliers in the data. From Figure 2, two periods stand out as times of high and rapidly changing volatility: the market break of September, 1973–December, 1974, and the last five months of 1987. Figures 4 and 5 plot returns and the one-day-ahead ex ante 99% prediction intervals implied by the estimated model for these periods. The 99% prediction intervals have a width of approximately $5.56 \cdot \sigma_t$.

In 1973–1974, the model seems to track the change in volatility quite closely, with no serious outliers (Figure 4). The model also succeeds quite well in picking up the volatility of the period *after* October 19, 1987, with no serious outliers (Figure 5). On October 19, 1987, however, the model's performance is mixed at best: the ex ante prediction intervals for the day are approximately $\pm 7\%$, the widest in the data set up to that time, brought on by the sharp drops in the market during the preceding week. Unfortunately, the drop in the index that day was approximately 20.25%, about $7.78 \cdot \sigma_t$, a serious outlier. If the estimated model were literally true, the expected number of outliers of this size or greater in a $25\frac{1}{2}$ year period (the length of the data set) is only about $1 \cdot 10^{-5}$ —i.e., the probability of observing such an outlier in a data set of this length is extremely low.

Although October 19, 1987 is the most extreme outlier in the sample, it is not the only large outlier. Table V lists the largest, ranked in order of the implied

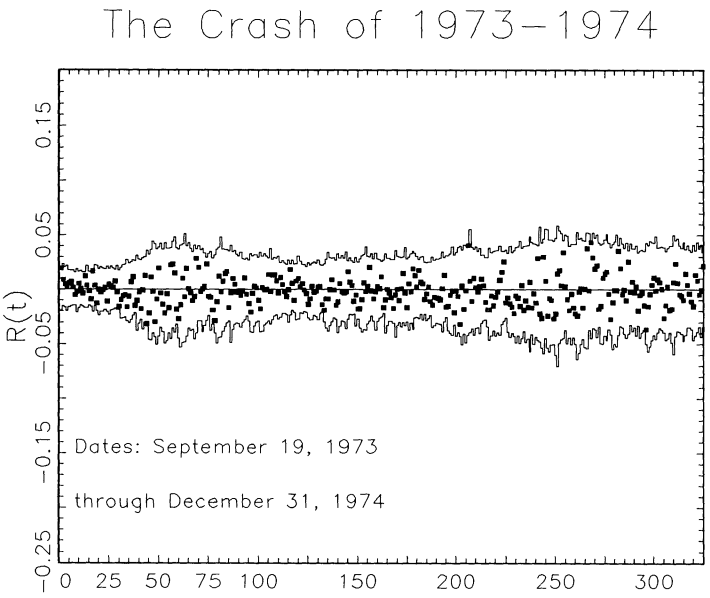

FIGURE 4

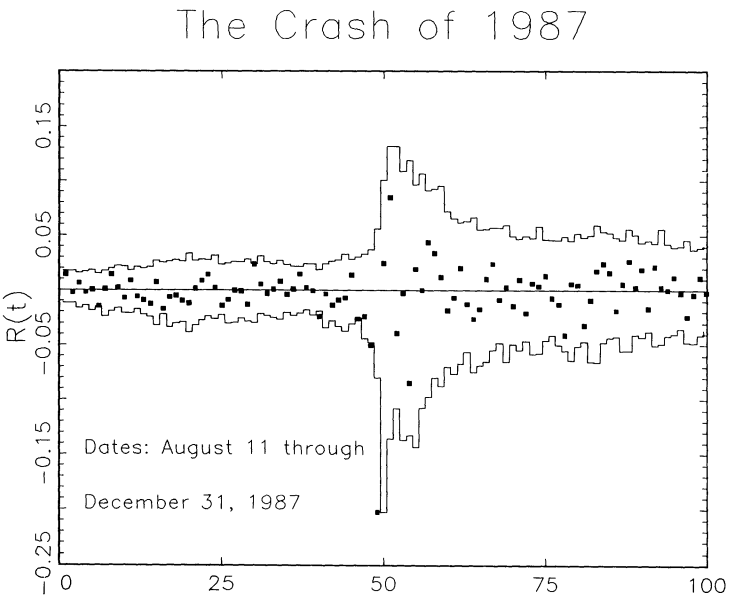

FIGURE 5

TABLE V

LARGEST OUTLIERS IN THE SAMPLE

Date	$R_t(\%)$	$\sigma_t(\%)$	z_t	Expected Frequency[a]
10/19/87	−20.25	2.44	−7.78	$1.11 \cdot 10^{-5}$
9/11/87	−4.58	0.84	−5.38	0.05
8/17/82	4.36	0.87	4.95	0.20
1/8/86	−2.42	0.57	−4.81	0.31
8/2/78	2.23	0.48	4.53	0.67
8/16/71	3.50	0.82	4.33	1.21
10/9/79	−3.37	0.73	−4.29	1.36
7/7/86	−3.02	0.72	−4.12	2.20

[a]The expected number of $|z_t|$ values of this size or greater in a $25\frac{1}{2}$ year sample (6408 observations).

value for $|z_t|$. The standardized GED has only one parameter, v, to control the shape of the conditional distribution, and this may well not be flexible enough—i.e., there are too many "large" $|z_t|$ values. Nonparametric methods (as in Engle and González-Rivera (1989)), or more flexible parametric families of distributions, would probably improve the model.

4. CONCLUSION

This paper has presented a new class of ARCH models that do not suffer from some of the drawbacks of GARCH models. Ideally, we would like ARCH models that allow the same degree of simplicity and flexibility in representing conditional variances as ARIMA and related models have allowed in representing conditional means. While this paper has made a contribution to this end, the goal is far from accomplished: it remains to develop a multivariate version of exponential ARCH, and a satisfactory asymptotic theory for the maximum likelihood parameter estimates. These tasks await further research.

Graduate School of Business, University of Chicago, 1101 E. 58th St., Chicago, IL 60637, U.S.A.

Manuscript received February, 1989; final revision received February, 1990.

APPENDIX I: THE MOMENTS OF σ_t^2 AND ξ_t

By (2.1) and the independence of the $\{z_t\}$, the joint moments and conditional moments of $\{\sigma_t^2\}$ and $\{\xi_t\}$ take either the form

$$\text{(A1.1)} \qquad E\left[\exp\left(a_t + \sum_{i=1}^{\infty} b_i g(z_{t-i})\right)\right] = \exp(a_t) \prod_{i=1}^{\infty} E[\exp(b_i g(z_{t-i}))] \quad \text{or}$$

$$\text{(A1.2)} \qquad E\left[z_{t-k}^p z_{t-j}^q \exp\left(a_t + \sum_{i=1}^{\infty} b_i g(z_{t-i})\right)\right] = \left[\prod_{i \geq 1,\, i \neq j,k} E[\exp(b_i g(z_{t-i}))]\right] \cdot \exp(a_t) \cdot E[z_{t-k}^p \exp(b_k g(z_{t-k}))] \cdot E[z_{t-j}^q \exp(b_j g(z_{t-j}))]$$

for nonnegative integers p, q, j, and $k \neq j$. For example, to get the unconditional expectation of σ_t^2, we set $\{a_t\} = \{\alpha_t\}$ and $\{b_t\} = \{\beta_t\}$ in (A1.1). For the conditional expectation of σ_t^2 given $z_{t-k}, z_{t-k-1}, \ldots$, set

$$a_t = \alpha_t + \sum_{i=k}^{\infty} \beta_i g(z_{t-i}), \qquad b_i = \beta_i \text{ for } 1 \leqslant i \leqslant k-1, \text{ and } b_i = 0 \text{ for } i \geqslant k.$$

To obtain the moments of ξ_t and $\xi_{t-k}\sigma_t^2$, we proceed similarly, using (A1.2).

To evaluate the expectations in (A1.1)–(A1.2), we must make a further distributional assumption about $\{z_t\}$. When $\{z_t\} \sim$ i.i.d. $N(0,1)$, the following result, combined with (A1.1)–(A1.2), gives the joint conditional and unconditional moments of the σ_t^2 and ξ_t processes:

THEOREM A1.1: *Let $z \sim N(0,1)$. For any finite, real scalar b and positive integer p,*

$$\text{(A1.3)} \qquad E[\exp(g(z)b)] = \left\{\Phi(\gamma b + \theta b)\exp\left[b^2(\theta+\gamma)^2/2\right] + \Phi(\gamma b - \theta b)\exp\left[b^2(\gamma-\theta)^2/2\right]\right\}\exp\left[-b\gamma(2/\pi)^{1/2}\right] < \infty,$$

and

$$\text{(A1.4)} \qquad E[z^p \exp(g(z)b)] = \exp\left[-b\gamma(2/\pi)^{1/2}\right] \cdot \Gamma(p+1) \cdot (2\pi)^{-1/2} \cdot \left\{\exp\left[b^2(\theta+\gamma)^2/4\right] \cdot D_{-(p+1)}[-b(\gamma+\theta)] + (-1)^p \exp\left[b^2(\gamma-\theta)^2/4\right] \cdot D_{-(p+1)}[-b(\gamma-\theta)]\right\} < \infty,$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\Gamma[\cdot]$ is the Gamma function, and $D_q[\cdot]$ is the parabolic cylinder function (Gradshteyn and Ryzhik (1980)).

PROOF OF THEOREM A1.1: (A1.3) follows by straightforward but tedious calculus. (A1.4) is easily proven with the help of Gradshteyn and Ryzhik (1980) Formula 3.462 #1. The finiteness of the expression in (A1.4) follows as a special case in the proof of Theorem A1.2 below. *Q.E.D.*

Theorem A1.2 deals with the more general GED case.

THEOREM A1.2: *Let p be a nonnegative integer, and let $z \sim \mathrm{GED}(v)$ with $E(z) = 0$, $\mathrm{Var}(z) = 1$, and $v > 1$. Then*

$$\text{(A1.5)} \qquad E[z^p \exp(g(z)b)] = \exp\left[-b\gamma\Gamma(2/v)\lambda 2^{1/v}/\Gamma(1/v)\right] \cdot 2^{p/v} \cdot \lambda^p \cdot \sum_{k=0}^{\infty} (2^{1/v}\lambda b)^k \left[(\gamma+\theta)^k + (-1)^p(\gamma-\theta)^k\right] \frac{\Gamma[(p+k+1)/v]}{2\Gamma(1/v)\Gamma(k+1)} < \infty.$$

If $z \sim \mathrm{GED}(v)$ with $v < 1$, or $z \sim$ Student t with d degrees of freedom ($d > 2$), and z is normalized to satisfy $E(z) = 0$, $\mathrm{Var}(z) = 1$, then $E[\exp(g(z)b)]$ and $E[z^p \exp(g(z)b)]$ are finite if and only if

$$\text{(A1.6)} \qquad b\gamma + |b\theta| \leqslant 0.$$

If $z \sim \mathrm{GED}(1)$, then $E[\exp(g(z)b)]$ and $E[z^p \exp(g(z)b)]$ are finite if and only if

$$\text{(A1.7)} \qquad b\gamma + |b\theta| < 2^{1/2}.$$

The restriction (A1.6) is rarely satisfied in practice. In computing the unconditional expectation of σ_t^2, b is one of the moving average coefficients $\{\beta_i\}$, at least some of which are positive, since $\beta_0 \equiv 1$. If $b > 0$, (A1.6) implies either $\gamma < 0$ or $\gamma = \theta = 0$. If $\gamma < 0$, residuals larger than expected decrease conditional variance, which goes against the intuition developed in Section 1. In the author's experience in fitting exponential ARCH models, the estimated value of γ is always positive.

PROOF OF THEOREM A1.2: The density of z given in (2.4) and Gradshteyn and Ryzhik (1980) Formula 3.381 #4 yield

$$E(|z|) = \lambda 2^{1/\iota}\Gamma(2/\upsilon)/\Gamma(1/\upsilon). \tag{A1.8}$$

Straightforward calculus then yields

$$\begin{aligned} E[z^p \cdot \exp(g(z)b)] &= \frac{\lambda^p 2^{p/\iota}}{2\Gamma(1/\upsilon)} \exp\left[-b\gamma\lambda 2^{1/\iota}\Gamma(2/\upsilon)/\Gamma(1/\upsilon)\right] \\ &\quad \cdot \int_0^\infty y^{(-1+(p+1)/\iota)} e^{-y}\left[e^{b\gamma(\theta+\gamma)(2y)^{1/\iota}} + (-1)^p e^{b\lambda(\gamma-\theta)(2y)^{1/\iota}}\right] dy. \end{aligned} \tag{A1.9}$$

Expanding the part of the integrand in square brackets in a Taylor series,

$$\begin{aligned} E[z^p \cdot \exp(g(z)b)] &= \frac{2^{p/\upsilon}\lambda^p}{2\Gamma(1/\upsilon)} \exp\left[-b\lambda 2^{1/\iota}\Gamma(2/\upsilon)/\Gamma(1/\upsilon)\right] \\ &\quad \cdot \int_0^\infty \sum_{k=0}^\infty \left[\left[(\theta+\gamma)2^{1/\iota}b\lambda\right]^k + (-1)^p\left[(\gamma-\theta)2^{1/\iota}b\lambda\right]^k\right] \frac{e^{-y}y^{-1+(p+1+k)/\upsilon}}{\Gamma(k+1)}\, dy. \end{aligned} \tag{A1.10}$$

If we can interchange the order of summation and integration in (A1.10), then Gradshteyn and Ryzhik (1980) Formula 3.381 #4 yields (A1.5). First, consider the related expression

$$\int_0^\infty \sum_{k=0}^\infty 2\Delta^k \frac{e^{-y}y^{-1+(p+1+k)/\iota}}{\Gamma(k+1)}\, dy, \qquad \text{where} \tag{A1.11}$$

$$\Delta \equiv \max\left\{\left|(\theta+\gamma)b\lambda 2^{1/\upsilon}\right|, \left|(\gamma-\theta)b\lambda 2^{1/\upsilon}\right|\right\}. \tag{A1.12}$$

The terms in (A.11) are nonnegative, so by monotone convergence (Rudin (1976, Theorem 11.28)), the order of integration and summation in (A1.11) can be reversed. If the integral in (A1.11) is finite, then by dominated convergence (Rudin (1976, Theorem 11.32)), we can interchange the order of summation and integration in (A1.10) and the integral in (A1.10) is finite. To prove (A1.5), therefore, it remains only to show that (A1.11) is finite if $\upsilon > 1$. By Gradshteyn and Ryzhik (1980) Formula 3.381 #4,

$$\int_0^\infty \sum_{k=0}^\infty 2\Delta^k \frac{e^{-y}y^{-1+(p+1+k)/\upsilon}}{\Gamma(k+1)}\, dy = \sum_{k=0}^\infty \frac{2\Delta^k\Gamma[(p+1+k)/\upsilon]}{\Gamma(k+1)}. \tag{A1.13}$$

By construction, $\Delta \geqslant 0$. When $\Delta = 0$, the convergence of (A1.13) is trivial. When $\Delta > 0$, then by the root test (Rudin (1976, Theorem 3.33)), the sum in (A1.13) converges if

$$\limsup_{k\to\infty} \ln(\Delta) + k^{-1}\ln(\Gamma[(k+1+p)/\upsilon]) - k^{-1}\ln(\Gamma(k+1)) < 0. \tag{A1.14}$$

By the asymptotic expansion for $\ln(\Gamma(x))$ in Davis (1965, Equation 6.1.41),

$$\begin{aligned} &\limsup_{k\to\infty} \ln(\Delta) + k^{-1}\ln(\Gamma[(k+1+p)/\upsilon]) - k^{-1}\ln(\Gamma(k+1)) \\ &\quad = \limsup_{k\to\infty} \ln(\Delta) + k^{-1}[(k+1+p)/\upsilon - 1/2]\cdot\ln(k+1+p) \\ &\qquad + (k+1)/k - k^{-1}[(k+1+p)/\upsilon - 1/2]\cdot\ln(\upsilon) \\ &\qquad - k^{-1}[k+1/2]\cdot\ln(k+1) - (k+p+1)/k\upsilon + O(k^{-1}). \end{aligned} \tag{A1.15}$$

Expanding $\ln(k+1+p)$ in a Taylor series around $p=0$ and substituting into (A1.15),

$$\text{(A1.16)}\quad \begin{aligned}&\limsup_{k\to\infty} \ln(\Delta) + k^{-1}\ln(\Gamma[(k+1+p)/v]) - k^{-1}\ln(\Gamma(k+1))\\ &\quad = \limsup_{k\to\infty} \ln(\Delta) - k^{-1}[(k+1+p)/v - 1/2]\cdot\ln(v) - (k+p+1)/kv\\ &\qquad + 1 + O(k^{-1}) - \left[(k+1+p)/kv - 1 - k^{-1}\right]\cdot\ln(k+1).\end{aligned}$$

The last term on the right-hand side of (A1.16) asymptotically dominates every other term if $v \neq 1$, diverging to $-\infty$ if $v > 1$, so the series converges when $v > 1$ and $\Delta < \infty$.

If $z \sim$ Student t, or GED with $v < 1$, it is easy to verify that $E[\exp(b\cdot g(z))] = \infty$ unless $(\gamma+\theta)b \leqslant 0$ and $(\gamma-\theta)b \leqslant 0$, which holds if and only if $b\gamma + |\theta b| \leqslant 0$. If this inequality is satisfied, $E[\exp(g(z)b] < \infty$. The GED case with $v = 1$ is similar. *Q.E.D.*

APPENDIX II: PROOFS

PROOF OF THEOREM 2.1: That $|\ln(\sigma_t^2 - \alpha_t|$ is finite almost surely when $\Sigma_{k=1}^{\infty}\beta_k^2 < \infty$ follows immediately from the independence and finite variance of the $g(z_t)$ terms in (2.1) and from Billingsley (1986, Theorem 22.6). Since $|\ln(\sigma_t^2) - \alpha_t|$ is finite almost surely, so are $\exp(-\alpha_t)\sigma_t^2$ and $\exp(-\alpha_t/2)\xi_t$. This, combined with Stout (1974, Theorem 3.5.8) and the representation in (1.1)–(1.2) and (2.1)–(2.2) implies that these series are strictly stationary and ergodic. For all t, the expectation of $(\ln(\sigma_t^2) - \alpha_t) = 0$, and the variance of $(\ln(\sigma_t^2) - \alpha_t)$ is $\mathrm{Var}(g(z_t))\Sigma_{k=1}^{\infty}\beta_k^2$. Since $\mathrm{Var}(g(z_t))$ is finite and the distribution of $\ln(\sigma_t^2) - \alpha_t$ is independent of t, the first two moments of $(\ln(\sigma_t^2) - \alpha_t)$ are finite and time invariant, so $\{\ln(\sigma_t^2) - \alpha_t\}$ is covariance stationary.

If $\Sigma_{k=1}^{\infty}\beta_k^2 = \infty$, then $|\ln(\sigma_t^2) - \alpha_t| = \infty$ almost surely by Billingsley (1986, Theorems 22.3 and 22.8). *Q.E.D.*

PROOF OF THEOREM 2.2: By Theorem 2.1, the distributions of $\exp(-\alpha_t)\sigma_t^2$ and $\exp(-\alpha_t/2)\xi_t$ and any existing moments are time invariant. We will show that $\exp(-\alpha_t)\sigma_t^2$ and $\exp(-\alpha_t/2)\xi_t$ have finite moments of arbitrary positive order. As shown in Appendix I, the conditional, unconditional, and cross moments of $\exp(-\alpha_t)\sigma_t^2$ and $\exp(-\alpha_t/2)\xi_t$ have the form (A1.1)–(A1.2). By Hölder's Inequality, if σ_t^2 and ξ_t have arbitrary finite moments, the cross moments are also finite. By the independence of the z_t, $E[|(\exp(-\alpha_t/2)\xi_t)|^d] = E|z|^d E[(\exp(-\alpha_t/2)\sigma_t)^d]$. Since z has arbitrary finite moments, we need only show that $E[(\exp(-\alpha_t/2)\sigma_t)^d]$ is finite for all $d > 0$ if $\{\beta_i\}$ is square-summable. This expectation is given by

$$\text{(A2.1)}\quad E\left[\left(\exp(-\alpha_t/2)\sigma_t\right)^d\right] = \prod_{t=1}^{\infty} E\exp\left[\tfrac{1}{2}d\beta_t g(z_{t-t})\right],$$

where the individual expectation terms in (A2.1) are obtained by setting $b = \beta_t d/2$ in Theorem A1.1.

A sufficient condition for an infinite product $\Pi_{t=1,\infty}a_t$ to converge to a finite, nonzero number is that the series $\Sigma_{t=1,\infty}|a_t - 1|$ converge (Gradshteyn and Ryzhik (1980, Section 0.25)). Let a_t equal the ith term in (A2.1). Define

$$\text{(A2.2)}\quad S(\beta) \equiv \exp\left[\tfrac{1}{2}d\beta\gamma\lambda\Gamma(2/v)2^{1/v}/\Gamma(1/v)\right]\cdot E\left(\exp\left[g(z)\beta\tfrac{1}{2}d\right]\right).$$

We then have

$$\text{(A2.3)}\quad \begin{aligned}&S(0) = 1, \qquad S'(0) = \tfrac{1}{2}d\gamma\lambda\Gamma(2/v)2^{1/v}/\Gamma(1/v), \qquad \text{and}\\ &S''(\beta) = O(1) \quad \text{as} \quad \beta \to 0.\end{aligned}$$

Expanding $S(\beta)$ and $\exp[-\tfrac{1}{2}d\beta\gamma\lambda\Gamma(2/v)2^{1/v}/\Gamma(1/v)]$ in Taylor series around $\beta = 0$ and substituting into (A1.5), we have

$$\text{(A2.4)}\quad \begin{aligned}a_t - 1 &= E\left(\exp\left[g(z)\beta_t\tfrac{1}{2}d\right]\right) - 1\\ &= \left[1 - \tfrac{1}{2}d\beta_t\gamma\lambda\Gamma(2/v)2^{1/v}/\Gamma(1/v) + O(\beta_t^2)\right]\\ &\quad\cdot\left[1 + \tfrac{1}{2}d\beta_t\gamma\lambda\Gamma(2/v)2^{1/v}/\Gamma(1/v) + O(\beta_t^2)\right] - 1\\ &= O(\beta_t^2) \quad \text{as} \quad \beta_t \to 0.\end{aligned}$$

$O(\beta_i^2)$ in (A2.4) means that for some $\varepsilon > 0$, there exists a finite M independent of i such that

$$\text{(A2.5)} \qquad \sup_{|\beta_i| < \varepsilon,\, \beta_i \neq 0} \beta_i^{-2} \left| O\left(\beta_i^2\right) \right| < M.$$

By (A2.4)–(A2.5), $\Sigma \beta_i^2 < \infty$ implies $\Sigma_{i=1,\infty} |a_i - 1| < \infty$ and thus $\Pi_{i=1,\infty} a_i < \infty$.

Finally, we must prove (2.6)–(2.7). The proofs of (2.6) and (2.7) are substantially identical, so we prove only (2.6). By Theorem A1.1,

$$\text{(A2.6)} \qquad E\left[\exp(-p\alpha_t)\sigma_t^{2p} \middle| z_0, z_{-1}, z_{-2}, \ldots\right] - E\left[\exp(-p\alpha_t)\sigma_t^{2p}\right]$$
$$= \left[\exp\left(p \sum_{j=t}^{\infty} \beta_j g(z_{t-j})\right) - \prod_{j=t}^{\infty} E\left[\exp\left(p\beta_j g(z_{t-j})\right)\right]\right]$$
$$\cdot \prod_{i=1}^{t-1} E\left[\exp\left(p\beta_i g(z_{t-i})\right)\right].$$

The last term on the right-hand side of (A2.6) is finite by Theorem A1.2. (2.6) will therefore be proven if we can show that

$$\text{(A2.7)} \qquad \operatorname*{plim}_{t \to \infty} \left[\exp\left(p \sum_{j=t}^{\infty} \beta_j g(z_{t-j})\right) - \prod_{j=t}^{\infty} E\left[\exp\left(p\beta_j g(z_{t-j})\right)\right]\right] = 0.$$

First, consider the unconditional variance of the log of the first term on the left-hand side of (A2.7). We have

$$\text{(A2.8)} \qquad \operatorname{Var}\left[\sum_{j=t}^{\infty} p\beta_j g(z_{t-j})\right] = p^2 \operatorname{Var}\left(g(z_t)\right) \sum_{j=t}^{\infty} \beta_j^2 \to 0 \quad \text{as} \quad t \to \infty.$$

Since convergence in L^2 implies convergence in probability,

$$\text{(A2.9)} \qquad \exp\left[\sum_{j=t}^{\infty} p\beta_j g(z_{t-j})\right] \xrightarrow{P} 1.$$

Finally,

$$\text{(A2.10)} \qquad \lim_{t \to \infty} \prod_{j=t}^{\infty} E\left[\exp\left(p\beta_j g(z_{t-j})\right)\right] = \lim_{t \to \infty} \exp\left[\sum_{j=t}^{\infty} \ln\left(E\left[\exp\left(p\beta_j g(z_{t-j})\right)\right]\right)\right]$$
$$= \exp\left[\sum_{j=t}^{\infty} \ln\left[1 + O\left(\beta_j^2\right)\right]\right]$$
$$= \exp\left[\sum_{j=t}^{\infty} O\left(\beta_j^2\right)\right] \to 1 \quad \text{as} \quad t \to \infty$$

by (A2.4)–(A2.5) and the square summability of $\{\beta_j\}$. *Q.E.D.*

REFERENCES

ABEL, A. B. (1988): "Stock Prices Under Time-Varying Dividend Risk: An Exact Solution in an Infinite-Horizon General Equilibrium Model," *Journal of Monetary Economics*, 22, 375–393.

AKAIKE, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory*, ed. by B. N. Petrov and F. Csáki. Budapest: Akadémiai Kiadó, 267–281.

BACKUS, D., AND A. GREGORY (1989): "Theoretical Relations Between Risk Premiums and Ccnditional Variances," Mimeo, Federal Reserve Bank of Minneapolis and Queen's University.

BARCLAY, M. J., R. H. LITZENBERGER, AND J. WARNER (1990): "Private Information, Trading Volume, and Stock Return Variances," *Review of Financial Studies*, 3, 233–254.

BARSKY, R. B. (1989): "Why Don't the Prices of Stocks and Bonds Move Together?" *American Economic Review*, 79, 1132–1145.

BERNANKE, B. S. (1983): "Irreversibility, Uncertainty and Cyclical Investment," *Quarterly Journal of Economics*, 98, 85–106.

BILLINGSLEY, P. (1986): *Probability and Measure*, Second Edition. New York: John Wiley and Sons.

BLACK, F. (1976): "Studies of Stock Market Volatility Changes," *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

BOLLERSLEV, T., R. F. ENGLE, AND J. M. WOOLDRIDGE (1988): "A Capital Asset Pricing Model with Time Varying Covariances," *Journal of Political Economy*, 96, 116–131.

BOX, G. E. P., AND G. C. TIAO (1973): *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley Publishing Co.

CHOU, R. Y. (1987): "Volatility Persistence and Stock Returns—Some Empirical Evidence Using GARCH," *Journal of Applied Econometrics*, 3, 279–294.

CHRISTIE, A. A. (1982): "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 10, 407–432.

CLARK, P. K. (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

COCHRANE, J. (1988): "How Big is the Random Walk in GNP?" *Journal of Political Economy*, 96, 893–920.

COX, J., J. INGERSOLL, AND S. ROSS (1985): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

DAVIS, P. J. (1965): "The Gamma Function and Related Functions," in *Handbook of Mathematical Functions*, ed. by M. Abramowitz and I. A. Stegun. New York: Dover Publications.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1008.

ENGLE, R. F., AND T. BOLLERSLEV (1986a): "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

——— (1986b): "Modelling the Persistence of Conditional Variances: A Reply," *Econometric Reviews*, 5, 81–87.

ENGLE, R. F., D. M. LILIEN, AND R. P. ROBINS (1987): "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391–408.

ENGLE, R. F., AND G. GONZÁLEZ-RIVERA (1989): "Semiparametric ARCH Models," U.C.S.D. Economics Department Discussion Paper 89-17.

FAMA, E. F. (1965): "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34–105.

FRENCH, K. R., AND R. ROLL (1986): "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17, 5–26.

FRENCH, K. R., G. W. SCHWERT, AND R. F. STAMBAUGH (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

GENNOTTE, G., AND T. A. MARSH (1987): "Variations in Economic Uncertainty and Risk Premiums on Capital Assets," Mimeo, U.C. Berkeley Graduate School of Business.

GEWEKE, J. (1986): "Modelling the Persistence of Conditional Variances: A Comment," *Econometric Reviews*, 5, 57–61.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. RUNKLE (1989): "Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," Banking Research Center Working Paper #166, Northwestern University.

GRADSHTEYN, I. S., AND I. M. RYZHIK (1980): *Table of Integrals, Series and Products*. New York: Academic Press.

HANNAN, E. J. (1980): "The Estimation of the Order of an ARMA Process," *The Annals of Statistics*, 8, 1071–1081.

HARVEY, A. C. (1981): *The Econometric Analysis of Time Series*. Oxford: Philip Allan.

HOSKING, J. R. M. (1981): "Fractional Differencing," *Biometrika*, 68, 165–176.

LO, A., AND C. MACKINLAY (1988): "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, 1, 41–66.

MANDELBROT, B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

MCDONALD, R. M., AND D. SIEGEL (1986): "The Value of Waiting to Invest," *Quarterly Journal of Economics*, 101, 707–728.

MERTON, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–888.

NELSON, D. B. (1988): "The Time Series Behavior of Stock Market Volatility and Returns," Unpublished Doctoral Dissertation, M.I.T. Economics Department.

——— (1989): "Modelling Stock Market Volatility Changes," *1989 Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 93–98.

——— (1990a): "Stationarity and Persistence in the GARCH(1, 1) Model," *Econometric Theory*, 6, 318–334.

——— (1990b): "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance With the Wrong Model," forthcoming, *Journal of Econometrics*.

NEWEY, W. K. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229–256.

PAGAN, A. R., AND Y. HONG (1988): "Non-Parametric Estimation and the Risk Premium," forthcoming in *Nonparametric and Semiparametric Methods in Econometrics*, ed. by W. Barnett, J. Powell, and G. Tauchen. Cambridge: Cambridge University Press.

PAGAN, A. R., AND H. C. L. SABAU (1987): "Consistency Tests for Heteroskedastic and Risk Models," Mimeo, University of Rochester Economics Department, October, 1987.

PANTULA, S. G. (1986): "Modelling the Persistence of Conditional Variances: A Comment," *Econometric Reviews*, 5, 71–73.

POTERBA, J. M., AND L. H. SUMMERS (1986): "The Persistence of Volatility and Stock Market Fluctuations," *American Economic Review*, 76, 1142–1151.

RUDIN, W. (1976): *Principles of Mathematical Analysis*. New York: McGraw-Hill Book Company.

SCHOLES, M., AND J. WILLIAMS (1977): "Estimating Betas from Nonsynchronous Data," *Journal of Financial Economics*, 5, 309–327.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

SCHWERT, G. W. (1989a): "Business Cycles, Financial Crises and Stock Volatility," *Carnegie-Rochester Conference Series on Public Policy*, 31, 83–126.

——— (1989b): "Why Does Stock Market Volatility Change over Time?" *Journal of Finance*, 44, 1115–1154.

STOUT, W. F. (1974): *Almost Sure Convergence*. London: Academic Press.

WEISS, A. A. (1986): "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107–131.

WIGGINS, J. B. (1987): "Option Values Under Stochastic Volatility: Theory and Empirical Estimates," *Journal of Financial Economics*, 19, 351–372.