THE TIME SERIES BEHAVIOR OF STOCK MMARKET VOLATILITY AND RETURNS

by

DANIEL BARLOW NELSON

B.A., Economics, University of Utah (1983)

SUBMITTED TO THE DEPARTMENT OF ECONOMICS IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1988

copyright Daniel Barlow Nelson 1988
The author hereby grants to M.I.T. permission to reproduce
and to distribute copies of this thesis in whole or in part.

Signature of Author ____________________________________________ Department of Economics

May 12, 1988

Certified by __________________________________________________ Jeffreý M. Wooldridge
Thesis Supervisor

Accepted by __________________________________________________ Peter Temin
Chairman, Graduate Committee

JUN 7 1988
ARCU LIBRARIES
THE TIME SERIES BEHAVIOR OF STOCK MARKET VOLATILITY AND RETURNS

by

Daniel Barlow Nelson

Submitted to the Department of Economics on May 12, 1988 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

ABSTRACT

The three papers that comprise this dissertation approach the econometric modelling of randomly changing stock market volatility from three different directions. The paper that comprises Chapter one, "The Time Series Behavior of Stock Market Volatility and Returns" models volatility as an unobserved state variable. Using a logarithmic transformation, this non-linear random coefficients model is transformed into a linear system and estimated using a Kalman Filter.

The models considered in chapters two and three belong to the family of ARCH models introduced by Robert Engle in 1982. Chapter two, "Stationarity and Persistence in the GARCH(1,1) Model" develops new results about the most commonly used ARCH model, GARCH(1,1). Finally, Chapter three, "Conditional Heteroskedasticity in Asset Returns: A New Approach" critiques the use of ARCH processes in modelling changing asset market volatility, and introduces a new member of the ARCH family that may be more appropriate for such modelling.

Thesis Supervisors: Jeffrey M. Wooldridge
John D. Cox

Third Reader: Jerry A. Hausman
ACKNOWLEDGEMENTS

A great deal of thanks is due my thesis committee, Jeff Wooldridge (chair) and John Cox. They were extraordinarily helpful. Jim Poterba gave many useful suggestions on both the substance and presentation of the papers. Chi fu Huang, Danny Quah, Daniel Stroock, Richard Dudley and Dan McFadden were also very helpful. As third reader, Jerry Hausman made a number of useful comments, and also read the thesis sooner than he had to, making it possible for me to graduate in May, for which I will be eternally grateful. Naturally, all of the above individuals are guiltless of any errors that may remain.

Participants in seminars at M.I.T., Princeton, Yale, Wharton, Berkeley, Stanford, Northwestern, Rochester, Chicago and the December 1987 A.E.A. meetings made many valuable comments and suggestions, many of which will be implemented in future versions of these papers. The National Science Foundation and the Department of Education provided research support.

Finally, I thank my long-suffering wife Therese and our children Caroline, Allen, and Scott for making the years at M.I.T. very happy. While this thesis is dedicated to them, they are hereby released from any obligation to read it.
# TABLE OF CONTENTS

**CHAPTER ONE:** The Time-Series Behavior of Stock Market
   Volatility and Returns .................................. 5
   Appendix .................................................. 60

**CHAPTER TWO:** Stationarity and Persistence in the
   GARCH(1,1) Model ........................................ 75

**CHAPTER THREE:** Conditional Heteroskedasticity in Asset
   Returns: A New Approach ................................ 108
   Appendix I ................................................ 150
   Appendix II .............................................. 164
CHAPTER ONE:

THE TIME SERIES BEHAVIOR OF STOCK MARKET VOLATILITY AND RETURNS

September 1987

Summary: This paper presents a state space model of conditional heteroskedasticity in stock market excess returns. The model is estimated using returns data from the CRSP tape for the value weighted market index from 1962 to 1985. We find very strong effects of leverage on volatility, weak evidence of mean reversion in conditional volatility, a significant relation between risk and expected return, and evidence for conditional normality of returns, with fat-tails accounted for by a slowly changing conditional variance.
I. INTRODUCTION

It has long been recognized in the finance literature that the volatility of stock market returns changes over time. These changes have important implications for financial economics. For example, the Black-Scholes options pricing formula is invalid if volatility changes randomly over time (Cox (1975), Wiggins (1985, 1986)). In addition, the recent work of Barsky (1986) and Abel (1986) has highlighted the importance of the perceived level of market risk in determining both riskless and risky required rates of return. When market risk rises, riskless rates fall and the risk premium rises. Required rates of return on risky investments may rise or fall, depending on the aggregate level of relative risk aversion. If shocks to volatility are expected to persist, then shifting volatility may have important implications for the term structure of interest rates. Finally, Bernanke (1983), McDonald and Siegel (1986) and Bertola (1987) have pointed out that if investment is irreversible and demand is uncertain, then the option that agents have to wait to invest rather than investing now will strongly affect the timing of investment. If volatility rises, then the option to wait to invest becomes more valuable. If asset volatilities in an economy tend to move together, then changing market volatility may play an important role in explaining business cycle
movements in investment. Some preliminary evidence (see Pindyck (1986)) suggests that it does.

In recent years, two basic econometric strategies have emerged for modeling the time series behavior of market volatility. The first, exemplified by the recent work of Poterba and Summers (1986) and of French, Schwert and Stambaugh (1986) is to chop a time series of market returns into blocks of, say, one month each, to assume that volatility is constant within each block (month), create a (monthly) time series of estimated variances, and then estimate the stochastic process followed by either the estimated monthly variances (Poterba and Summers) or by their logs (French, Schwert and Stambaugh.)

The dynamics of the estimated variance series, however, will differ from the dynamics of the true underlying conditional variance series both because of measurement error (i.e. variance is not observable and is estimated each period with error) and because of time aggregation (i.e. it seems unlikely that variance changes suddenly at the end of each month; this is only an approximation to what is probably a process of continuous adjustment in conditional variance.) Both time aggregation (Sims 1971) and measurement error (Granger and Morris 1976) will distort the dynamics of the estimated variance series from dynamics of the true variance series. Unfortunately, the Poterba-Summers-French-Schwert-Stambaugh approach cannot reduce one
source of distortion without increasing the other. For example, we can reduce time aggregation by assuming that variance is constant over a shorter period, but this increases the measurement error in estimated variance. Conversely, we can assume that variance is constant over a longer period, and if this assumption is correct, measurement error is reduced. But if it is not correct, we increase bias due to time aggregation.

Reducing time aggregation may be particularly crucial in modeling the effect of changing variances on options valuation. i.e. since the typical option has a life of only a few months, modeling variance as being constant between monthly jumps seems likely to lead to substantial distortions if in fact variance changes continuously.

To improve on the Poterba-Summers-French-Schwert-Stambaugh methodology therefore requires that we explicitly account for measurement error and reduce time aggregation as much as possible. Finally we would certainly like to impose the condition that volatility is non-negative, which the procedure employed by Poterba and Summers (i.e. estimating an ARMA process for standard deviation or for variance) does not do.

ARCH models provide the second major approach to modeling processes with changing conditional volatilities. The theory underlying ARCH processes and their estimation was developed by Engle (1982) and has since been extended in
many interesting ways. Engle, Lillien and Robbins (1987) extended the basic ARCH model to allow conditional means to depend on conditional variance and applied this to model time varying risk premia in the term structure of interest rates, and Bollerslev, Engle, and Wooldridge (1986) have further extended these results to estimate a dynamic capital asset pricing model with time varying asset covariances.

Since ARCH models do not suffer from the time aggregation and measurement error problems of the methods discussed above, they represent a major improvement in the realism of time series models of changing variance. Nevertheless, it is not clear that ARCH models are appropriate for modeling conditional heteroskedasticity in financial asset markets. A basic assumption of ARCH models (see Engle 1982, section 4 and theorem 4), is that only the magnitude of lagged residuals—and not their signs—determine conditional variance. i.e. residuals of -2 or 2 have the same effect on future conditional variance, and increase future variance relative to residuals of either -1 or 1.

This symmetry assumption runs afoul of a widely noted (see for example, Black (1976) and Christie (1982)) stylized fact about asset markets, namely that positive returns today are associated with lower volatility tomorrow, and that negative returns today are associated with higher volatility tomorrow. For example, if the market experiences a large
positive gain today, then an ARCH model would predict that conditional variance would be higher tomorrow than it was today. The results of Black and Christie, however, would lead us to suspect that conditional variance should drop in this case, not rise. This restrictive assumption of ARCH models can have serious implications: for example, if the ARCH symmetry assumption (along with further regularity conditions) is correct in an ARCH regression, then regression parameters and covariance parameters are asymptotically independent (Engle 1982 theorem 4). If the symmetry condition does not hold, we have no guarantee of such independence, so that misspecification of conditional variance could cause inconsistent estimation of regression parameters. For ARCH-M models, any misspecification of the conditional variance will imply misspecification for the conditional mean. Therefore, in using ARCH models, we would like to be able to test the symmetry condition, and have an alternate form of conditional heteroskedasticity that allows conditional variance to depend on both the size and the sign of past returns.

In the remainder of this paper, we will develop and test an alternative model of conditional heteroskedasticity in asset markets that meets the objections to the econometric methods outlined above. In addition, we will be able to relax the assumption of conditional normality, account for the effect of non-trading periods on conditional
volatility, have a simple form for forecasting future risk premia and volatilities, and allow for positive or negative serial correlation of realized excess returns. The models in the existing literature that are most closely related to what is presented in this paper are the models of Merton (1980) and Wiggins (1986), which may be viewed as a special cases of our model.

In section II, we will present the model. In section III, we discuss estimation and testing. In section IV, we present our statistical results. In section V we discuss smoothing and forecasting of volatility and market risk premia. In section VI, we briefly discuss extensions to the model. Section VII concludes.
II. A MODEL OF MARKET EXCESS RETURNS

The model of market excess returns that we develop in this section is inspired by the models of Merton (1980.) Merton assumed that the variance of market returns was constant for blocks of several years at a time, and then investigated the relation between returns and variance. In his first model, he assumed that the required risk premium on the market is proportional to the variance of market returns. In his second model, he assumed that the required risk premium is proportional to the standard deviation of market returns (i.e. that the capital market line has a constant slope.) The model developed extends Merton's model 2 by allowing variance to change with each observation. We use model 2 instead of model 1 because it turns out to be quite easy to estimate with changing, unobserved conditional variance.

Let \( e_{t} \) be the realized logarithmic excess return on the market portfolio at time \( t \). Let \{\( z_{t} \)\} and \{\( e_{t} \)\} be independent sequences of i.i.d. normal random variables, where for all \( t \), \( z_{t} \) is \( N(0,1) \) and \( e_{t} \) is \( N(0,\sigma_{e}^{2}) \). Let \{\( \alpha_{t} \)\} be a non-stochastic but possibly time varying sequence of real numbers. Our model is then represented by the following equations:

\[
2.1) \quad e_{t} = \sigma_{t} b + z_{t} \sigma_{t} \\
2.2) \quad \ln(\sigma_{t}^{2}) = \alpha_{t} + \sum_{k=0}^{\infty} \lambda_{k} (\theta z_{t-1-k} + e_{t-k})
\]

where
2.3) \[ \sum_{k=0}^{\infty} \lambda_k^2 < \infty \]

All parameters are assumed to be real and finite. \( \sigma_t \) is the instantaneous standard deviation of excess returns which is assumed known to the market (but not to the econometrician) at time \( t \). To the econometrician, it should be interpreted as an unobserved state variable. Note that by construction, \( \sigma_t \) and \( z_t \) are independent at time \( t \). Therefore at time \( t \) \( \varepsilon_t \) has conditional mean \( \sigma_t \beta \) and conditional variance \( \sigma_t^2 \).

Equation 2.2 gives the moving average representation for the process \( \ln(\sigma_t^2) \). Under the above assumptions \( \ln(\sigma_t^2) \) is stationary around its deterministic component \( \alpha_t \). \( \{\theta z_{t-1} + \varepsilon_t\} \) is its innovations sequence, which has a component \( \{\theta z_{t-1}\} \) that may be positively or negatively correlated with lagged returns and a component \( \{\varepsilon_t\} \) that is independent of lagged returns. In essence the \( z_t \) component picks up volatility changes that are correlated with lagged returns (i.e. through leverage effects), and the \( \varepsilon_t \) term picks up changes in volatility from other sources.

The means and covariograms of \( \{\varepsilon_t\} \) and \( \{\sigma_t^2\} \) are given in theorem 2.1.

**THEOREM 2.1**

Define

2.4) \[ A_k = \sum_{i=0}^{k-1} \lambda_i (\theta z_{t-1-i} + \varepsilon_{t-1}) + \sum_{i=k}^{\infty} (\lambda_i + \lambda_{i-k}) (\theta z_{t-1-i} + \varepsilon_{t-1}) \]

then
2.5) \( \text{Var}(A_k) = [\theta^2 + \sigma_\varepsilon^2] \sum_{l=0}^{k-1} \lambda_1^l + \sum_{l=k}^{\infty} (\lambda_1 + \lambda_{1-k})^2 \)

2.6) \( \text{E}(\sigma_1^2) = \exp[\alpha_t + \text{Var}(A_0)/8] \)

2.7) \( \text{Var}(\sigma_1^2) = \exp[2\alpha_t + \text{Var}(A_0)/2] - \exp[2\alpha_t + \text{Var}(A_0)/4] \)

2.8) \( \text{Cov}(\sigma_1^2, \sigma_{1-k}^2) = \exp[\alpha_t + \alpha_{1-k} + \text{Var}(A_k)/2] \)

\[ - \exp[\alpha_t + \alpha_{1-k} + \text{Var}(A_0)/4] \]

2.9) \( \text{E}(\varepsilon_t) = b \exp[\alpha_t/2 + \text{Var}(A_0)/32] \)

2.10) \( \text{Var}(\varepsilon_t) = (b^2 + 1) \exp[\alpha_t + \text{Var}(A_0)] \)

\[ - b^2 \exp[\alpha_t + \text{Var}(A_0)/16] \]

2.11) \( \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = b^2 \text{Cov}(\sigma_1, \sigma_{1-k}) \)

\[ + b\alpha_{k-1} \exp[\alpha_t + \alpha_{1-k} + \text{Var}(A_k)/8] \]

If \( \alpha_t = \alpha \) for all \( t \), then \( \{\varepsilon_t\}, \{\ln(\sigma_t^2)\}, \{\sigma_t\} \) and \( \{\sigma_1^2\} \) are jointly covariance stationary. If \( \alpha_t \) is not constant over time, then \( \{\varepsilon_t\}, \{\ln(\sigma_t^2)\}, \{\sigma_t\} \) and \( \{\sigma_1^2\} \) are still individually correlation stationary, in that although their covariograms are a function of time, their correlograms are not.

**Proof:** See appendix.

From 2.11 we can see that excess returns can be positively or negatively correlated, or positively correlated at some lags and negatively correlated at others. The basic intuition is that \( \varepsilon_t \) consists of two components: drift and noise. The drift component is proportional to the instantaneous standard deviation of \( \varepsilon_t \). If volatility is
positively serially correlated, as we might suspect, then 
the drift component will be positively serially correlated. 
The noise component is, by assumption, serially 
uncorrelated. If, however, $\theta \neq 0$, then noise at time $t$ will 
affect volatility at times greater than $t$. This, in turn, 
will affect future drift. If $\lambda_k > 0$ and $\theta < 0$, then bad 
news today brings high volatility tomorrow, which in turn 
brings a higher risk premium tomorrow. This tends to induce 
a negative serial correlation in excess returns. Which 
tendency will prevail at which lag is a question we can hope 
to answer with the model. Unfortunately, it will turn out 
that our estimated parameters will imply negative 
correlation of $\{e_{t}\}$ at all lags, whereas recent work by Lo 
and MacKinlay(1987) finds positive correlation at short lags 
(up to a month or so) and negative correlation at longer 
lags.
III ESTIMATION AND TESTING

Although the model introduced in section II is highly non-linear and includes unobservable variables such as \( \ln(\sigma_t^2) \) and \( z_t \), it turns out that the model can be put into a linear state space form. We can then use a Kalman Filtering techniques for estimation, forecasting and smoothing. Even if the stochastic sequences \( \{z_t\} \) and \( \{e_t\} \) are normal, however, linearizing the model will introduce non-normal errors in the state space form, and the least squares estimates we employ are quasi maximum likelihood estimators, whose theory has been developed in Domowitz and White (1982), Ljung and Caines (1979), Gourieroux, Monfort and Holly (1984), and Wooldridge (1987).

i) Putting the Model in Linear Form:

Recall equations 2.1 and 2.2:

2.1) \[ e_{t+1} = \sigma_t (z_t + b) \]

2.2) \[ \ln(\sigma_t^2) = \alpha_t + \sum_{k=0}^{\infty} \lambda_k (\theta z_{t-1-k} + e_{t-k}) \]

First we will consider the simple case when \( b = 0 \) for all \( t \), so that we have

3.1) \[ e_{t+1} = \sigma_t z_t \]

3.2) \[ \ln(\sigma_t^2) = \alpha_t + \sum_{k=0}^{\infty} \lambda_k (\theta z_{t-1-k} + e_{t-k}) \]

Both the left and right hand sides of 3.2 are unobservable. But note that, from 3.1

3.3) \[ (e_{t+1})^2 = (\sigma_t z_t)^2 \]

Taking logs, we have
3.4) \[ \ln(\epsilon_{t}^{2}) = \ln(\sigma_{t}^{2}) + v_{t} \]

where

3.5) \[ v_{t} = \ln(z_{t}^{2}) \]

Since the for all \( t \), \( v_{t} \) is a simple function of \( z_{t} \), \( \{v_{t}\} \) is an i.i.d. sequence. Further, it has easily computable moments. In fact, using formulae 9.3880 # 3 and 8.384 # 1 from Gradshteyn and Ryzhik (1965) we have

3.6) \[ E(v_{t}) = -\gamma - \ln(2) = -1.27 \]

where \( \gamma \) is Euler's constant (≈ 0.577) and

3.7) \[ \text{Var}(v_{t}) = \pi^{2}/2 = 4.93 \]

We can now restate 3.2 as

3.2* \[ Y_{t}^{*} = \ln(\epsilon_{t}^{2}) = \gamma_{t} + v_{t} + \sum_{k=0}^{\infty} \lambda_{k}(\theta z_{t-k} + \epsilon_{t-k}) \]

The left hand side of 3.2* is observable, and the right hand side is linear process. Our next step is to linearize

3.1. Define

3.8) \[ S_{zt} = \text{sign}(z_{t}) \]

Using 3.1 and the fact that for all \( t \), \( \sigma_{t} > 0 \) a.s., we have

3.9) \[ \text{Sign}(\epsilon_{t}) = \text{Sign}(\sigma_{t} z_{t}) = \text{Sign}(z_{t}) = S_{zt} \]

Thus, although \( \{z_{t}\} \) is not observable, \( \{S_{zt}\} \) is. Using the fact that \( E|z_{t}| = (2/\pi)^{1/2} \), it is easy to check that

3.10) \[ z_{t} = S_{zt} (2/\pi)^{1/2} + W_{t} = Q_{t} + W_{t} \]

where \( E(W_{t}) = 0, \text{Var}(W_{t}) = 1 - (2/\pi) \), and \( \text{Cov}(W_{t}, S_{zt}) = 0 \). In
other words, 3.10 decomposes \( z_t \) into two orthogonal components, an observable component \( Q_t = S_z E(iz_t) \) and an unobservable component \( (W_t) \). \( Q_t \) can be interpreted as the conditional expectation of \( z_t \) given the sign of \( z_t \). Perhaps surprisingly, its variance accounts for about 2/3 of the variance of \( z_t \). Our linearized system can now be written as:

\[
3.11) \quad S_z \sim \text{i.i.d.} = 1 \text{ with probability } 1/2 \\
= -1 \text{ with probability } 1/2
\]

\[
3.12) \quad Y_t^* = \alpha_t + \sum_{k=0}^{\infty} \lambda_k Q_{t-k-1} + v_t + \sum_{k=0}^{\infty} \lambda_k (W_{t-k} + e_{t-k})
\]

We also define:

\[
3.13) \quad Y_t = \ln(\sigma_t^2) = Y_t^* - v_t
\]

It is also easy to check that in the \( b=0 \) case we have been looking at, \( v_t, Q_t \) and \( W_t \) are all mutually orthogonal. They are certainly not independent, however, as they are all functions of \( z_t \).

Next, consider the log transform of equations 3.3 to 3.5, this time allowing \( b \neq 0 \). We have

\[
3.3^* \quad (e r_t)^2 = (\sigma_t \{z_t+b\})^2
\]

\[
3.4^* \quad \ln(e r_t^2) = \ln(\sigma_t^2) + v_t
\]

where

\[
3.5^* \quad v_t = \ln([z_t+b]^2)
\]

Again, for all \( t \), \( v_t \) is a simple function of \( z_t \) and \( b \), and \( \{v_t\} \) is an i.i.d. sequence. Unfortunately, there are no longer simple formulae for the moments of \( v_t, Q_t \) and \( W_t \) when
b does not equal zero. In addition, we no longer have orthogonality between \( v_t \), \( Q_t \) and \( \tilde{W}_t \) (although \( Q_t \) and \( \tilde{W}_t \) will still be orthogonal.) Fortunately, however, it is straightforward to compute these moments numerically.\(^1\)

**ii The Model in State Space Form:**

Next we need to choose a functional form for the error terms in 3.1 and 3.2. In this paper we will restrict these errors to follow low order ARMA processes. We will examine AR(1), ARMA(1,1), and ARMA(2,1) specifications for the log variance error process. Since the AR(1) and ARMA(1,1) models are special cases of the ARMA(2,1), we will write out the state space form only for this model. We therefore assume

\[
Y_t = \alpha_t + \Delta_1 Y_{t-1} + \Delta_2 Y_{t-2} + (\theta Z_{t-1} + e_t) + \phi(\theta Z_{t-2} + e_{t-1})
\]

For details on how to put ARMA models in state space form, see Harvey (1982).

The transition equation is given by

\(^1\) One problem that does arise, however, is that if we allow \( b \neq 0 \), \( s(var(v_t))/ab \) evaluated at \( b=0 \) is not defined. Therefore the regularity conditions for asymptotic normality will require that either we impose \( b=0 \), or else we restrict \( b \) to be bounded away from zero. This means that we don't know the asymptotic distributions for the test statistics of the hypothesis that \( b=0 \).
\[
\begin{align*}
3.14) \quad & \begin{bmatrix}
Y_t^* \\
Y_t \\
Y_{t-1} \\
Q_t \\
W_t
\end{bmatrix} =
\begin{bmatrix}
0 & \Delta_1 & 1 & \theta & \theta & 0 \\
0 & \Delta_1 & 1 & \theta & \theta & 0 \\
0 & \Delta_2 & 0 & \theta & \psi & \theta & \psi & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
Y_{t-1}^* \\
Y_{t-1} \\
Q_{t-1} \\
Q_t \\
W_t
\end{bmatrix} \\
+ \begin{bmatrix}
(1-\Delta_1 L-\Delta_2 L^2) \alpha_t + E(v_t) - \theta b \\
(1-\Delta_1 L-\Delta_2 L^2) \alpha_t - \theta b \\
- \theta \psi b \\
b \\
0
\end{bmatrix} + \begin{bmatrix}
1 & 0 & 0 \\
0 & \sigma_e & 0 & 0 \\
0 & \psi & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
v_t - E(v_t) \\
e_t \\
q_{1t} \\
q_{2t}
\end{bmatrix}
\end{align*}
\]

\[\equiv X_t = \phi X_{t-1} + C_t + \tau \varepsilon_t\]

The vector \([(v_t - E(v_t)), e_t, q_{1t}, q_{2t}]'\) has mean zero and is serially uncorrelated, with covariance matrix

\[
3.15) \quad \text{cov} \{[(v_t - E(v_t)), e_t, q_{1t}, q_{2t}]'\} = \begin{bmatrix}
\sigma_v^2(b) & 0 & \Gamma_1(b) & \Gamma_2(b) \\
0 & 1 & 0 & 0 \\
\Gamma_1(b) & 0 & \sigma_q^2(b) & 0 \\
\Gamma_2(b) & 0 & 0 & 1 - \sigma_q^2
\end{bmatrix}
\]

As indicated, the elements in the covariance matrix are functions of \(b\). For \(b \neq 0\), they must be computed numerically. The precise functional forms are given in the appendix. Our measurement equation is given by:

\[
3.16) \quad \begin{bmatrix}
Y_t^* \\
S_{(x+b)t}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & J^0
\end{bmatrix}
\begin{bmatrix}
Y_t^* \\
Y_t \\
Y_{t-1} \\
Q_t \\
W_t
\end{bmatrix} + \begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

\[= X_t^* = Z_t X_t + M_t\]

where
3.17) \[ J_0 = J_0(b) = \frac{(Q^-) + (Q')}{(Q^-) - (Q')} \]

3.18) \[ K_0 = K_0(b) = -2/[(Q^-) - (Q')] \]

and

3.19) \[ Q' = E[(z+b)|S(z+b)_t = 1] \]

\[ Q^- = E[(z+b)|S(z+b)_t = -1] \]

Integral formulae for computing \( J_0 \) and \( K_0 \) are found in the appendix.

To close the model, we need to specify a data set, plus a specific form for \( b \) and \( \alpha_1 \). The data set we will use is the logarithmic returns series for the Market-weighted index from the CRSP tape from June 1962 through December 1985.

It is well known that the fact that stocks do not trade continuously induces a spurious autocorrelation in observed returns on a market index. Lo and MacKinlay (1986) argue that taking returns over blocks of 4 or 5 days at a time should eliminate essentially all of this effect. Therefore we take returns over blocks of four trading days each. There are 1475 such blocks in our data set. For the riskless rate of interest, the Federal Reserve Discount rate was used. Treasury bill returns would have been better, but daily (or four day) treasury returns are unavailable.

It seemed wise not to set \( \alpha_1 \) equal to a constant, since a number of authors (i.e. French and Roll (1984)) have pointed out that there seems to be very different contribution to variance from trading days and non-trading days. For example, price movement on a Tuesday presumably
reflect information arriving from the Market close on Monday
to its close on Tuesday, a period of 24 hours. Price
movements on a typical Monday, however, reflect information
that has arrived between the Friday market closing and the
Monday closing, a period of about 72 hours. Yet the
variance of returns on Mondays, while greater than on
Tuesdays, is much less than three times as great. In fact
this difference is so great that French and Roll claim (for
reasons that space does not permit us to explore here) that
it may well be inconsistent with rational expectations in
the stock market. To account for this, we will let our
information clock run at two speeds, at a speed of one on
trading days and at a speed of $\beta$, on non-trading days.
Intuitively, we would expect that $0 < \beta < 1$. (i.e. some
information arrives on non-trading days, but not as much as
on trading days.) Let $NT_t$ be the number of non-trading days
in the (four trading day) block $t$. Our specification for $\alpha_t$
is then

$$3.20 \quad \alpha_t = \alpha + \ln(NT_t \beta + 4).$$

iv State Space Quasi Maximum Likelihood Estimation:

The model described in 3.14-3.19 is in the familiar
State Space/Kalman Filter form, except that the error terms
are not normally distributed. In the remainder of this
section, I will assume that the reader is familiar with the
basics of Kalman filtering. For a detailed reference on the
Kalman Filter, the reader is referred to Anderson and Moore (1979). For a succinct summary of the Kalman filter results, see Judge et. al. (1985) appendix C.) With or without normality, given the true parameter values, the Kalman filter will provide linear minimum mean squared error estimates of the state variables $X_t$, and linear minimum mean squared error forecasts of the observables $X_t^*$. We will be able to use these forecasts of $X_t^*$ to form a sequence of prediction errors. Minimizing a suitable function of these prediction errors will then allow us to consistently estimate the model parameters.

In our model, the equations of the Kalman Filter are:

Prediction:

3.21) $S_{t+1|-1} = \Phi S_{t-1} + C_t$

3.22) $\Pi_{t+1|-1} = \Phi \Pi_{t-1} \Phi' + \Sigma \Gamma'$

Measurement Update:

3.23) $S_t = S_{t+1|-1} + \Pi_{t+1|-1} (Z') F_t^{-1} (X_t^* - ZS_{t+1|-1} - M_t)$

3.24) $\Pi_t = \Pi_{t+1|-1} - \Pi_{t+1|-1} (Z') F_t^{-1} Z \Pi_{t+1|-1}$

where $S_{t+1|-1}$ is the optimal linear predictor of $X_t$ given information at time $t-1$, $S_t$ is the optimal linear predictor of $X_t$ given information at time $t$, $\Pi_t$ is the time $t$ covariance matrix of $(X_t - S_t)$, $\Pi_{t+1|-1}$ is the covariance matrix of $(X_t - S_{t+1|-1})$, and $F_t$ is the covariance matrix of the prediction error

3.25) $u_t = (X_t^* - ZS_{t+1|-1} - M_t)$

$F_t$ is given by
3.26) \[ F_t = Z\gamma_{t-1}Z' \]

Every period, the Kalman Filter generates predicted values for the observable variables \( X_t \). These predicted values for \( X_t \) equal \( Z(S_{t-1}) + M_t \). This is fairly intuitive when we recall the measurement equation 3.23 and the fact that our best one-period ahead forecast of what the state variables will be at time \( t \) is \( S_{t-1} \).

Subtracting the predicted value of the observables from the observed, we get the prediction error \( u_t \), which has mean zero and covariance \( F_t \). If the error terms in \( \xi_t \) were normally distributed for all \( t \), then the log likelihood function of the state space model would be

3.27) \[ L_T = -(K^{T/2})\ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} [\ln F_t + u_t'F_t^{-1}u_t] \]

where \( K \) is the number of state variables (5 in our model) and \( T \) is the number of observations (1475). We could then maximize the likelihood function to obtain consistent, efficient estimates of the model's parameters.

Unfortunately, as we have mentioned, our error terms are non-normal. However, Wooldridge (1987) has shown that if our model correctly specifies the first two conditional moments, then under regularity conditions reported in the appendix, the quasi maximum likelihood estimator obtained by maximizing the normal likelihood function \( L_T \) will yield consistent estimates of our model's parameters. Since the first two conditional moments are precisely what the model
specifies, our estimates will be consistent under the null hypothesis of correct specification.

Ljung and Caines (1979) showed that given consistency of a quasi-maximum likelihood estimator in a prediction error setup, and given mild regularity conditions, our parameter estimates will be asymptotically normal, with

\[
\begin{bmatrix}
\Delta_1 & \Delta_1 \\
\Theta & \Theta \\
\alpha & \alpha \\
\beta & \beta \\
\sigma^2_e & \sigma^2_e \\
b & b \\
\psi & \psi \\
\Delta_2 & \Delta_2
\end{bmatrix} \begin{pmatrix} \lambda \end{pmatrix} \overset{d}{\sim} N \left(0, H_0^{-1}V_0H_0^{-1}\right)
\]

where

\[
\begin{align*}
3.29) \quad H_0 &= \lim T^{-1} E[\nabla^2 L_T] \\
3.30) \quad V_0 &= \lim T^{-1} \sum_{t=1}^{T} E\left[(\nabla R_t)'(\nabla R_t)\right]
\end{align*}
\]

In 3.29 and 3.30, \( \nabla \) denotes partial differentiation with respect to the parameter vector, and \( \lambda_t = \ln|F_{t-1}| + u_t'F_{t-1}^{-1}u_t \).

Equations 3.28-3.30 give the now familiar form for the consistent covariance estimator of a quasi maximum likelihood model (see, for example, Gourieroux, Monfort and Trognon (1984).) \( H_0 \) is the limiting value of the hessian matrix and \( V_0 \) is the inner product of the score evaluated at the true parameter vector. \( H_0 \) is consistently estimated by the sample hessian divided by \( T \). \( V_0 \) is the score evaluated at the true parameter values, and is consistently estimated
by the inner product of the score evaluated at the estimated parameter values. The regularity conditions for asymptotic normality of our parameter estimates require \( L_1 \) to be three times differentiable with respect to the parameters and error terms, (this requires us to assume that \( b \) is bounded away from zero—see note 1) and rule out unit roots in the transition equation. They also require the finiteness of the expected values of sufficiently high powers of the error terms. It is straightforward to verify that these moment conditions are met as long as \( b \) does not equal zero. Hypothesis testing in the presence of unit roots is a topic we will take up in the next section.
IV RESULTS

The estimation was carried out using the GAUSS subroutine MAXMUM (Edlefsen and Jones (1986).) The Kalman filter equations are easily written in GAUSS. The numeric integrals were evaluated using the GAUSS routine INTQUAD. Tests indicate that the numeric integrals of the moments are accurate to about five significant digits. More computational details about the evaluation of the integrals are found in the appendix.

AR(1), ARMA(1,1) and ARMA(2,1) models were estimated. Given the ARMA forms chosen, we were able to easily compute the unconditional state covariances given the model parameters. This computed unconditional covariance matrix was used to initialize the Kalman filter, along with the unconditional mean of the state vector.

In the remainder of this section, we will present the basic parameter estimates discuss choosing between the models and then discuss several important issues addressed by the models: i) the relation between conditional variance and expected returns, ii) mean reversion in conditional volatility, iii) the effect of leverage on volatility, iv) conditional normality of the errors and v) the results of model specification tests.

The AR(1) Model: (standard errors are in parentheses)

\[ \text{er}_t = (0.074 + z_t) \sigma_t \]

(0.028)
\[ \ln \sigma_t^2 = -9.707 + \ln(N_t \times 0.280 + 4) + 0.966 \times \ln \sigma_{t-1}^2 \]
\[
(0.180) \\ (0.289) \\ (0.019)
\]
\[-0.161 \times z_{t-1} + e_t \]
\[
(0.046)
\]
\[
[z_{t-1} \ e_t]' \sim N \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.003 \end{bmatrix} \]
\[
(0.009)
\]
Value of the log quasi-likelihood function = -5353.272

The ARMA(1,1) Model:

4.2) \[ e_{t} = [0.077 + z_{t}] \times \sigma_t \]
\[
(0.027)
\]
\[ \ln \sigma_t^2 = -10.187 + \ln(N_t \times 0.306 + 4) + 0.974 \times \ln \sigma_{t-1}^2 \]
\[
(0.443) \\ (0.295) \\ (0.012)
\]
\[-0.300 \times z_{t-1} + e_t \]
\[
(0.074)
\]
\[
[z_{t-1} \ e_t]' \sim N \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.006 \end{bmatrix} \]
\[
(0.026)
\]
Value of the log quasi-likelihood function = -5349.206

The ARMA(2,1) Model:

4.3) \[ e_{t} = [0.057 + z_{t}] \times \sigma_t \]
\[
(0.029)
\]
\[ \ln \sigma_t^2 = -10.847 + 1.487 \times \ln \sigma_{t-1}^2 - 0.498 \times \ln \sigma_{t-2}^2 \]
\[
(0.392) \\ (0.010) \\ (0.010)
\]
\[-0.298 \times z_{t-1} + e_t - 0.801 \times (-0.298 \times z_{t-2} + e_{t-1}) \]
\[
(0.067) \\ (0.059)
\]
\[
[z_{t-1} \ e_t]' \sim N \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.004 \end{bmatrix} \]
\[
(0.021)
\]
Value of the log quasi-likelihood function = -5345.839

Asymptotic covariance estimates (and the results of the
specification tests discussed below) for the three models are found in tables I through III.

There are a conflicting criterion in the statistics literature for choosing between ARMA parameterizations. Two of the most widely-used criteria are the Schwartz criterion (SC) (Schwartz (1978)), which chooses the model to maximize the function

\[ L - (k/2)\ln(T) \]  
(\text{where } L \text{ is the log-likelihood function, } k \text{ is the number of parameters and } T \text{ is the number of observations,}) \text{ and the Akaike Information Criterion (AIC) (Akaike (1974)) which maximizes}

\[ L - k \]

The Schwartz criterion favors a more parsimonious parameterization than the AIC, and has the advantage of consistently estimating the order of the model (see Judge, et al 1985.) For the models we have estimated, the AIC and classical Likelihood ratio tests favor the ARMA(2,1) model, while the Schwartz criterion favors the ARMA(1,1) model. Of course, we cannot rule out that a parameterization of higher order than has been estimated in this paper is required. In any case, the dynamics of the smoothed series and the forecasting properties of the estimated ARMA(1,1) and ARMA(2,1) models are very similar, so that for most purposes there may not be a substantial difference between them.

Next, lets turn to some specific issues that the models
address:

i) Conditional Volatility and Expected Returns:

The estimated coefficients for $b$ are about the right order of magnitude (we will see this in section V when we examine the smoothed risk premia estimates.) The coefficients are all significantly greater than zero at the 95% level with a standard one tailed $t$ test, and all but one coefficient has a $t$ statistic over 2. Recall, however, that if $b=0$, one of the regularity conditions that we need for asymptotic normality is not satisfied, $(\text{VAR}(v_t))$ is not differentiable with respect to $b$ when $b=0$) so that the standard $t$ test may not have the right size.

ii) Mean Reversion in Conditional Volatility

There have been great strides in recent years in the econometric theory of testing in the presence of unit AR roots (see, for example, Phillips (1987) and the references therein.) In fact, tests for unit roots are available for a wide variety of linear models. Unfortunately, however, the theory of testing for unit roots in non-linear models has not yet been worked out, so that in applying the standard Dickey-Fuller $t$ tests in our model, the results must be interpreted with caution. With this in mind, let's look at the $t$ test statistics for the hypothesis that the largest AR root equals one. The $t$ test statistic is given by

\[(\Delta-1)/\sigma(\Delta),\]

where $\Delta$ is the largest AR root and $\sigma(\Delta)$ is its
estimated standard error.

<table>
<thead>
<tr>
<th>4.6) Model</th>
<th>Largest AR Root</th>
<th>Standard Error σ(Δ)</th>
<th>t=(Δ-1)/σ(Δ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.966</td>
<td>0.019</td>
<td>-1.789</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.974</td>
<td>0.012</td>
<td>-2.167</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>0.977</td>
<td>0.008</td>
<td>-2.919</td>
</tr>
</tbody>
</table>

The Dickey-Fuller test for models with a constant term rejects a unit root at the 1% level for \( t \leq -3.43 \), at the 5% level for \( t \leq -2.86 \) and at the 10% level for \( t \leq -2.57 \). Because the standard errors for the AR coefficients are relatively high, only the ARMA(2,1) models reject the presence of a unit root at the 5% level, despite the fact that the ARMA(2,1) models have AR roots larger than the roots for the AR(1) and ARMA(1,1) models.

These tests must be regarded with caution, however, since as Schwert(1985) and Quah(1987) have pointed out, tests for unit roots in time series models are very sensitive to the exact specification of the model, and if the model is misspecified, the tests for unit roots will often be inconsistent. Quah (1987), and Phillips and Perron(1986) have developed non-parametric tests for unit roots against stationary alternatives. We will apply the Phillips-Perron test of the null hypothesis

4.7) \( Y_t = Y_{t-1} + \tau + u_t \)

against the alternative

4.8) \( Y_t = \mu + \alpha Y_{t-1} + u_t \), with \( |\alpha| < 1 \)

\( \{u_t\} \) is a zero mean, process that is assumed to follow
strong mixing conditions, which allow it to exhibit considerable temporal dependence and variance heterogeneity. For example, \( \{u_t\} \) could follow any stationary ARMA process, as well as many processes with much greater shock persistence or heteroskedasticity than allowed by the ARMA framework. The exact restrictions that \( \{u_t\} \) is assumed to satisfy are detailed in Phillips and Perron.

The Phillips-Perron test involves regressing \( Y_t \) on \( Y_{t-1} \) and a constant, and then constructing the test statistic \( Z_t \alpha \) as defined in their paper. \( Z_t \alpha \) is basically a t-test of the hypothesis test that \( \alpha = 1 \) in 4.9 corrected for the effects of a unit root and the possible temporal heterogeneity of the \( u_t \) series. Critical values for \( Z_t \alpha \) are found in table 8.5.2 in Fuller (1976). For a sample size of 1475, a unit root is rejected at the 1% level for \( Z_t \alpha < -3.43 \).

For our model, recall

\[
2.2) \quad \ln(\sigma_t^2) = \alpha_t + \sum_{k=0}^{\infty} \lambda_k \eta_{t-k}
\]

The non-stationary alternative of interest is

\[
4.9) \quad \ln(\sigma_t^2) = \ln(\sigma_{t-1}^2) + (\alpha_t - \alpha_{t-1}) + u_t
\]

Where \( u_t \) satisfies 4.10-4.13. In terms of observables, we have

\[
4.10) \quad \ln(\sigma_t^2) = \ln(\sigma_t^2) + \ln[(NT_t B+4.0)/(NT_{t-1} B+4.0)] + u_t^* \]

where \( u_t^* = u_t + v_t - v_{t-1} \).

The only thing that now prevents us from directly applying the Phillips-Perron test is the presence in 4.16 of
the non-linear term \( \ln[(NT: B+4.0)/(NT_{t-1} B+4.0)] \). The asymptotics of unit roots in non-linear systems have not yet been worked out. We do, however, have strong priors about the value of \( B \), namely that \( 0 \leq B \leq 1 \), and if we are willing to assume a value for \( B \), we can apply the Phillips-Perron results directly. We therefore applied the Phillips-Perron test with \( B=0 \), \( B=0.1 \), \( B=0.2 \), and so on through \( B=1.0 \). For all values of \( B \) and all of the alternate methods of constructing \( Z_t \alpha \) tried, the results were nearly identical, with the null hypothesis of a unit root being overwhelmingly rejected in all cases. In every case, \( Z_t \alpha \) was below -50. Since all the results were so close, just one will be formally presented, for \( B=0.2 \) and lag truncation 4. (see the Phillips-Perron paper for an explanation of the lag truncation.)

Performing the regression in equation 4.14 yields

\[
\begin{align*}
\mu &= -9.9635 \quad \alpha = 0.0970 \quad Z_t \alpha = -53.9511, \\
\end{align*}
\]

which rejects \( \alpha = 1 \) at any reasonable confidence level.

The rejection of the null hypothesis using the Phillips-Perron test is so emphatic relative to the tests for a unit root based on the ARMA specifications as to call into question the appropriateness of using the Phillips-Perron tests in our sample. The Phillips-Perron test is based on the fact that asymptotically, the variance arising in the series from the presence of a unit root swamps all other sources of variance (i.e. in the \( \{u_t\} \) series) so that
regressing $Y_t$ on $Y_{t-1}$ will detect a unit root with probability one in the infinite sample. If, however, the non-stationary component of $Y_t$ is small and the variance and serial dependence of the $\{u_i\}$ is large, then in finite sample the Phillips-Perron test may be quite misleading. In our model, if there is a stochastic trend with low innovations variance, then the measurement error $v_t$ and the stationary component of log variance may be dominant when we regress $\ln(\text{er}_t)^2$ on $\ln(\text{er}_{t-1})^2$. To detect a slowly-changing stochastic trend therefore may require a substantially longer data series than we have used in this paper. We can say, however, that in the 1962-1985 data set that we have used, the only evidence we find for a unit root is the failure to reject in the AR(1) and ARMA(1,1) models.

iii Leverage Effects vs. Everything Else:

When the price of the stock of a levered firm drops, the firm becomes more heavily levered, which in turn should increase the volatility of its stock. Black (1976) and Christie (1982) showed, however, that if this pure leverage effect was all that moves volatility, the elasticity of the variance with respect to the stock price should be no more than -2. That is, a 1% in the stock price should lead to no more than a 2% drop in the returns variance of the stock. Black found this bound violated by measured changes in stock price volatility, and we do as well. The elasticity of $\sigma_t^2$
with respect to $S_{t-k}$ is easily found to be

$$
4.12 \quad \varepsilon(\sigma_1^2, S_{t-k}) = \left(\theta/\sigma_{1-k}\right)^k \lambda_k
$$

where $\lambda_k$ is the $k^{th}$ coefficient in the moving average representation for log variance. When $\sigma_{1-k}$ is low (high), this elasticity is high (low.) Figure 1 plots the percentage decrease in variance resulting from a one percent increase in $S$. The horizontal line at elasticity = -2 gives the Black-Christie lower bound on elasticity through leverage effects. The highest curved line represents the elasticity at horizons of 0 through 39 months when $\sigma_1$ is two standard deviations above its mean, the middle curve when $\sigma_1$ is at its mean, and the lower curve when it is two standard deviations below its mean. A 1% increase in the stock price leads to a large decrease in variance initially (over 20%)

The impact on variance decays over time, but the 1% stock price increase is still causing a 7% to 12% decrease in variance one month later, 5% to 9% four months later, and 1% to 3% a year later. Eventually, however, (as long as there is not a unit root in log volatility) the shocks die out. Our results confirm Black’s and Christie’s findings of a paradoxically large response of variance to returns, at least in the short run.

Perhaps the biggest surprise in the estimated coefficients is the minuscule role that $e_t$ seems to play. In all of the models estimated, $\sigma_e^2$ is close to zero, and in none of the models can we reject the hypothesis that $\sigma_e^2 =$
variance and returns

Solid line: Black–Christie lower elasticity bound.
Dashes: elasticity when sigma is at its mean.
Dots and dashes: elasticity when sigma
is two standard deviations above its mean.
Dots: elasticity when sigma is two
standard deviations below its mean.

FIGURE 1
0. At the same time \( \theta \) is significantly below zero at any standard significance level in all of the models estimated. The proportion of the innovations variance accounted for by \( \theta z_{t-1} \) is given by

\[
\alpha = \frac{\theta^2}{\theta^2 + \sigma_e^2}
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimated ( \alpha )</th>
<th>( \sigma_e^2 )</th>
<th>( \theta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.883</td>
<td>0.0034</td>
<td>0.0260</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.938</td>
<td>0.0060</td>
<td>0.0901</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>0.953</td>
<td>0.0044</td>
<td>0.0888</td>
</tr>
</tbody>
</table>

It seems, therefore, that the primary determinant of conditional volatility on the stock market as a whole is the recent direction of movement in the market. Indeed, the question seems not to be whether leverage effects matter, but rather whether anything else matters. The ARCH assumption of symmetric effects of residuals on conditional variance seems not to be appropriate for the stock market.

iv Testing Conditional Normality:

A well known and much investigated (i.e. see Mandelbroit (1963), Fama (1963),) fact about stock returns is that they are leptokurtotic--i.e. they have fatter tails than does a normal distribution. Up to this point we have assumed that \( \{z_t\} \) is i.i.d. \( N(0,1) \). In this subsection, we will modify this assumption and allow \( \{z_t\} \) to be an i.i.d. sequence with a non-normal distribution with tails either thinner or fatter than those of the normal distribution. We will find no evidence of non-normality. Unless the reader
is interested in the issue of the fatness of the tails of the returns distribution, s/he is advised to skip to the next subsection.

It has been known at least since the work of Clark (1973) that the unconditional distribution of a sequence of conditionally normal random variables \{x_t\} with randomly changing conditional variances \{\sigma_t^2\} will be leptokurtotic. It is easy to check that \varepsilon_t is leptokurtotic in the model of this paper. In fact, Clark's basic model is a special case of 2.1 through 2.4, with, for all t, \alpha_t = \alpha, \theta = 0 and \lambda_k = 0 for all k > 0 -- i.e. Clark's model had \{\ln(\sigma_t^2)\} be white noise. The low order ARMA models we estimated for \ln(\sigma_t^2) imposed a condition that conditional volatility changes relatively slowly compared with volatility in the white noise model of Clark. One logical question to ask about our model is if we will be able to explain all of the fat tails in the returns data, or if, after we have allowed for a slowly changing conditional volatility, conditional returns are still leptokurtotic. We may, therefore, want to relax the conditional normality assumption, so that we can allow for whatever fat or thin tails remain after we allow for a slowly changing variance.

The specific alternative to conditional normality that we test is the generalized error distribution (GED) (Harvey (1981.)) The density of a GED random variable is given by

\[ f(x) = C 2^{(1 - 1/c)} \cdot \exp\left[-0.5^{*1/c} \cdot (x-u)/\sigma_1^{c}\right]/[\sigma r(1/C)] \]
where \( u \) is the mean of \( x \), \( \sigma \) is a spread parameter, \( r(\cdot) \) is the gamma function, and \( C \) regulates the thickness of the tails of the distribution. The reason that the GED family of distributions represents an attractive alternative to the normal is that while the GED includes the normal (for \( C=2 \),) it also includes distributions with thinner tails than the normal (i.e. if \( C=\infty \) then the distribution is uniform) and fatter tails than the normal (i.e. if \( C=1 \), the distribution is double exponential.) In general, GED distributions with \( C < 2 \) have thicker tails than the normal, and distributions with \( C > 2 \) have thinner tails than the normal. The system we now test is identical to that considered earlier, except that instead of assuming that \( \{z_t\} \sim \text{i.i.d. } N(0,1) \), we instead assume that \( \{z_t\} \sim \text{i.i.d. with a standard GED density with } u=0, \sigma=1 \) and \( C \) a parameter to be estimated.

Recall that \( \ln(\sigma_1^2) = \ln(\sigma_t^2) + v_t \). It turns out that the variance of \( v_t \) is an increasing function of \( C \). In fact, it is easy to show, using Gradshteyn and Ryzhik (1980) formulae 4.352 #1 and 4.358 #2, that

\[
4.15) \quad E(v_t) = \frac{2}{C}[\psi(1/C) + \ln(2)]
\]

and

\[
4.16) \quad \text{Var}(v_t) = \frac{(2/C)^2}{[\psi'(1/C)]}
\]

where \( \psi(\cdot) \) is the Euler Psi function. Table VII gives the values of \( E(v_t) \) and \( \text{Var}(v_t) \) corresponding to a number of
different values of C. Allowing C to differ from 2 affects the dynamics of the system primarily through its effect on the variance of \( v_t \).

We re-estimated our three models using the GED parameterization just outlined, and the results are found in tables IV through VI. In no case do the estimates of any parameter change substantially. In every case, the point estimates for C were very close to 2, the C-value associated with the normal distribution. Unfortunately, the standard errors on the GED parameter estimates were quite high. The estimated coefficients, standard errors and t statistics for the hypothesis C=2 were:

\[
\begin{array}{cccc}
\text{Model} & C & \text{standard error} & \text{t-statistic} \\
\text{AR(1)} & 1.825 & 0.466 & -0.376 \\
\text{ARMA(1,1)} & 1.839 & 0.469 & -0.343 \\
\text{ARMA(2,1)} & 1.878 & 0.485 & -0.252 \\
\end{array}
\]

In no case do we reject the null hypothesis of normality, although the power of the test is low.

\( v \) Specification Tests

Each period, the Kalman filter produces an prediction error vector \( u_t \). If our model is correctly specified, then at the true parameter values, the prediction error sequence \{\( u_t \)\} will have mean zero for all t and will have conditional covariance given by the sequence \{\( F_t \)\}. Let \( F_t^{1/2} \) denote the Cholesky decomposition of \( F_t \). Define

\[
4.18) \quad \omega_t = F_t^{-1/2}u_t
\]

We can summarize the restrictions imposed by correct
specification as

4.19) for all \( t \):

a) \( E(\omega_t | I_t) = [0] \)

b) \( E(\omega_t \omega_t' | I_t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \)

c) \( E(\omega_t \omega_{t-k}' | I_t) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \) for all \( k \neq 0 \)

where \( I_t \) is information available at time \( t \).

These restrictions form a natural basis for conditional moment specification tests. Newey (1985) developed the theory of conditional moment specification testing for models estimated by maximum likelihood. Wooldridge (1987) has extended the Newey results by developing conditional moment tests for models estimated by quasi maximum likelihood methods. The models may exhibit time varying conditional heteroskedasticity. The Wooldridge results can be summarized as follows: Suppose we estimate a dynamic model parameterized by a \( m \) by \( 1 \) vector of parameters \( \theta \). In the dynamic model we predict each period a vector \( y_t \) conditional on information available at \( t \) (summarized by \( X_t \)). We then estimate the model using the one step ahead prediction errors \( u_t \) and a normal quasi likelihood function, i.e.,

4.20) \( \theta^* = \arg\max \{ \sum_{t=1}^{T} \lambda_t \} \)

where

4.21) \( \lambda_t = -(1/2)\ln|F_t(\theta)| - (1/2)u_t(\theta)'F(\theta)^{-1}u_t(\theta) \)
4.22) \( u_t = y_t - m_t(X_t, \theta) \)

where \( m_t \) is the conditional mean of \( y_t \) given \( X_t \), and \( F(\theta) \) is the conditional covariance matrix of \( y_t \) at time \( t \). \( F_t \) may also depend on \( X_t \), but in our model, it depends only on parameters. Suppose our null hypothesis is that for all \( t \) we have specified the conditional means \( m_t \) and the conditional variances correctly, and that this null hypothesis implies a set of \( n \) conditional moment restrictions \( r_{jt} \), \( j=1, \ldots, n \). That is, for all \( t, t=1, \ldots, T \), we have

4.23) \( E(r_{jt}(y_t, X_t, \theta) | X_t) = 0, \ j=1, \ldots, n \)

For notational convenience we will write \( r_{jt} = r_{jt}(\theta) \).

To test the orthogonality conditions implied by 3.44, we first define the test statistic

4.24) \( R_T = T^{-1/2} \sum_{t=1}^{T} [r_{1t}(\theta^v), \ldots, r_{nt}(\theta^v)]' \)

\[ \equiv T^{-1/2} \sum_{t=1}^{T} R_t(\theta^v) \]

and

4.25) \( O_T = T^{-1} \sum_{t=1}^{T} \begin{bmatrix} R_t(\theta^v) \\ \nabla \lambda_t \end{bmatrix} [R_t(\theta^v)'] = R_t(\theta^v) \nabla \lambda_t ' \]

4.26) \( G_T = T^{-1} \sum_{t=1}^{T} \nabla R_t(\theta^v) \)

4.27) \( I_n = \text{an } n \text{ by } n \text{ identity matrix} \)

4.28) \( H_T = T^{-1} \sum_{t=1}^{T} \nabla^2 \lambda_t \)
4.29) \[ P_T = \begin{bmatrix} I_n & \cdot \\ -G_TH_T^{-1} \end{bmatrix} O_T \begin{bmatrix} I_n & -H_T^{-1}G_T' \end{bmatrix} \]

Then

4.30) \[ R_T'P_T^{-1}R_T \geq x^2 c, \quad \text{where} \quad c \quad \text{is the rank of plim} \quad P_T. \]

As suggested by 4.19, correct specification of the model places restrictions on the first two conditional moments of the prediction errors. We accordingly test for serial correlation and conditional heteroskedasticity in the prediction errors. We noted above that our model for the returns process could account for serial correlation in returns and volatility. But if our model does not adequately capture the nature of these changes, this should show up in serially correlated prediction errors for \( S_{(z+b)} \) (the sign of excess returns) and \( Y^*_t \) (the log of squares excess returns.) We test for such serially correlated prediction errors at lags one through four. Formally, the conditional moment restriction we test is

4.31) \[ E \left( F_t^{-1/2}u_{t-j} \otimes I_{22} \right) F_t^{-1/2}u_t = 0_4 \quad j=1,\ldots,4 \]

Here \( I_{22} \) is the 2 by 2 identity matrix, and \( 0_4 \) is a 4 by 1 vector of zeros.

Finally, we test for conditional heteroskedasticity in the prediction errors:

4.32) \[ E \text{vech} \left[ (u_t u_t' - F_t)(u_{t-1} u_{t-1}' - F_{t-1}) \right] = 0_8 \]

These restrictions test for ARCH (of the Engle variety) in the log variance equation. If, for example, there was an arch effect for predicting \( Y_t^* \), we would take that as
evidence that the assumption that the \( z_i \) all came form the same (i.e. normal) distribution is incorrect.

The results, which are found in tables I through VI, are practically identical for all models estimated. On eighteen out of nineteen orthogonality conditions, no significant evidence of misspecification was found. On condition number four, however, which tested for serial correlation in \( \text{Sign}(er_t) \), very strong serial correlation (of about 13%) was found, and this serial correlation had a \( t \) statistic of about 5 for each of the models. The chi square statistic (with 19 degrees of freedom) was between 38 and 41 for each model, which rejects at the 1% level, since the 1% critical value for a chi square with 19 degrees of freedom is approximately 36.19. The nineteenth orthogonality condition, which requires that \( \text{Sign}(er_t) \) be homoskedastic, also has a fairly high \( t \) statistic, about 1.9. This is a natural consequence of the failure of the fourth condition, however: \( \text{Sign}(er_t) \) is a binomial random variable, and the variance of a binomial is a simple function of its mean. For a sequence of binomial random variables therefore, autocorrelation implies conditional heteroskedasticity.

In summary, the model therefore seems to fit the data very well, except for the one big problem of not being able to explain the degree of positive serial correlation in returns, for which a satisfactory explanation does not exist anywhere in the literature so far as I know. (see Lo and
MacKinlay (1987).) On every other score the models do quite well—i.e. there is no evidence of remaining ARCH effects in the log variance equation in any of the models examined, no evidence of serial correlation beyond first order in returns, and no evidence that a higher-order ARMA model is needed to model the log variance process.

We know that the model can, for appropriate coefficient values, accommodate positive returns correlation. Might it not be the case that with a higher order ARMA or some other small specification change that the puzzle might go away? Unfortunately this is not the case. The basic assumption that \( b \) is constant and that expected excess returns are linear in \( \sigma_t \) imply that \( \text{Sign}(er_t) = \text{Sign}((b+Z_t)\sigma_t) \) should be i.i.d.. But \( \text{Sign}(er_t) \) isn't i.i.d., and this won't be altered by a change in the specification of the log variance process: although our model does allow for serial correlation in \( er_t \), it does not allow for such correlation in \( \text{Sign}(er_t) \).
V. FORECASTING AND SMOOTHING

The theory of optimal smoothing and prediction with the Kalman filter is well known (see, for example, Anderson and Moore, 1979.) Since log variance is a state variable in our model, standard Kalman filter techniques will allow us to construct optimal linear smoothers and forecasters for the log variance series. Since, in our model, the log of variance is normal, the variance series and the risk premium series are lognormal. If all of the error terms $\xi_t$ in our linearized transition equation 3.14 were normally distributed, then the forecast and smoothing errors for log variance would be normally distributed as well, and we would have natural smoothers and forecasters for our the risk premium and variance series, namely:

5.1) $E[b\sigma_t | I_{t-k}] =
\exp \{E(\ln(\sigma_t^2) | I_{t-k}) + (1/8) \text{Var}(\ln(\sigma_t^2) | I_{t-k})\}$

and

5.2) $E[\sigma_t^2 | I_{t-k}] =
\exp \{E(\ln(\sigma_t^2) | I_{t-k}) + (1/2) \text{Var}(\ln(\sigma_t^2) | I_{t-k})\}$

where $E[a | I_{t-k}]$ and $\text{Var}[a | I_{t-k}]$ denote the mean and variance of a random variable a conditional on information at time $t-k$. (k will be negative for smoothing and positive for forecasting.) Formulae 5.2 and 5.3 just use the well known fact that if

5.3) $X \sim N(u, \sigma^2)$, then

$E[\exp(X)] = \exp[u + (1/2)\sigma^2]$
Unfortunately, some of the errors in the transition equation are not normally distributed, so that our use of 5.2 and 5.3 in smoothing and forecasting is only approximate.

i SMOOTHING:

We smooth the log variance series with a fixed lag smoother, in which we construct a sequence \( \{E(X_t \mid I_{t-k})\}_{t=1}^T \). That is, during each time period \( t \), we construct an optimal estimate of what the unobserved state \( X \) was \( k \) periods ago given the information we have at time \( t \). This filter is sub-optimal, in that in forming our smoothed estimate of \( X_t \), we are not using any information arriving after time \( t + k \). The optimal ("fixed interval") filter uses information in the entire sample to form the smoothed estimates \( X_t \) for each \( t \). Unfortunately, as Anderson and Moore (1979) discuss, fixed interval filters take a great deal of computer memory, and can also suffer from accumulation of round off errors which can cause serious computational problems. Anderson and Moore suggest that in many cases, a fixed lag smoother can approximate a fixed interval smoother without the same computational and memory problems. After some experimentation, it appeared that the variance of \( X_t - E(X_t \mid I_t, k) \) seemed to stop dropping much with increases in \( k \) for \( k \) bigger than about 10 or so, and \( k \) was therefore set equal to 10.
We construct our fixed lag smoother for log variance by augmenting our transition equation as follows:

5.4) Augmented transition equation:

\[
\begin{bmatrix}
Y_t^a \\
Y_{t-1}^a \\
Y_t^b \\
Q_t \\
W_t \\
Y_{t-1} \\
Y_{t-2} \\
\vdots \\
Y_{t-10}
\end{bmatrix}
= 
\begin{bmatrix}
0 & \Delta_1 & 1 & \theta & \theta & 0 & 0 & \ldots & 0 & 0 \\
0 & \Delta_1 & 1 & \theta & \theta & 0 & 0 & \ldots & 0 & 0 \\
0 & \Delta_2 & 0 & \theta & \psi & \psi & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\
\vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0
\end{bmatrix}
\begin{bmatrix}
Y_{t-1}^a \\
Y_{t-1} \\
Y_{t-1}^b \\
Q_{t-1} \\
W_{t-1} \\
Y_{t-2} \\
\vdots \\
Y_{t-11}
\end{bmatrix}
\]

\[
\begin{bmatrix}
(1-\Delta_1 L-\Delta_2 L^2)\alpha_t + E(v_t) - \Theta b \\
(1-\Delta_1 L-\Delta_2 L^2)\alpha_t - \Theta b \\
- \Theta \psi \beta \\
0 \\
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
+ 
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & \psi & 0 \\
0 & 0 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\vdots \\
0
\end{bmatrix}
\begin{bmatrix}
\epsilon_t \\
q_{1t} \\
q_{2t}
\end{bmatrix}
+ 
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
= 
X^a_t = \Phi^a X^a_{t-1} + C^a_t + \Gamma^a \xi_t
\]

Essentially, we add lags 1 through 10 of the log variance series as state variables. The transition equation sets \( Y_{t-2} = Y_{t-1} \) lagged once (which of course is \( Y_{t-2} \)) \( Y_{t-3} = Y_{t-2} \) lagged once and so on. The Kalman filter produces for us \( E(Y_{t-k} | I_t) \) and \( \text{Var}(Y_{t-k} | I_t) \) for \( k=0 \) to \( 10 \). For our smoothed series for \( Y_t = \ln(\sigma_t^2) \), therefore, we just use \( E(Y_t | I_{t+10}) \). Our smoothed series for \( \sigma_t^2 \) and \( b \sigma_t \) are constructed using \( E(Y_t | I_{t+10}), \text{Var}(Y_t | I_{t+10}) \) and formulas 5.2 and 5.3.
We convert the variance data to annual units\(^2\) and remove the non-trading days effect by defining

\[ Y_t^{**} = Y_t - \ln(N) \beta + 4) + \ln(365) \]

Figure 2 plots the raw (unfiltered) sequence \( \ln(e_{1,t}^2) = \ln(\sigma_t^2) + v_t \). Figures 3 through 5 plot the smoothed series for the log of variance, variance and risk premium for the ARMA(1,1) model. The ARMA(1,1) model was selected on the basis of the Schwartz criterion discussed earlier. The plots are substantially identical for all of the models, except that they are slightly smoother for the AR(1) model and also that the risk premium series is consistently about 2% lower for the ARMA(2,1) model.

The first thing we should notice about the smoothed log variance series is that it is really quite jagged. For example, between September 1965 and April 1967 (observations 200 to 300,) the movements in the log of variance included a rise of 3 and a subsequent drop of about 2, suggesting that variance rose by a factor of 20 and fell by a factor of 7.4 in the space of about a year and a half. It seems, therefore, that the assumption of approximately constant volatility over periods of even a month or so may not be a good one.

\(^2\) This conversion to annual units must be regarded as an approximation only, since it assumes that variance increases linearly with time, which is only true if there is no serial correlation in returns.
Raw data series: $\ln(\epsilon r^2)$

FIGURE II
FIGURE III
ARMA(1, 1) model

FIGURE IV
The next thing we should notice about the log variance series is that it does not seem to exhibit conditional heteroskedasticity—i.e. big moves do not seem to be followed by big moves and small moves by small moves. This is in line with our specification test results, and is favorable evidence that the process for variance is log linear. This is in contrast to what we see in figure 4, which gives the smoothed variance series using the ARMA(1,1) model. This series does seem conditionally heteroskedastic, and also quite volatile.

It is also instructive to compare the variance estimator with alternative variance estimators. In figure 6 we plot the smoothed annual variance from the ARMA(1,1) model versus variance estimates obtained by dividing the returns series into blocks of 25 4-day intervals and then estimating the variance (assumed constant) within each interval. The two series track each other quite closely, although there are some anomalies, most notably in late 1973, when the filtered estimates show a very large rise in variance while the interval estimates show a much smaller rise. The two series are impossible to compare directly, however, since both contain estimation error, and the two estimation errors are related to each other in a highly non-linear way.
ARMA(1,1) model

Figure VI
The Kalman Filter equations also provide us with an optimal linear forecast of log variance, and we can again use the approximations given by equations 5.2 and 5.3 to forecast volatility and risk premia. An expectations theory of the term structure of risk premia might suggest, for example, that the discount rate applied for k periods in the future equals the one-period interest rate expected to prevail k periods in the future. We might, therefore, be interested in forecasting future risk premia. In figure 7, we plot the expected value of the future risk premium using the estimated coefficients for the ARMA(1,1) model. The top curve traces the path of the expected future risk premium given that the current risk premium is two standard deviations above its mean. The lower curve is similar, except that the starting point is a risk premium two standard deviations below the mean. The middle curve starts at the mean. From this figure it appears that most of a shock to the risk premium is expected to die out within a year.
Forecast Risk Premia

expected risk premium (%)

Dots and dashes: forecast starting from mean risk premium.
Solid line: forecast starting two standard deviations above mean.
Dashes: forecast starting from two standard deviations below mean.

months

FIGURE VII
VI. EXTENSIONS

The state space setup developed thus far can easily be extended in a number of interesting ways. We could, for example, introduce other variables into the state space system. There has, for example, been attention given to the interaction of trading volume and price volatility (Huffman (1987), French and Roll (1984)). Given a specification for the dynamic relation between trading and volume, it would be a straightforward exercise to add a volume measure as another state variable in our model.

The model we have presented can also be extended to model the CAPM with changing covariances along the lines of the recent work by Lillien, Engle and Wooldridge (1986), who modeled changing asset covariances with ARCH. If we are willing to assume, 1) a linear, time invariant capital market line 2) constant instantaneous correlation between the market returns and the return on a given portfolio i and 3) that the one-period CAPM holds, then it is easy to show that the risk premium on portfolio i is linear in the instantaneous standard deviation of the return on portfolio i. This would allow us to apply the methods developed in this paper for portfolios other than the market. Carrying this out remains to be done.
VII. CONCLUSION

What do we learn from the model? First of all, we confirmed the strong relation found by Black (1976) and Christie (1982) between leverage and volatility. This effect is very strong, and highly statistically significant. We also found nothing to contradict the finding of Poterba and Summers (1986) that shocks to volatility die out fairly quickly. It may well be, however, that in a data set spanning only twenty three years we will not be able to detect the existence of a stochastic trend in variance.

We found a significant relation between risk and return, and found no evidence against conditional normality of returns. The biggest failing of the model is its inability to account for serial correlation in market returns. Perhaps this problem could be addressed by allowing more general dependence of conditional mean on conditional variance. This possibility is left for future research.
APPENDIX:

Proof of Theorem 2.1

The algebra is completely straightforward when we use the following facts:

A.1) If $X \sim N(\mu, \sigma^2)$, then

$$E[\exp(X)] = \exp(\mu + \sigma^2/2)$$

A.2) If $Z \sim N(0,1)$, then

$$E[Z\exp(\theta Z)] = \theta \exp(\theta^2/2)$$

QED.

Functional Forms for $J_0$, $K_0$, $Q'$, $Q^{-}$ and the elements of $\Sigma$:

$$3.15) \quad I = \begin{bmatrix} \sigma_y^2 (C,b) & 0 & \gamma_1 (C,b) & \gamma_2 (C,b) \\ 0 & \sigma_z^2 & 0 & 0 \\ \gamma_1 (C,b) & 0 & \sigma_y^2 (C,b) & 0 \\ \gamma_2 (C,b) & 0 & 0 & 1 - \sigma_y^2 \end{bmatrix}$$

$$3.17) \quad J_0 = J_0 (C,b) = \frac{(Q^{-}) + (Q')}{(Q^{-}) - (Q')}$$

$$3.18) \quad K_0 = K_0 (C,b) = -2/[(Q^{-}) - (Q')]$$

$$3.19) \quad Q' = E[(z+b)|S_{(z+b) i}: t = 1]$$

$$Q^{-} = E[(z+b)|S_{(z+b) i}: t = -1]$$

Define $X$ to be a random variable that is distributed GED with mean=$b_1$, variance=1 and tail fatness parameter $C$. Let $f(X)$ be the probability density of $X$, which was given in section IV. Then define

$$A.3) \quad P_1 = \int_0^\infty f(X) dX$$
and $P_2 = 1 - P_1$.

Then $Q' = \left(\frac{1}{P_1}\right) \int_0^\infty X f(X) \, dX$

$Q^- = (b - P_1 Q') / P_2$

$\sigma^2_0 (C, b) = P_1 (Q' - b)^2 + P_2 (Q^- - b)^2$

$E[\ln(X^2)] = \int_{-\infty}^\infty \ln(X^2) f(X) \, dX$

$E[\ln^2(X^2)] = \int_{-\infty}^\infty \ln^2(X^2) f(X) \, dX$

$\sigma^2_v (C, b) = E[\ln^2(X^2)] - E^2[\ln(X^2)]$

$\Gamma_1(C, b) = -b \ E[\ln(X^2)] + P_1 \int_0^{Q'} \ln(X^2) f(X) \, dX$

$+ P_2 \int_{-\infty}^{Q^-} \ln(X^2) f(X) \, dX$

$\Gamma_2(C, b) = -b \ [\ln(X^2)] - \Gamma_1 - \int_{-\infty}^\infty X \ln(X^2) f(X) \, dX$

Since several of the above integrals are improper (i.e. $\ln(X^2) = -\infty$ when $X^2 = 0$), it was necessary to truncate the bound near zero. After some experimentation, it was found that truncating the bound nearest zero at about $\pm 10^{-9}$ gave a very close approximation to the correct answer. Since
\( \ln(X^2) \) changes rapidly near this lower bound, it was also necessary to break up the integrals into several pieces. i.e., the functions were integrated from \( 10^{-9} \) to \( 10^{-8} \), from \( 10^{-8} \) to \( 10^{-7} \) and so on until about \( 10^{-2} \). After this point, larger blocks were used.
References:


Generating?," Mimeo.


TABLE I:
Covariance and Specification test results for AR(1) with normality:

<table>
<thead>
<tr>
<th>parameter estimate</th>
<th>t-statistic</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta$</td>
<td>0.96551800</td>
<td>51.09560634</td>
</tr>
<tr>
<td>$\theta$</td>
<td>-0.16116300</td>
<td>-3.49282065</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-9.70734700</td>
<td>-53.88021932</td>
</tr>
<tr>
<td>$b$</td>
<td>0.27962000</td>
<td>0.96829283</td>
</tr>
<tr>
<td>$\sigma^2_e$</td>
<td>0.00345203</td>
<td>0.38262320</td>
</tr>
<tr>
<td>$b_0$</td>
<td>0.07444600</td>
<td>2.66553968</td>
</tr>
</tbody>
</table>

The asymptotic covariance matrix of the parameter estimates is

\[
\begin{bmatrix}
0.00035707 & 0.00077679 & 0.00074705 \\
-0.00032835 & -0.00012702 & 2.24151935E-005 \\
0.00077679 & 0.00212901 & 0.00174879 \\
-0.00120694 & -0.00024033 & 3.93791524E-005 \\
0.00074705 & 0.00174879 & 0.03245954 \\
-0.03309574 & -0.00024527 & -0.00250431 \\
-0.00032835 & -0.00120694 & -0.03309574 \\
0.08339174 & -9.31948940E-005 & 0.00031878 \\
-9.31948940E-005 & 8.13966492E-005 & -1.85195487E-005 \\
2.24151935E-005 & 3.93791524E-005 & -0.00250431 \\
0.00031878 & -1.85195487E-005 & 0.00078003
\end{bmatrix}
\]

Specification test statistic = 41.4823564
which is asymptotically chi square with 19 degrees of freedom.

<table>
<thead>
<tr>
<th>test#</th>
<th>test statistics</th>
<th>standard errors</th>
<th>t statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00184319</td>
<td>0.03596250</td>
<td>0.05125323</td>
</tr>
<tr>
<td>2</td>
<td>-0.00553222</td>
<td>0.02604738</td>
<td>-0.21239067</td>
</tr>
<tr>
<td>3</td>
<td>-0.04486690</td>
<td>0.03758390</td>
<td>-1.19377971</td>
</tr>
<tr>
<td>4</td>
<td>0.13103606</td>
<td>0.02610298</td>
<td>5.01996585</td>
</tr>
<tr>
<td>5</td>
<td>-0.01064890</td>
<td>0.03161658</td>
<td>-0.33681375</td>
</tr>
<tr>
<td>6</td>
<td>-0.03661302</td>
<td>0.02614490</td>
<td>-1.40038867</td>
</tr>
<tr>
<td>7</td>
<td>-0.03814814</td>
<td>0.03776592</td>
<td>-1.01012083</td>
</tr>
<tr>
<td>8</td>
<td>0.00999194</td>
<td>0.02608511</td>
<td>0.38305138</td>
</tr>
<tr>
<td>9</td>
<td>0.00515678</td>
<td>0.02998926</td>
<td>0.17195415</td>
</tr>
<tr>
<td>10</td>
<td>0.02288076</td>
<td>0.02636315</td>
<td>0.86790689</td>
</tr>
<tr>
<td>11</td>
<td>-0.01688727</td>
<td>0.03689442</td>
<td>-0.45771864</td>
</tr>
<tr>
<td>12</td>
<td>0.04312726</td>
<td>0.02615089</td>
<td>1.64916978</td>
</tr>
<tr>
<td>13</td>
<td>-0.00914841</td>
<td>0.02903525</td>
<td>-0.31507947</td>
</tr>
<tr>
<td>14</td>
<td>0.02953423</td>
<td>0.02636385</td>
<td>1.12025505</td>
</tr>
<tr>
<td>15</td>
<td>-0.01407743</td>
<td>0.03530697</td>
<td>-0.39871539</td>
</tr>
<tr>
<td>16</td>
<td>0.04075423</td>
<td>0.02610234</td>
<td>1.56132498</td>
</tr>
</tbody>
</table>
TABLE II:
Covariance Results for ARMA(1,1) with normality

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>T-statistic</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta$</td>
<td>0.9740600</td>
<td>82.23195036</td>
<td>0.01184527</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>-0.30023200</td>
<td>-4.03258157</td>
<td>0.07445156</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-10.18725500</td>
<td>-23.00682159</td>
<td>0.44279280</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>0.30582500</td>
<td>1.03831245</td>
<td>0.29454043</td>
</tr>
<tr>
<td>$\sigma^2_\epsilon$</td>
<td>0.00594719</td>
<td>0.23198755</td>
<td>0.02563580</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.07650900</td>
<td>2.85368205</td>
<td>0.02681063</td>
</tr>
<tr>
<td>$\psi$</td>
<td>-0.54696400</td>
<td>-3.43445814</td>
<td>0.15925773</td>
</tr>
</tbody>
</table>

The asymptotic covariance matrix of the parameter estimates is

\[
\begin{bmatrix}
0.00014031 & 2.54543300E-005 & -0.00288612 \\
-0.00019376 & -0.00015359 & -1.54974698E-005 \\
-0.00117124 & & \\
2.54543300E-005 & 0.00554304 & 0.01469791 \\
-0.00057899 & -0.00019348 & 0.00019606 \\
0.00785104 & & \\
-0.00288612 & 0.01469791 & 0.19606547 \\
-0.03250503 & 0.00271843 & -0.00609289 \\
0.04931220 & & \\
-0.00019376 & -0.00057899 & -0.03250503 \\
0.08675407 & -0.00055116 & 0.00038091 \\
0.00208615 & & \\
-0.00015359 & -0.00019348 & 0.00271843 \\
-0.00055116 & 0.00065719 & -4.80396625E-005 \\
0.00055822 & & \\
-1.54974698E-005 & 0.00019606 & -0.00609289 \\
0.00038091 & -4.80396625E-005 & 0.00071881 \\
0.00046880 & & \\
-0.00117124 & 0.00785104 & 0.04931220 \\
0.00208615 & 0.00055822 & 0.00046880 \\
0.02536302 & & 
\end{bmatrix}
\]

Specification test statistic = 40.49216852
which is asymptotically chi square with 19 degrees of freedom.

<table>
<thead>
<tr>
<th>Test</th>
<th>Test Statistics</th>
<th>Standard Errors</th>
<th>T Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00199250</td>
<td>0.03419866</td>
<td>0.05826260</td>
</tr>
<tr>
<td>2</td>
<td>-0.00542017</td>
<td>0.02605349</td>
<td>-0.20804010</td>
</tr>
<tr>
<td>3</td>
<td>0.00234509</td>
<td>0.05208867</td>
<td>0.04502111</td>
</tr>
<tr>
<td>4</td>
<td>0.12925105</td>
<td>0.02620723</td>
<td>4.93188611</td>
</tr>
<tr>
<td>5</td>
<td>-0.00864405</td>
<td>0.03124858</td>
<td>-0.27662199</td>
</tr>
<tr>
<td>6</td>
<td>-0.03434788</td>
<td>0.02611659</td>
<td>-1.31517509</td>
</tr>
<tr>
<td>7</td>
<td>-0.04442204</td>
<td>0.03309272</td>
<td>-1.34235096</td>
</tr>
<tr>
<td>8</td>
<td>0.01012475</td>
<td>0.02606027</td>
<td>0.38851287</td>
</tr>
<tr>
<td>9</td>
<td>0.00509274</td>
<td>0.02887481</td>
<td>0.17637319</td>
</tr>
<tr>
<td>10</td>
<td>0.02556700</td>
<td>0.02636117</td>
<td>0.96987342</td>
</tr>
<tr>
<td>11</td>
<td>-0.02939811</td>
<td>0.03448412</td>
<td>-0.85251158</td>
</tr>
<tr>
<td>12</td>
<td>0.04340651</td>
<td>0.02611007</td>
<td>1.66244330</td>
</tr>
<tr>
<td>13</td>
<td>-0.00303880</td>
<td>0.02867470</td>
<td>-0.10597499</td>
</tr>
<tr>
<td>14</td>
<td>0.02754449</td>
<td>0.02632954</td>
<td>1.04614432</td>
</tr>
<tr>
<td>15</td>
<td>-0.02388714</td>
<td>0.03347226</td>
<td>-0.71363968</td>
</tr>
<tr>
<td>16</td>
<td>0.04108972</td>
<td>0.02607160</td>
<td>1.57603398</td>
</tr>
<tr>
<td>17</td>
<td>5.38158929</td>
<td>4.74973181</td>
<td>1.13303014</td>
</tr>
<tr>
<td>18</td>
<td>-0.53895887</td>
<td>0.96863563</td>
<td>-0.55641033</td>
</tr>
<tr>
<td>19</td>
<td>0.27298656</td>
<td>0.14766176</td>
<td>1.84872888</td>
</tr>
</tbody>
</table>

**TABLE III:**
Covariance results for ARMA(2,1) with normality

<table>
<thead>
<tr>
<th>parameter estimates</th>
<th>t-statistics</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta_1$</td>
<td>0.97744700</td>
<td>124.05775074</td>
</tr>
<tr>
<td>$\theta$</td>
<td>-0.29806800</td>
<td>-4.65887070</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-10.84705700</td>
<td>-19.54341975</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.29557700</td>
<td>1.00402032</td>
</tr>
<tr>
<td>$\sigma^2_\epsilon$</td>
<td>0.00441135</td>
<td>0.18274649</td>
</tr>
<tr>
<td>$\psi$</td>
<td>-0.80095200</td>
<td>-7.03518015</td>
</tr>
<tr>
<td>$b$</td>
<td>0.05740100</td>
<td>1.48785680</td>
</tr>
<tr>
<td>$\Delta_2$</td>
<td>0.50961500</td>
<td>2.64253593</td>
</tr>
</tbody>
</table>

Where $(1 - \Delta_1 L - \Delta_2 L^2) = (1 - \Delta_1 L)(1 - \Delta_2 L)$.

The asymptotic covariance matrix of the parameter estimates is

<table>
<thead>
<tr>
<th>6.20781306E-005</th>
<th>2.68881659E-005</th>
<th>-0.00014196</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.00023397</td>
<td>-6.03181791E-005</td>
<td>-0.00046602</td>
</tr>
<tr>
<td>-0.00018374</td>
<td>0.00047790</td>
<td></td>
</tr>
<tr>
<td>2.68881659E-005</td>
<td>0.00409326</td>
<td>0.00621074</td>
</tr>
<tr>
<td>9.21156053E-006</td>
<td>-0.00015901</td>
<td>0.00247763</td>
</tr>
<tr>
<td>0.00054845</td>
<td>0.00024933</td>
<td></td>
</tr>
<tr>
<td>-0.00014196</td>
<td>0.00621074</td>
<td>0.30805107</td>
</tr>
<tr>
<td>-0.03284404</td>
<td>0.0031479</td>
<td>0.01233194</td>
</tr>
<tr>
<td>-0.000775432</td>
<td>-0.01608448</td>
<td></td>
</tr>
<tr>
<td>-0.00023397</td>
<td>9.21156053E-006</td>
<td>-0.03284404</td>
</tr>
<tr>
<td>0.08666750</td>
<td>-0.00057953</td>
<td>0.00387965</td>
</tr>
<tr>
<td>0.00120398</td>
<td>-0.00495216</td>
<td></td>
</tr>
<tr>
<td>-6.03181791E-005</td>
<td>-0.00015901</td>
<td>0.0031479</td>
</tr>
<tr>
<td>-0.00057953</td>
<td>0.00058270</td>
<td>0.0010271</td>
</tr>
<tr>
<td>0.00011397</td>
<td>-0.00042906</td>
<td></td>
</tr>
<tr>
<td>-0.00046602</td>
<td>0.00247763</td>
<td>0.01233194</td>
</tr>
<tr>
<td>0.00387965</td>
<td>0.00010271</td>
<td>0.01296172</td>
</tr>
<tr>
<td>0.000351806</td>
<td>-0.01873002</td>
<td></td>
</tr>
</tbody>
</table>
The specification test statistic = 39.29930951 \cdot x_{19}^2, 

<table>
<thead>
<tr>
<th>test</th>
<th>statistics</th>
<th>standard errors</th>
<th>t statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00011205</td>
<td>0.03474733</td>
<td>0.00322458</td>
</tr>
<tr>
<td>2</td>
<td>-0.00332352</td>
<td>0.02616432</td>
<td>-0.12702479</td>
</tr>
<tr>
<td>3</td>
<td>0.00575158</td>
<td>0.04952318</td>
<td>0.11613915</td>
</tr>
<tr>
<td>4</td>
<td>0.12986239</td>
<td>0.02613569</td>
<td>4.96877557</td>
</tr>
<tr>
<td>5</td>
<td>-0.00965233</td>
<td>0.03063628</td>
<td>-0.31506220</td>
</tr>
<tr>
<td>6</td>
<td>-0.03385831</td>
<td>0.02607283</td>
<td>-1.29860494</td>
</tr>
<tr>
<td>7</td>
<td>-0.01659430</td>
<td>0.03993240</td>
<td>-0.41555969</td>
</tr>
<tr>
<td>8</td>
<td>0.00965158</td>
<td>0.02603951</td>
<td>0.37065132</td>
</tr>
<tr>
<td>9</td>
<td>0.00655096</td>
<td>0.02824795</td>
<td>0.23190908</td>
</tr>
<tr>
<td>10</td>
<td>0.02474842</td>
<td>0.02623451</td>
<td>0.94335367</td>
</tr>
<tr>
<td>11</td>
<td>-0.01564749</td>
<td>0.03655225</td>
<td>-0.42808558</td>
</tr>
<tr>
<td>12</td>
<td>0.04399497</td>
<td>0.02612896</td>
<td>1.68376243</td>
</tr>
<tr>
<td>13</td>
<td>-0.00227114</td>
<td>0.02764777</td>
<td>-0.08214551</td>
</tr>
<tr>
<td>14</td>
<td>0.02842299</td>
<td>0.02624049</td>
<td>1.08317275</td>
</tr>
<tr>
<td>15</td>
<td>-0.02094846</td>
<td>0.03353987</td>
<td>-0.62458388</td>
</tr>
<tr>
<td>16</td>
<td>0.04200669</td>
<td>0.02606146</td>
<td>1.61183209</td>
</tr>
<tr>
<td>17</td>
<td>5.16490637</td>
<td>4.55337336</td>
<td>1.13430329</td>
</tr>
<tr>
<td>18</td>
<td>-0.49394440</td>
<td>0.94871220</td>
<td>-0.52064725</td>
</tr>
<tr>
<td>19</td>
<td>0.25350097</td>
<td>0.14886735</td>
<td>1.70286480</td>
</tr>
</tbody>
</table>

TABLE IV:
Covariance Results for AR(1) without Normality

<table>
<thead>
<tr>
<th>parameter estimate</th>
<th>t-statistic</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma )</td>
<td>0.96696800</td>
<td>53.61416916</td>
</tr>
<tr>
<td>( \theta )</td>
<td>-0.14677500</td>
<td>-2.56615902</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>-9.83836200</td>
<td>-20.80556266</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.27750100</td>
<td>0.95834768</td>
</tr>
<tr>
<td>( \sigma_e^2 )</td>
<td>0.00562290</td>
<td>0.51316441</td>
</tr>
<tr>
<td>( \omega )</td>
<td>0.08244000</td>
<td>2.14683342</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>1.82524300</td>
<td>3.91536517</td>
</tr>
</tbody>
</table>

The asymptotic covariance matrix of the parameter estimates is

\[
\begin{bmatrix}
0.00032529 & 0.00080827 & -0.00118046 \\
-0.00050190 & -9.73907529E-005 & 0.00012246 \\
-0.00189977 & 0.00080827 & -0.01600549
\end{bmatrix}
\]
\[-0.00280911 \quad 1.33418065E-005 \quad 0.00100205\]
\[-0.01841180\]
\[-0.00118046 \quad -0.01600549 \quad 0.22360765\]
\[-0.01639662 \quad -0.00305216 \quad -0.01349334\]
\[0.20490804\]
\[-0.00050190 \quad -0.00280911 \quad -0.01639662\]
\[0.08384611 \quad -0.00027710 \quad -0.00071036\]
\[0.01932913\]
\[-9.73907529E-005 \quad 1.33418065E-005 \quad -0.00305216\]
\[-0.00027710 \quad 0.00012006 \quad 0.00013851\]
\[-0.00304507\]
\[0.00012246 \quad 0.00100205 \quad -0.01349334\]
\[-0.00071036 \quad 0.00013851 \quad 0.00147462\]
\[-0.01134886\]
\[-0.00189977 \quad -0.01841180 \quad 0.20490804\]
\[0.01932913 \quad -0.00304507 \quad -0.01134886\]
\[0.21731857\]

Specification test statistic = $40.25469177 \sim x^2_{19}$

<table>
<thead>
<tr>
<th>test statistics</th>
<th>standard errors</th>
<th>t statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00391571</td>
<td>0.03405140</td>
</tr>
<tr>
<td>2</td>
<td>-0.00537404</td>
<td>0.02573857</td>
</tr>
<tr>
<td>3</td>
<td>-0.04556461</td>
<td>0.03606829</td>
</tr>
<tr>
<td>4</td>
<td>0.13084352</td>
<td>0.02611814</td>
</tr>
<tr>
<td>5</td>
<td>-0.00846545</td>
<td>0.03203286</td>
</tr>
<tr>
<td>6</td>
<td>-0.03604766</td>
<td>0.02590572</td>
</tr>
<tr>
<td>7</td>
<td>-0.03906207</td>
<td>0.03629795</td>
</tr>
<tr>
<td>8</td>
<td>0.00971789</td>
<td>0.02609214</td>
</tr>
<tr>
<td>9</td>
<td>0.00677919</td>
<td>0.02931764</td>
</tr>
<tr>
<td>10</td>
<td>0.02279387</td>
<td>0.02605770</td>
</tr>
<tr>
<td>11</td>
<td>-0.01791275</td>
<td>0.03646051</td>
</tr>
<tr>
<td>12</td>
<td>0.04296764</td>
<td>0.02613035</td>
</tr>
<tr>
<td>13</td>
<td>-0.00735682</td>
<td>0.02838076</td>
</tr>
<tr>
<td>14</td>
<td>0.02935041</td>
<td>0.02602360</td>
</tr>
<tr>
<td>15</td>
<td>-0.01506248</td>
<td>0.03450211</td>
</tr>
<tr>
<td>16</td>
<td>0.04060257</td>
<td>0.02609363</td>
</tr>
<tr>
<td>17</td>
<td>5.56158093</td>
<td>4.81531274</td>
</tr>
<tr>
<td>18</td>
<td>-0.58687760</td>
<td>0.99997481</td>
</tr>
<tr>
<td>19</td>
<td>0.28627376</td>
<td>0.14839766</td>
</tr>
</tbody>
</table>
TABLE V:

Covariance Results for ARMA(1,1) without Normality

<table>
<thead>
<tr>
<th>parameter</th>
<th>estimate</th>
<th>t-statistic</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta$</td>
<td>0.97489000</td>
<td>87.88861499</td>
<td>0.01109234</td>
</tr>
<tr>
<td>$\theta$</td>
<td>-0.27727700</td>
<td>-3.35518091</td>
<td>0.08264145</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-10.32823300</td>
<td>-14.57017607</td>
<td>0.70886123</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.30396100</td>
<td>1.02980181</td>
<td>0.29516456</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.01292928</td>
<td>0.38270191</td>
<td>0.03378421</td>
</tr>
<tr>
<td>$\psi$</td>
<td>-0.54773400</td>
<td>-3.52507163</td>
<td>0.15538237</td>
</tr>
<tr>
<td>$b$</td>
<td>0.08357900</td>
<td>2.30920059</td>
<td>0.03619391</td>
</tr>
<tr>
<td>$C$</td>
<td>1.83854000</td>
<td>3.91643561</td>
<td>0.46944216</td>
</tr>
</tbody>
</table>

The asymptotic covariance matrix of the parameter estimates is

\[
\begin{pmatrix}
0.00012304 & 0.00015531 & -0.00403035 \\
-0.00027560 & -5.00961127E-005 & -0.00110963 \\
4.03263455E-005 & -0.00121054 & \\
0.00015531 & 0.00682961 & -0.01711483 \\
-0.000310976 & 0.00104657 & 0.00463009 \\
0.00133119 & -0.02380043 & \\
-0.00403035 & -0.01711483 & 0.50248425 \\
-0.01513241 & -0.01208462 & 0.06406019 \\
-0.01924612 & 0.26446555 & \\
-0.0027560 & -0.00310976 & -0.01513241 \\
0.08712212 & 0.00123749 & 0.00213820 \\
-0.00048131 & 0.01737765 & \\
-5.00961127E-005 & 0.00104657 & -0.01208462 \\
-0.00123749 & 0.00114137 & -0.0057887 \\
0.00052308 & -0.01196973 & \\
-0.00110963 & 0.00463009 & 0.06406019 \\
0.00213820 & -0.00057887 & 0.02414368 \\
-0.00029912 & 0.01653273 & \\
4.03263455E-005 & 0.00133119 & -0.01924612 \\
-0.00048131 & 0.00052308 & -0.00029912 \\
0.00131000 & -0.01060137 & \\
-0.00121054 & -0.02380043 & 0.26446555 \\
0.01737765 & -0.01196973 & 0.01653273 \\
-0.01060137 & 0.22037594 &
\end{pmatrix}

The specification test statistic = 39.73716664 ~ $\chi^2_{19}$

<table>
<thead>
<tr>
<th>test#</th>
<th>test statistics</th>
<th>standard errors</th>
<th>t statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00346033</td>
<td>0.03237556</td>
<td>0.10688084</td>
</tr>
<tr>
<td>2</td>
<td>-0.00516384</td>
<td>0.02578541</td>
<td>-0.20026200</td>
</tr>
<tr>
<td>3</td>
<td>0.00113105</td>
<td>0.05018082</td>
<td>-0.2253948</td>
</tr>
<tr>
<td>4</td>
<td>0.12914746</td>
<td>0.02620334</td>
<td>4.92866336</td>
</tr>
<tr>
<td>5</td>
<td>-0.00665596</td>
<td>0.03156467</td>
<td>-0.21086736</td>
</tr>
<tr>
<td>6</td>
<td>-0.03383944</td>
<td>0.02592083</td>
<td>-1.30549212</td>
</tr>
<tr>
<td>7</td>
<td>-0.04465017</td>
<td>0.03215887</td>
<td>-1.38842457</td>
</tr>
<tr>
<td></td>
<td>parameter estimate</td>
<td>t statistic</td>
<td>standard error</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>8</td>
<td>0.00985587</td>
<td>0.02607423</td>
<td>0.37799275</td>
</tr>
<tr>
<td>9</td>
<td>0.00668339</td>
<td>0.02824206</td>
<td>0.23664676</td>
</tr>
<tr>
<td>10</td>
<td>0.02547052</td>
<td>0.02608565</td>
<td>0.97641906</td>
</tr>
<tr>
<td>11</td>
<td>-0.02965311</td>
<td>0.03363766</td>
<td>-0.88154504</td>
</tr>
<tr>
<td>12</td>
<td>0.04324873</td>
<td>0.02609804</td>
<td>1.65716400</td>
</tr>
<tr>
<td>13</td>
<td>-0.00144281</td>
<td>0.02809627</td>
<td>-0.05135248</td>
</tr>
<tr>
<td>14</td>
<td>0.02747325</td>
<td>0.02603573</td>
<td>1.05521333</td>
</tr>
<tr>
<td>15</td>
<td>-0.02424691</td>
<td>0.03262309</td>
<td>-0.74324398</td>
</tr>
<tr>
<td>16</td>
<td>0.04094065</td>
<td>0.02607034</td>
<td>1.57039191</td>
</tr>
<tr>
<td>17</td>
<td>5.31107021</td>
<td>4.77683096</td>
<td>1.11183968</td>
</tr>
<tr>
<td>18</td>
<td>-0.51111022</td>
<td>0.98362412</td>
<td>-0.51961945</td>
</tr>
<tr>
<td>19</td>
<td>0.27384275</td>
<td>0.14722015</td>
<td>1.86009013</td>
</tr>
</tbody>
</table>

**TABLE VI:**
Covariance results for ARMA(2,1) without normality.

parameter estimate  | t statistic  | standard error
\[ \Delta_1^* \] 0.97763900 | 131.87967695 | 0.00741311
\[ \theta \] -0.28093700 | -3.62230616 | 0.07755750
\[ \alpha \] -10.94432000 | -14.97967845 | 0.73061114
\[ \beta \] 0.29505800 | 1.00038709 | 0.29494383
\[ \sigma_e^2 \] 0.00966820 | 0.27806800 | 0.03476919
\[ \phi \] -0.79635400 | -6.87827503 | 0.11577816
\[ \sigma \] 0.06200800 | 1.34598370 | 0.04606891
\[ \Delta_2^* \] 0.50309700 | 2.52928892 | 0.19890847
\[ C \] 1.87839200 | 3.87180893 | 0.48514584

where \((1 - \Delta_1 L - \Delta_2 L^2) \approx (1 - \Delta_1^* L)(1 - \Delta_2^* L)\)

The asymptotic covariance matrix of the parameter estimates is

\[
\begin{bmatrix}
5.49542502E-005 & 1.43135925E-005 & -0.00010151 \\
-0.00022330 & -4.73876537E-005 & -0.00042158 \\
-0.00017675 & 0.00040871 & 3.09841453E-005 \\
\end{bmatrix}
\]
\begin{tabular}{ccc}
-0.00042158 & 0.00349745 & 0.00191161 \\
0.00283571 & 0.00062121 & 0.01340458 \\
0.00413544 & -0.01989075 & -0.01037690 \\
-0.00017675 & 0.00176110 & -0.01890962 \\
0.00048998 & 0.00069822 & 0.00413544 \\
0.00212234 & -0.00650662 & -0.01097412 \\
0.00040871 & -0.00319748 & 0.00904254 \\
-0.00328995 & -0.00177123 & -0.01989075 \\
-0.00650662 & 0.03956458 & 0.02686262 \\
3.09841453E-005 & -0.02533331 & 0.23885356 \\
0.01469339 & -0.01340406 & -0.01037690 \\
-0.01097412 & 0.02686262 & 0.23536649 \\
\end{tabular}

Specification test statistic = \( 38.60656820^{*} \times x^{2}_{19} \).

<table>
<thead>
<tr>
<th>test statistics</th>
<th>standard errors</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00155696</td>
<td>0.03305986</td>
</tr>
<tr>
<td>2</td>
<td>-0.00322802</td>
<td>0.02594529</td>
</tr>
<tr>
<td>3</td>
<td>0.00465297</td>
<td>0.04789280</td>
</tr>
<tr>
<td>4</td>
<td>0.12971082</td>
<td>0.02614129</td>
</tr>
<tr>
<td>5</td>
<td>-0.00820201</td>
<td>0.03096523</td>
</tr>
<tr>
<td>6</td>
<td>-0.03347889</td>
<td>0.02596618</td>
</tr>
<tr>
<td>7</td>
<td>-0.01743927</td>
<td>0.03874230</td>
</tr>
<tr>
<td>8</td>
<td>0.00944702</td>
<td>0.02605059</td>
</tr>
<tr>
<td>9</td>
<td>0.00760489</td>
<td>0.02775681</td>
</tr>
<tr>
<td>10</td>
<td>0.02469799</td>
<td>0.02602915</td>
</tr>
<tr>
<td>11</td>
<td>-0.01616061</td>
<td>0.03632600</td>
</tr>
<tr>
<td>12</td>
<td>0.04381347</td>
<td>0.02612203</td>
</tr>
<tr>
<td>13</td>
<td>-0.00120121</td>
<td>0.02727070</td>
</tr>
<tr>
<td>14</td>
<td>0.02835972</td>
<td>0.02601138</td>
</tr>
<tr>
<td>15</td>
<td>-0.02116429</td>
<td>0.03294259</td>
</tr>
<tr>
<td>16</td>
<td>0.04181618</td>
<td>0.02607382</td>
</tr>
<tr>
<td>17</td>
<td>5.12334270</td>
<td>4.57101769</td>
</tr>
<tr>
<td>18</td>
<td>-0.47500632</td>
<td>0.96156918</td>
</tr>
<tr>
<td>19</td>
<td>0.25486362</td>
<td>0.14821772</td>
</tr>
</tbody>
</table>

73
<table>
<thead>
<tr>
<th>C</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 (double exponential)</td>
<td>0.23</td>
<td>6.58</td>
</tr>
<tr>
<td>1.25</td>
<td>-0.44</td>
<td>5.89</td>
</tr>
<tr>
<td>1.5</td>
<td>-0.83</td>
<td>5.45</td>
</tr>
<tr>
<td>1.75</td>
<td>-1.09</td>
<td>5.15</td>
</tr>
<tr>
<td>2.0 (normal)</td>
<td>-1.27</td>
<td>π² ~ 4.93</td>
</tr>
<tr>
<td>2.25</td>
<td>-1.40</td>
<td>4.78</td>
</tr>
<tr>
<td>2.5</td>
<td>-1.50</td>
<td>4.66</td>
</tr>
<tr>
<td>3.0</td>
<td>-1.63</td>
<td>4.49</td>
</tr>
<tr>
<td>3.5</td>
<td>-1.71</td>
<td>4.38</td>
</tr>
<tr>
<td>4.0</td>
<td>-1.77</td>
<td>4.30</td>
</tr>
<tr>
<td>5.0</td>
<td>-1.84</td>
<td>4.20</td>
</tr>
<tr>
<td>6.0</td>
<td>-1.88</td>
<td>4.15</td>
</tr>
<tr>
<td>∞ (uniform)</td>
<td>-2.00</td>
<td>4.00</td>
</tr>
</tbody>
</table>
CHAPTER TWO:

STATIONARITY AND PERSISTENCE IN THE GARCH(1,1) MODEL

April 22, 1988

Summary: This paper establishes necessary and sufficient conditions for stationarity and ergodicity of the GARCH(1,1) process. As a special case, it is shown that the IGARCH(1,1) with no drift converges almost surely to zero, while IGARCH(1,1) with a positive drift possesses a strictly stationary and ergodic limiting distribution. We also examine the persistence of shocks in the GARCH(1,1) model, and show that whether shocks "persist" or not depends crucially on the definition of persistence. We develop conditions for transience and persistence of shocks almost surely, in probability, and in $L^p$. 

75
I. INTRODUCTION

Since the seminal work of Engle (1982), ARCH (autoregressive conditionally heteroskedastic) models have been widely used to model time-varying volatility and the persistence of shocks to volatility. One member of the family of ARCH processes, GARCH(1,1) has, since its introduction by Bollerslev (1986), been especially popular in econometric modelling (see, for example, Bollerslev and Engle (1986), French, Schwert and Stambaugh (1987), Chou (1987) and Mustafa (1987)).

The error process for the GARCH(1,1) model is generally written as:

1.1) \( \xi_t \mid \psi_{t-1} \sim N(0, \sigma^2_t) \)

1.2) \( \sigma^2_t = \omega + \beta \sigma^2_{t-1} + \alpha \xi_{t-1}^2 \)

where \( \psi_t \) denotes the information set at time \( t \), and \( \omega, \beta \) and \( \alpha \) are non-negative real numbers, with \( \alpha \) strictly positive. In applications, \( \xi_t \) is generally an innovation in a time series regression. For example, a simple model of the realized risk premium on an asset could be given by (1.1), (1.2) and

1.3) \( R_t = \theta \sigma^2_t + \xi_t \)

In (1.3), the realized excess return (\( R_t \)) on an asset is equal to a zero-mean error term \( \xi_t \) and an ex ante risk
premium assumed proportional to $\sigma^2_t$.

Models such as (1.1)-(1.3) have a number of appealing features. First, the likelihood function is simple to compute, so that estimation is straightforward. Forecasting is also easily carried out, with

1.4) $E(\sigma^2_{t+k} | \sigma^2_t) = (\delta + \alpha)^k \sigma^2_t + \omega [ \sum_{i=0}^{k-1} (\delta + \alpha)^i ], \ k \geq 1$

Finally, the model seems to provide an easy interpretation of the persistence of shocks. Bollerslev (1986) showed that when $\alpha + \delta < 1$, the $\xi_t$ process is covariance stationary. But when $\alpha + \delta = 1$, we have

1.5) $E(\sigma^2_{t+k} | \sigma^2_t) = \sigma^2_t + \omega k, \ k \geq 1$

When $\omega = 0$, the conditional variance process $\sigma^2_t$ is a martingale; that is, for all $k \geq 1$

1.6) $E(\sigma^2_{t+k} | \sigma^2_t) = \sigma^2_t$.

GARCH(1,1) models with $\alpha + \delta = 1$ are known as IGARCH(1,1) models (Integrated Autoregresssive Conditional Heteroskedasticity.) Insofar as the behavior of the first moment is concerned, cases (1.5) and (1.6) are analogous to a random walk with and without drift respectively. When $\alpha + \delta = 1$, shocks to $\sigma^2_t$ persist indefinitely in $L^1$. The sum $\alpha + \delta$ seems to be a natural index of the persistence of shocks to volatility, and it has been so used by a number of authors (i.e. Engle and Bollerslev (1986), Chou (1987).
French, Schwert and Stambaugh (1987)).

However Geweke (1986) and Bollerslev and Engle (1986b) showed that the structure of the higher moments of $\sigma^2_{t}$ when $\alpha + \beta = 1$ and $\omega = 0$ implies that the mass of $\sigma^2_{t}$ becomes more and more concentrated around zero with fatter and fatter tails. The behavior of $\sigma^2_{t}$ is therefore not exactly analogous to a random walk.

In the remainder of this paper we formalize and extend these results. We will see that in the IGARCH(1,1) model with no drift ($\omega = 0$), $\sigma^2_{t}$ and $\xi_{t}$ converge to zero almost surely. We will also see that IGARCH(1,1) models with drift do not behave like random walks with drift (i.e. they do not diverge to infinity.) In fact, IGARCH(1,1) models with drift are strictly stationary, as are a fairly broad class of models with $\alpha + \beta > 1$. The basic intuition is that while shocks in IGARCH(1,1) models persist in $L^1$, they die out almost surely, so that IGARCH(1,1) models are strictly stationary but not covariance stationary.\(^1\) In section II,

\(^1\) It may seem counter-intuitive that a process could be strictly stationary but not weakly (covariance) stationary. Despite what the names seem to imply, "weak" stationarity is neither necessary nor sufficient for "strict" stationarity. To see that covariance stationarity is not necessary for strict stationarity, consider an iid sequence of Cauchy random variables. To see that covariance stationarity is not sufficient for strict stationarity, consider a independent sequence $\{x_{t}\}$ such that for all $t$, $E(x_{t}) = 0$ and $E(x_{t}^2) = 1$, but the $x_{t}$'s are drawn from one distribution on even periods and from another distribution on odd periods.
we examine the GARCH(1,1) process with $\omega = 0$ and state
necessary and sufficient conditions for $\sigma_{t}^2$ to go to 0 or $\infty$ almost surely, in probability and in $L^p$. Our results will suggest another model of persistence of shocks, in which $\ln(\sigma_{t}^2)$ follows a simple random walk. In section III, we present our main results on the conditions for stationarity and ergodicity of $\{\sigma_{t}^2, \xi_t\}$ and the persistence of shocks when $\omega > 0$. In section IV, we test different hypotheses about the persistence of shocks using the exchange rate data previously analyzed in Diebold and Mertens (1986) and Engle and Bollerslev (1986). Section V discusses extensions of the results developed in sections I through IV and concludes the paper.
II. GARCH(1,1) WHEN $\omega = 0$:

The system in 1.1 and 1.2 can be written in a form more convenient for the analysis in this paper:

2.1) $\{z_t\}, t=\ldots, \infty \sim \text{iid } N(0,1)$
2.2) $\varepsilon_t = z_t \sigma_t$
2.3) $\sigma^2_t = \omega + (\beta + \alpha z_{t-1}^2) \sigma^2_{t-1}$

It is straightforward to verify that (2.1)-(2.3) imply (1.1)-(1.2). Substituting recursively back to time 0, we have

2.4) $\sigma^2_t = \sigma^2_0 \prod_{i=1}^{t} (\beta + \alpha z_{i-1}^2) + \omega \left[ 1 + \sum_{k=1}^{t} \prod_{i=1}^{k} (\beta + \alpha z_{i-1}^2) \right]$

when $\omega = 0$, this simplifies to

2.5) $\sigma^2_t = \sigma^2_0 \prod_{i=1}^{t} (\beta + \alpha z_{i-1}^2)$

Condition 2.1 can easily be relaxed:

2.1)' $\{z_t\}, t=\ldots, \infty \sim \text{iid, with } E(z_t) = 0 \text{ and } E(z_t^2) = 1$.

To close the system, we assume that $\sigma^2_0$ is a random variable that is strictly positive and finite almost surely, and is independent of $z_t$ for all $t \geq 0$.

When $\beta + \alpha = 1$, $E(\beta + \alpha z_t^2) = 1$, so that in (2.5), $\sigma^2_t$ is a martingale. Nevertheless, we have

Proposition 2.1:
In the IGARCH(1,1) model with no drift (i.e. $\omega=0$), $\sigma^2_t \to 0$ a.s..
Proof:

Raise each side of (2.5) to the $t^{th}$ power and take logs to obtain

$$2.5) \quad (1/t) \ln(\sigma^2_t) = (1/t) \ln(\sigma^2_0) + (1/t) \sum_{i=1}^{t} \ln(\beta + \alpha z_{t-i}^2)$$

As $t \to \infty$, the first term on the right-hand side of (2.5) goes to zero. By Jensen's inequality and the strict concavity on $\ln(x)$, we have

$$2.6) \quad E[\ln(\beta + \alpha z_t^2)] < \ln(E[\beta + \alpha z_t^2]) = 1 \cdot (1) = 0$$

Consider first the case in which $-\infty < E[\ln(\beta + \alpha z_t^2)]$. Then applying the strong law of large numbers,

$$2.7) \quad \frac{1}{t} \sum_{i=1}^{t} \ln(\beta + \alpha z_{t-i}^2) \to E[\ln(\beta + \alpha z_t^2)] \quad a.s.$$.

Combining (2.5)-(2.7) yields

$$2.8) \quad \ln(\sigma^2_t) \to -\infty \quad a.s.$$.

So that, using the relation $\xi_t = \sigma_t z_t$, we have

$$2.9) \quad [\sigma^2_t, \xi_t] \to [0,0] \quad a.s.$$.

The argument for the case $E[\ln(\beta + \alpha z_t^2)] = -\infty$ proceeds similarly: This is quite a special case, which can only occur if $\beta=0$ and $E[\ln(z_t^2)] = -\infty$. First consider the case in which $P[z_t = 0] = \theta > 0$. In this case, $\ln(\sigma^2_t) = -\infty$ for all $t$ after a single zero-valued $z$ is drawn, which happens eventually with probability one, and the proposition follows immediately. Finally, consider the case where $P[z_t = 0] =$
0, but $E[\ln(\alpha z_t^2)] = -\infty$. For $E[\ln(\alpha z_t^2)]$ to exist, it must be the case that $E[\ln(\alpha z_t^2) | \ln(\alpha z_t^2) > 0] < \infty$ and that

2.10) $\lim_{M \to -\infty} E[\ln(\alpha z_t^2) | \ln(\alpha z_t^2) < 0] = -\infty$

We therefore choose an $M$ such that

2.11) $E[\ln(\alpha z_t^2) | \ln(\alpha z_t^2) < 0] = \mu < 0$

2.12) Define $x_t = \max\{ M, \ln(\alpha z_t^2) \}$.

$x_t$ is iid with a finite, negative mean. By the strong law of large numbers and the same argument given in the main proof of the proposition,

2.13) $\sigma^2_t \leq \sigma^2_0 \prod_{i=1}^{t} \exp(x_t) \to 0$ a.s.

so that $\sigma^2_t \to 0$ a.s.. Q.E.D.

Proposition 2.2 is an extension of proposition 2.1:

Proposition 2.2

In the GARCH(1,1) model with $\omega = 0$:

2.14) If $E[\ln(\beta + \alpha z_t^2)] < 0$, then $\sigma^2_t \to 0$ a.s..

2.15) If $E[\ln(\beta + \alpha z_t^2)] > 0$, then $\sigma^2_t \to \infty$ a.s..

Proof: virtually identical to the proof of Proposition 2.1.

Q.E.D.

2) This proof was inspired by a similar argument in Dudley (1987).
To test for convergence or non-convergence, we must therefore be able to evaluate $E[\ln(\beta + \alpha z_t^2)]$. In the case that

$\{z_t\} \sim iid N(0,1)$, we have

**Proposition 2.3**

If $z \sim N(0,1)$, then

$$2.16 \quad E[\ln(\beta + \alpha z^2)] = \left(2\pi \beta / \alpha \right)^{1/2} \; _1F_1(1/2; 3/2; \beta / 2 \alpha) +$$

$$\ln(2\alpha) + \psi(1/2) - (\beta / \alpha) \; _2F_2(1, 1; 2, 3/2; \beta / 2 \alpha)$$

where $\; _1F_1$ is a confluent hypergeometric function, $\; _2F_2$ is a generalized hypergeometric function (Lebedev (1972)) and $\psi()$ is the Euler Psi function (Davis (1965)).

**Proof:** a direct application of Prudnikov et. al. (1986) formula 2.6.23 # 4.

Q.E.D.

Figure 1 uses (2.16) to plot the different regions of convergence for $\sigma^2_t$ when $\{z_t\} \sim N(0,1)$. In regions I and II and on their common boundary (excluding the point $\beta = 1, \alpha = 0$), $\sigma^2_t \to 0$ a.s. when $\psi = 0$. IGARCH(1,1) processes lie on the boundary between regions I and II. In region III, $\sigma^2_t \to \infty$ a.s.. On the boundary, $\ln(\sigma^2_t)$ is a simple random walk, with

$$2.17 \quad \ln(\sigma^2_t) = \ln(\sigma^2_{t-1}) + v_t$$

where $v_t = \ln(\beta + \alpha z_{t-1}^2)$ is an iid, zero mean error term.

The results in this section show that if we do not want
FIGURE 1
conditional variance to go to zero asymptotically. IGARCH(1,1) models with no drift are not an appropriate way to model the persistence of shocks to $\sigma^2_t$. (2.17) suggests that a more appropriate way to model persistent changes in $\sigma^2_t$ may be to choose $\alpha$ and $\beta$ to satisfy $E[\ln(\beta + \alpha z_t^2)] = 0$ and thereby make $\ln(\sigma^2_t)$ a simple random walk.

Confusion about the long run properties of $\{\sigma^2_t, \xi_t\}$ can arise because different notions of convergence yield different answers to the question "do shocks to $\sigma^2_t$ persist?" Just because $\sigma^2_t$ is a martingale does not mean that it will behave at all like a random walk, since shocks may persist in one norm and vanish in another.

To end this section, we will review the main notions of convergence of a random sequence to zero and develop simple criterion for the convergence of $\{\sigma^2_t, \xi_t\}$ to zero in the GARCH(1,1) model with $\omega = 0$.

Definitions:

We say that a random sequence $\{x_t\} \to 0$

a) almost surely

2.18) if $P[|x_t| \to 0] = 1$

b) in probability

2.19) if for each $M > 0$, $\lim_{t \to \infty} P[|x_t| > M] = 0$

c) in $L^p$, $0 < p < \infty$

2.20) if $\lim_{t \to \infty} E[|x_t|^p] = 0$
Similarly, we say that \(|x_t|\) diverges

a) almost surely

2.18)' if \(P[|x_t| \to \infty] = 1\)

b) in probability

2.19)' if for each \(M > 0\), \(\lim_{t \to \infty} P[|x_t| > M] = 1\)

c) in \(L^p\), \(0 < p < \infty\)

2.20)' if \(\lim_{t \to \infty} E[|x_t|^p] = \infty\)

**Proposition 2.4**

Conditionally on \(\sigma^2_0\),

2.21) if \(E[(\beta + \alpha z_t^2)^{p^*}] < 1\)

then \(\{\sigma^2_t, |\xi_t|\} \to [0,0]\) in probability, almost surely and in \(L^p\) for all \(p \leq p^*\).

2.22) if \(E[(\beta + \alpha z_t^2)^{p^*}] > 1\)

then \(\{\sigma^2_t, |\xi_t|\} \to [\infty, \infty]\) in \(L^p\) for all \(p \geq p^*\), but may converge to \([0,0]\) for some \(p < p^*\), in probability and almost surely.

2.23) if \(E[\ln(\beta + \alpha z_t^2)] < 0\)

then \(\{\sigma^2_t, |\xi_t|\} \to [0,0]\) in probability and almost surely, but may diverge in \(L^p\) for some \(p\).

2.24) if \(E[\ln(\beta + \alpha z_t^2)] > 0\)

then \(\{\sigma^2_t\} \to [\infty]\) in probability, almost surely, and in \(L^p\) for every \(p > 0\).

2.25) if \(E[\ln(\beta + \alpha z_t^2)] = 0\)
then \( \{\sigma^2_t, |\xi_t|\} \rightarrow (\infty, \infty) \) in \( L^p \) for every \( p > 0 \), and \( \{\ln(\sigma^2_t)\} \) is a random walk which fails to converge to any finite number or random variable almost surely, in probability or in distribution.

Proof:

2.23): The first part of (2.23) is proposition 2.2. That \( \{\sigma^2_t, |\xi_t|\} \) may nevertheless not converge in \( L^p \) is illustrated by IGARCH(1,1).

2.21), 2.22), 2.24!:

From (2.5) we have

\[
2.26 \quad E[(\sigma^2_t)^p | \sigma^2_0] = (\sigma^2_0)^p \prod_{i=1}^{1} E[(\beta + \alpha z_{i-1}^2)^p] \\
\]

\[
2.27 \quad = (\sigma^2_0)^p (E[(\beta + \alpha z_{i-1}^2)^p])^t
\]

The sum on the right-hand side of (2.27) converges to zero (diverges to \( \infty \)) if \( E[(\beta + \alpha z_{i-1}^2)^p] \) \( \leq 1 \) \((>1)\).

By the integral version of the means inequality (Hardy et al (1934)), if \( p > q > 0 \), then

\[
2.28 \quad \left[ E(\beta + \alpha z_{i}^2)^p \right]^{1/p} > \left[ E(\beta + \alpha z_i^2)^q \right]^{1/q} > \exp[E(\ln(\beta + \alpha z_i^2))]
\]

And

\[
2.29 \quad \left[ E(\beta + \alpha z_{i-1}^2)^p \right]^{1/p} < 1 \text{ iff } E[(\beta + \alpha z_{i-1}^2)^p] < 1.
\]

(2.21), (2.22), and (2.24) follow immediately from (2.27)-(2.28) and (2.23).
2.25): Since

\[ \ln(\sigma^2_t) = \ln(\sigma^2_0) + \sum_{i=1}^{t} \ln(\beta + \alpha z_{t-1}^2) \]

the non-convergence of \( \ln(\sigma^2_t) \) follows immediately by the Kolmogorov three series theorem and the Levy equivalence theorem. (Dudley (1987)). Divergence in \( L^p \) for all \( p > 0 \) follows from (2.28). \( \text{Q.E.D.} \)

To evaluate conditions for \( L^p \) convergence when \( |z_t| \) is iid \( N(0,1) \), we have

**Proposition 2.5:**

If \( z \sim N(0,1), \alpha, \beta > 0 \),

\[ E[(\beta + \alpha z^2)^p] = [2\alpha]^{-1/2} \beta^{p+1/2} \Psi(\frac{1}{2}, \frac{p+3/2}{2}, \frac{\beta}{2\alpha}) \]

where \( \Psi(\cdot, \cdot, \cdot) \) is a confluent hypergeometric function of the second kind (Lebedev (1972)).

**Proof:** Follows directly from the integral representation of \( \Psi \) given in Lebedev (1972).

\( \text{Q.E.D.} \)

Intuitively, the reason that \( \sigma^2_t \) can die out almost surely but persist in \( L^1 \) is that almost sure convergence to zero of \( \sum_{t=1}^{\infty} (\beta + \alpha z_{t-1}^2) \) requires only that the geometric mean of \( (\beta + \alpha z_{t-1}^2) \) is less than one whereas convergence in \( L^1 \) requires that the arithmetic mean be less than one. If \( \alpha > 0 \), the arithmetic mean is strictly greater than the geometric mean, so that if the geometric mean is less than
one but the arithmetical mean is greater than one, the series will diverge in $L^1$ but converge almost surely.
III. GARCH(1,1) WITH $\omega > 0$

In this section we establish conditions for strict stationarity of GARCH(1,1) with $\omega > 0$, and examine via Monte Carlo the densities and distributions of various IGARCH(1,1) processes.

First, recall (2.1)-(2.4):

2.1) $\{z_t\}, \ t=--\infty, \infty$ iid $N(0,1)$

2.2) $\xi_t = z_t \sigma_t$

2.3) $\omega \sigma^2_t = \omega + (\beta + \alpha z_{t-1}^2) \sigma^2_{t-1}$

2.4) $\sigma^2_t = \sigma^2_0 \prod_{l=1}^{t} (\beta + \alpha z_{l-1}^2) + \omega \sum_{k=1}^{t} \prod_{l=1}^{k} (\beta + \alpha z_{l-1}^2)$

Note that the first term on the right-hand side of (2.4) is just the value of $\sigma^2_t$ that would prevail if $\omega$ were equal to zero, and that the second term on the right-hand side of (2.4) is independent of $\sigma^2_0$. Consider a shock at time zero as a change in $\sigma^2_0$. (2.4) implies that the criteria for persistence or decay of a shock entering at time zero are exactly the same as the criteria in proposition 2.4 -- i.e. if $E[\ln(\beta + \alpha z_1^2)] < 0$, then shocks decay almost surely, if $E[(\beta + \alpha z_1^2)^p] < 1$ they decay in $L^p$ and etc.. The main result of this section is that the condition $E[\ln(\beta + \alpha z_1^2)] < 0$ is also necessary and sufficient for $\sigma^2_t$ to have a strictly stationary and ergodic limiting distribution. Proposition 3.1 also characterizes the limiting distribution of a GARCH(1,1) with initial state.
\( \sigma^2_0 \) in terms of a random variable that does not depend on \( \sigma^2_0 \).

**Proposition 3.1:**

Define the stochastic processes

3.1)  
\[ \xi_t^* = z_t \sigma_t^* \]

3.2)  
\[ \sigma_t^* = \omega[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{\pi} (\beta + \alpha z_{t-i}^2)] \]

If \( \omega > 0 \), then \( \{\sigma_t^*, \xi_t^*\} \) is stationary and ergodic iff \( E[\ln(\beta + \alpha z_{t}^2)] < 0 \). Further, if \( E[\ln(\beta + \alpha z_{t}^2)] < 0 \) then 
\[ \{\sigma_t^* - \sigma_t^*, \xi_t^* - \xi_t^*\} \to \{0,0\} \text{ a.s.} \text{, and } \{\sigma_t^*, \xi_t^*\} \to \{\sigma_t^*, \xi_t^*\} \text{ in distribution.} \]

**Proof:**

When \( E[\ln(\beta + \alpha z_{t}^2)] \geq 0 \), \( \prod_{i=1}^{\pi} (\beta + \alpha z_{t-i}^2) \) diverges by proposition 2.4. For \( \{\sigma_t^*, \xi_t^*\} \) to be stationary and ergodic it is therefore necessary that \( E[\ln(\beta + \alpha z_{t}^2)] < 0 \). To show this condition is also sufficient, suppose first that for all \( t \), \( \sigma_t^* \), \( \xi_t^* \) was finite almost surely. Then the stationarity and ergodicity of \( \{\sigma_t^*, \xi_t^*\} \) would follow directly from the representation (3.1)-(3.2), and Stout (1974) theorem 3.5.8.

To prove that for each \( t \), \( \{\sigma_t^*, \xi_t^*\} \) are finite almost surely, we will prove that the sequence

\[ \{ \prod_{i=1}^{\pi} (\beta + \alpha z_{t-i}^2) \}, i=1,\infty = o(\exp(-b)) \text{ a.s. for some } b > 0, \]
so that the sum in (3.2) converges a.s. Using the strong law of large numbers the same way that we did in the proof of proposition 2.1, we have

\[ 3.3) \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} \ln(\beta + \alpha z_{i-1}^{2}) = E[\ln(\beta + \alpha z_{1}^{2})] \text{ a.s.} \]

That is, for each $\delta > 0$, there exists an integer $M$ such that for all $k > M$,

\[ 3.4) \left| \frac{1}{k} \sum_{i=1}^{k} \ln(\beta + \alpha z_{i-1}^{2}) - E[\ln(\beta + \alpha z_{1}^{2})] \right| < \delta \]

$M$ is, in general, stochastic, but the strong law guarantees that it is finite almost surely. Set $\delta = (1/2)E[\ln(\beta + \alpha z_{1}^{2})]$. Then by (3.3) and (3.4), we have for all $k > M$

\[ 3.5) \left[ \prod_{i=1}^{k} (\beta + \alpha z_{i-1}^{2}) \right]^{1/k} < \exp[(1/2)E[\ln(\beta + \alpha z_{1}^{2})]] < 1 \]

so that

\[ 3.6) \prod_{i=1}^{k} (\beta + \alpha z_{i-1}^{2}) = o(e^{-b}) \text{ a.s., where } b = (1/4)E[\ln(\beta + \alpha z_{1}^{2})] \]

By definition 3.1, the almost sure finiteness of $\xi_{1}^*$ follows from the almost sure finiteness of $z_{1}$ and $\sigma_{1}^{2,*}$, which concludes the proof that $[\xi_{1}^*, \sigma_{1}^{2,*}]$ is stationary and ergodic.

Finally, we must prove that $[\xi_{t} - \xi_{t}^*, \sigma_{1}^{2} - \sigma_{1}^{2,*}] \to (0,0)$ a.s. when $E[\ln(\beta + \alpha z_{1}^{2})] < 0$. Again by definition 3.1, $[\xi_{t} - \xi_{t}^*] \to 0$ a.s. will follow immediately if $[\sigma_{1}^{2} - \sigma_{1}^{2,*}] \to 0$ a.s.. From (2.4) and definition (3.2), we have
3.7) \[ \sigma^2 - \sigma^2 = \sigma^2_0 \prod_{i=1}^{t} (\beta + \alpha z_{i-1}^2) + \omega \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\beta + \alpha z_{i-1}^2) \]

The first term on the right-hand side of (3.7) vanishes a.s. by Proposition 2.2. The second term vanishes a.s. since, as we just proved, 

\[ \prod_{i=1}^{k} (\beta + \alpha z_{i-1}^2) = o(e^{-b}) \text{ a.s. for } a, b > 0. \]

Therefore \( \{\sigma^2 - \sigma^2\} \to 0 \text{ a.s.} \), so that \( \{e - e, \sigma^2 - \sigma^2\} \to \{0,0\} \text{ a.s.} \). But by White (1984) Lemma 4.7, this implies that \( \{e, \sigma^2\} \) converges in distribution to \( \{e_0, \sigma^2_0\} \).

Q.E.D.

Consider figure 1 again. Bollerslev (1986) showed that \( \{e\} \) is covariance stationary iff \( (\alpha, \beta) \) lie in region I. Proposition 3.1 says that \( \{e, \sigma^2\} \) has a strictly stationary and ergodic (but not necessarily covariance stationary) steady state distribution if \( (\alpha, \beta) \) lie either in region I or region II or on their common boundary. Since \( (\alpha, \beta) \) in an IGARCH(1,1) process lie on the boundary between regions I and II, an IGARCH(1,1) process with \( \omega > 0 \) has a strictly stationary and ergodic limiting distribution, while IGARCH(1,1) has a trivial limiting distribution with 

\[ \{e, \sigma^2\} \to \{0,0\} \text{ a.s.} \]

Since IGARCH(1,1) processes have steady state distributions, a natural question to ask is what these distributions look like. Figures 2 and 3 summarize a Monte
CDFs for IGARCH(1,1)

From left to right:
Alpha = 1.0, .75, .5, .25, .1

Figure II

Density Estimates

From left to right:
Alpha = 1.0, .75, .5, .25, .1

Figure III
Carlo experiment with 2000 draws from IGARCH(1,1) distributions for $\sigma_{t}^{2}$ with $(\alpha, \beta) = (1,0), (.75,.25), (.5,.5), (.25,.75), (.1,.9)$. The $z_t$'s were drawn from a $N(0,1)$ distribution using the random number generator in GAUSS (Edlefonson and Jones (1986)).

From definition 3.2, it is clear that the support of $\sigma_{t}^{2}$ is $[\omega/(1-\beta), \infty)$ and that $\omega$ is just a scale parameter. For simplicity therefore we set $\omega = 1-\beta$ and plot in figure 2 the empirical cdfs of $\ln(\sigma_{t}^{2})$.

Strictly speaking, we cannot draw from the $\sigma_{t}^{2}$ distribution, since $\sigma_{t}^{2}$ is a function of an infinite number of $z_t$'s. However by proposition 3.1 we know that the dependence on initial conditions vanishes exponentially in the limit. Trial and error established that setting $\sigma_0^2 = 1$ and then drawing $\sigma_{2000}^2$ as a proxy for $\sigma_{t}^{2}$ resulted in only minuscule errors, since for the values of $\alpha$ and $\beta$ we used and $t = 2000$, $\prod_{1=1}^{t} (\beta + \alpha z_{t-1}^2)$ is almost always a very tiny number. (Not surprisingly, however, the closer $E[\ln(\beta + \alpha z_t^2)]$ is to zero the larger $t$ value we need to use to get dependence on initial conditions to effectively vanish. For the coefficients we have chosen $t=2000$ proved to be sufficient.)

Figure 3 plots density estimates for $\ln(\sigma_{t}^{2})$. These estimates were obtained using a kernel estimator. Since with $\omega = 1-\beta$ the support of $\ln(\sigma_{t}^{2})$ is bounded below by zero, we first took $\ln(\ln(\sigma_{t}^{2}))$ and estimated $'s$ density
using a normal kernel with bandwidth selected using the procedure in Silverman (1986) chapter 3. We then did a change of variables to obtain a density estimate for $\ln(\sigma^2_t)$. This estimator is consistent, and its mean integrated squared error of the density estimator based on $n$ draws from the $\ln(\sigma^2_t)$ distribution converges to zero at rate $n^{-4/3}$ (Silverman (1986) chapter 3.) Note that the lower $\alpha$ is, the more strongly skewed to the right the $\sigma^2_t$ density is. This is not surprising, since the smaller $\beta$ is, the more influence a single small draw for $z_t^2$ has in making $\sigma^2_t$ small. Note also that $\sigma^2_t$ has very fat upper tails, with very large values for $\sigma^2_t$ found in all the IGARCH(1,1) distributions.
IV. THE SWISS FRANC EXCHANGE RATE DATA REVISITED: TESTING ALMOST SURE PERSISTENCE

In this section we reexamine the dollar/swiss franc exchange rate data previously analyzed in Diebold and Nerlove (1985) and Bollerslev and Engle (1986). Bollerslev and Engle used this data to test hypotheses about the persistence of volatility shocks using the IGARCH(1,1) model with $\omega = 0$. In section II we suggested that a more appropriate model of persistence may be to select $(\alpha, \beta)$ such that $E[\ln(\beta + \alpha z_i^2)] = 0$, so that $\ln(\sigma_i^2)$ becomes a simple random walk. We reestimate the model and compare the fits under various restrictions on $(\alpha, \beta, \omega)$.

The data consists of Wednesday interbank closing spot bid prices for the Swiss franc from July 1973 to August 1985. The data is originally from the International Money Markets Yearbook, and was kindly supplied to the author of this paper by Robert Engle. As discussed in Bollerslev and Engle (1986), the first differences in the logs of the exchange rate have a mean near zero and exhibit no significant serial correlation, so that we can concentrate entirely on changes in the higher moments. Estimation was carried out using the GAUSS procedure MAXMUM.

Since in (2.16) $\alpha$ is an implicit function of $\beta$ and visa versa, some numerical procedure is required to derive $\beta$ given $\alpha$ at each iteration in the estimation procedure in order to satisfy the constraint that $E[\ln(\beta + \alpha z_i^2)] = 0$. 

97
The generalized hypergeometric functions \( _1F_1 \) and \( _2F_2 \) have infinite series representations given in Lebedev (1972).

When \( \alpha \) is not too small (say \( \alpha > .2 \)) truncated versions of these series and its derivatives give fast and accurate approximations of the right-hand side of (2.16) and its derivatives. These can then be used in combination with Newton's rule to find, for a given \( \alpha \), the \( \beta \) that satisfies

\[ \text{E}[\ln(\beta + \alpha z^2)] = 0. \]

When \( \alpha \) is small, however, the truncated series is inaccurate and slow to converge, so an alternate means of evaluating (2.16) and its derivatives is needed. Since \( z^2 = x^2 \) with 1 degree of freedom, we have

\[ 4.1) \text{E}[\ln(\beta + \alpha z^2)] = (2\pi)^{-1/2} \int_0^\infty e^{-y/2} y^{-1/2} \ln(\beta + \alpha y) \, dy \]

\[ 4.2) = \ln(\beta) + (2\pi)^{-1/2} \int_0^\infty e^{-y/2} y^{-1/2} \ln[1 + (\alpha/\beta)y] \, dy \]

Integrating by parts yields

\[ 4.3) = \ln(\beta) + (\alpha/\beta) + (\alpha/\beta)^2 \int_0^\infty (t[F(t)-1])/(1+t\alpha/\beta) \, dt \]

where \( F(t) \) is the cdf of a chi-square random variable with one degree of freedom. Numerical quadrature, for example using the INTQUAD program in GAUSS, yields a fast, accurate evaluation of (4.3) for small \( \alpha \). Over a substantial range (say \( 0.2 < \alpha < 0.4 \)) the two methods agree to several decimal places, which is accurate enough to allow constrained estimation. Differentiating (4.3) with respect to \( \alpha \) or \( \beta \)
and integrating via quadrature yields fast, accurate derivatives for use in Newton's rule.

In table 4.1 below we report results on coefficients and likelihoods under different hypotheses.\(^3\) Standard errors for coefficients, which are computed from the score, are in parenthesis. \(\lambda\) is the log likelihood, and \(x^2\) is the likelihood ratio test statistic with \(d\) degrees of freedom for each of the restricted GARCH(1,1) models against the unrestricted GARCH(1,1).

Little is known about the asymptotic distribution of the estimators in GARCH models without quite restrictive moment conditions that are not satisfied by the coefficients estimated below. The test statistics below are computed assuming asymptotic normality of the coefficient estimates. For the state of the art in proving asymptotic normality in ARCH models, see Wiess (1986).

\(^3\) The reported coefficient estimates differ slightly from the estimates reported in Engle and Bollerslev for models that were estimated in both papers. Since the likelihoods reported here are slightly higher, it seems likely that Bollerslev and Engle found a local maximum. Of course, there is no guarantee that I have found the global maximum either.
TABLE 4.1

<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$(\alpha + \beta)$</th>
<th>$\omega \times 1e+5$</th>
<th>$\gamma$</th>
<th>$\chi^2$</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>unrestricted</td>
<td>0.115975</td>
<td>0.883155</td>
<td>0.99913</td>
<td>0.2732</td>
<td>1750.944</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.0163)</td>
<td>(0.0126)</td>
<td></td>
<td></td>
<td>(0.0127)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega = 0$</td>
<td>0.115081</td>
<td>0.899671</td>
<td>1.0175</td>
<td>0</td>
<td>1748.013</td>
<td>5.863</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.0144)</td>
<td>(0.010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGARCH(1,1)</td>
<td>0.116713</td>
<td>0.883287</td>
<td>1</td>
<td>0.2647</td>
<td>1750.941</td>
<td>11.44</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.0163)</td>
<td>(0.0125)</td>
<td></td>
<td></td>
<td>(0.0125)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGARCH(1,1), $\omega = 0$</td>
<td>0.090783</td>
<td>0.909217</td>
<td>1</td>
<td>0</td>
<td>1745.226</td>
<td>0.006</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(0.0103)</td>
<td>(0.0088)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E[\ln(\beta + \alpha z^2)]$</td>
<td>0.108979</td>
<td>0.900564</td>
<td>1.0095</td>
<td>0</td>
<td>1747.700</td>
<td>6.448</td>
<td>2</td>
</tr>
<tr>
<td>$=0$ $\omega = 0$</td>
<td>(0.0130)</td>
<td>(0.0096)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IGARCH(1,1) with no drift is rejected at any standard significance level, while IGARCH(1,1) with $\omega > 0$ is not rejected at any reasonable significance level. The otherwise unrestricted GARCH(1,1) model fit with $\omega = 0$ is actually explosive, since $E[\ln(\beta + \alpha z^2)] = 0.0042 > 0$. This model has a probability value of 0.015, and so is rejected at the 1.5% level. The model with $E[\ln(\beta + \alpha z^2)] = 0$ and $\omega = 0$, in which $\ln(\sigma^2_t)$ is a random walk, does somewhat better: it imposes another restriction and so gains a degree of freedom in the $\chi^2$ test, but sacrifices little in the way of fit, so that its probability value is 0.039. In sum, the IGARCH(1,1) model with $\omega=0$, which makes $\sigma^2_t$ a martingale, does very badly, while the model suggested by section II, in which $\ln(\sigma^2_t)$ is a random walk, does
substantially better, though it is rejected at the 5% level. IGARCH(1,1) with \( \omega > 0 \), in which \( \sigma^2 \), is strictly stationary but with no finite moments in steady state, fits almost as well as the unrestricted model. It seems, then, that the non-covariance stationary GARCH(1,1) models in which \( \omega > 0 \) and \((\alpha, \beta)\) lie in region II in figure 1, may be important empirically.
V. CONCLUSION

In this paper we have exhaustively examined the properties of GARCH(1,1). What can a practitioner using ARCH models learn from our results?

First, unless ones economic model calls for variance to die out asymptotically, IGARCH(1,1) with no drift is probably not a good way to model changing volatility and the persistence of shocks to volatility. If one wishes to have shocks persist almost surely but not have the probability mass of $\sigma^2_t$ pile up around a particular point, a better choice seems to be to make $\omega = 0$ and $E[\ln(\beta + \alpha z_t^2)] = 0$, so that $\ln(\sigma^2_t)$ follows a random walk.

Second, if one estimates a GARCH(1,1) model and finds $\alpha + \beta > 1$, this does not necessarily constitute evidence of misspecification: the process may still be stationary, even if it has no finite moments. Stationarity imposes certain inequality constraints on $(\alpha, \beta)$ which can be tested.

One interesting way to extend the results of this paper would be to show that a unique limiting density (as opposed to a limiting distribution) exists for $\{\sigma^2_t\}$ and find an efficient numerical method to solve for this density given $\alpha$ and $\beta$. We could then estimate GARCH(1,1) processes by exact maximum likelihood, rather than by choosing $\sigma^2_0$ arbitrarily as is the current practice. Such a limiting density $f(\sigma^2_t)$, if it can be shown to exist, must be the solution to the
integral equation

\[ 5.1 \quad f(\sigma^2_t) = \int_0^\infty g(\sigma^2_t | \sigma^2_{t-1}) f(\sigma^2_{t-1}) \, d\sigma^2_{t-1} \]

subject to

\[ 5.2 \quad 1 = \int_0^\infty f(\sigma^2_t) \, d\sigma^2_t \]

and

\[ 5.3 \quad f(\sigma^2_t) \geq 0 \text{ for all } \sigma^2_t, \quad 0 \leq \sigma^2_t < \infty \]

where \( g(\sigma^2_t | \sigma^2_{t-1}) \) is the transition density of \( \sigma^2_t \) given \( \sigma^2_{t-1} \).

There is a vast mathematics literature on integral equations, and it may well possible to prove existence and uniqueness of a solution to (5.1)-(5.3). The author has tried, but has not yet succeeded.

Probably the most important extension of the results of this paper would be to develop conditions for stationarity and ergodicity of higher-order GARCH processes. In a GARCH\((p,q)\) process, we have, in place of (1.2)

\[ 5.4 \quad \sigma^2_t = \omega + \sum_{i=1}^p \alpha_i \xi^2_{t-i} + \sum_{j=1}^q \beta_j \sigma^2_{t-j} \]
When \( q \geq 1 \) and \( p \geq 2 \), we can re-write (5.4) as

\[
\begin{bmatrix}
\sigma^2_{t+1} \\
\sigma^2_t \\
\vdots \\
\sigma^2_{t-q} \\
\xi^2_t \\
\xi^2_{t-p+2}
\end{bmatrix} =
\begin{bmatrix}
\alpha_1 + \beta_1 z^2_1 & \alpha_2 & \cdots & \alpha_{q-1} & \alpha_q & \beta_2 & \cdots & \beta_{p-1} & \beta_p \\
1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\
\xi^2_t & z^2_t & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \\
\xi^2_{t-p+2} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\sigma^2_t \\
\sigma^2_{t-1} \\
\vdots \\
\sigma^2_{t-q+1} \\
\xi^2_t \\
\xi^2_{t-p+1}
\end{bmatrix}
\]

or, in matrix notation

\[
H_{t+1} = W + A_t H_t
\]

Substituting recursively for \( H_{t-1} \) yields the equivalent of (2.4):

\[
H_t = \left[ \prod_{i=1}^{t} A_{t-i} \right] H_0 + W + \left[ \sum_{k=1}^{t} \prod_{i=1}^{k} A_{t-i} \right] W
\]

Just as in section III, \( H_t \) will be possess a stationary and ergodic limiting distribution if the first term on the right-hand side of (5.7) vanishes to zero a.s. and the terms in the summation in (5.7) die out quickly enough. In

\[4\] A similar matrix representation is easily derived when \( p < 2 \) and \( q \leq 1 \) or when \( q = 0 \) and \( p > 2 \).
sections II and III, we developed a fairly simple criterion for stationarity and ergodicity for the GARCH(1,1) model. Since $A_t$ was a scalar in this model, we were able to take logs of the product terms and turn them into summation terms. We then were able to normalize these sums and apply the strong law of large numbers to show that they died out exponentially. Unfortunately, when $A_t$ is a matrix, we can no longer take logs to convert the products into sums, so that the method used in sections II and III no longer works.

Fortunately, however, there is an enormous mathematics literature on convergence of matrix products (see, for example, Bougerol (1987), Cohen and Newman (1984), Kesten and Spitzer (1984) and Cohen et. al. eds. (1984) and the Cohen (1984) Bibliography.) Using the techniques developed in this literature, it may well be possible to develop a set of necessary and sufficient conditions for stationarity and ergodicity of $H_t$. This remains to be carried out.
REFERENCES:


Diebold, F., and M. Nerlove, 1985, "ARCH models of Exchange Rate Fluctuations", Unpublished manuscript, University of Pennsylvania Department of Economics.


CHAPTER THREE:

CONDITIONAL HETEROSKEDASTICITY IN ASSET RETURNS:
A NEW APPROACH

December 1, 1987

Summary: ARCH models have been fruitfully applied in modeling the relation between conditional variance and asset risk premia. Unfortunately, these models have at least three features which are unappealing in asset pricing applications: 1) ARCH models impose the restriction that only the magnitude of residuals and not their direction determines future conditional variance. This contradicts the findings of Black (1976) and many other researchers that the volatility of stock returns tends to rise when the market falls and fall when the market rises. 2) ARCH models impose restrictions on the dynamics of conditional variance that are theoretically unappealing and also cause numerical problems in estimation. 3) There is no continuous time version of ARCH, whereas most asset pricing models are written in terms of stochastic differential equations. A new form of conditional heteroskedasticity is proposed that meets these objections. The method is used to estimate a simple model of the risk premium on the S&P 500 from 1928 to 1987.
I. INTRODUCTION

After the events of October 1987, few would argue with the proposition that stock market volatility changes over time. Understanding the way in which it changes is crucial to our understanding of many areas in macroeconomics and finance. Consider some of the following examples:

a) Options pricing: The Black–Scholes options pricing formula is invalid if volatility changes randomly. The implications of stochastic conditional volatility for options pricing have been investigated by Cox and Rubenstein (1985), and Wiggins (1986).

b) Term structure of interest rates: The recent work of Barsky (1986) and Abel (1986) has highlighted the importance of the perceived level of market risk in determining both riskless and risky required rates of return. In the Barsky and Abel models, riskless rates fall and the risk premium rises when there is a rise in market volatility. Required rates of return on risky investments may rise or fall, depending on the aggregate level of relative risk aversion. If shocks to volatility are expected to persist, then shifting volatility may have important implications for the term structure of interest rates, and thus for capital investment decisions.

c) Investment Theory: Bernanke (1983), McDonald and Siegel (1986), and Bertola (1987) have pointed out that if investment is irreversible and demand is uncertain, then the
option that agents have to wait to invest rather than investing now will strongly affect the timing of investment. If volatility rises, then the option to wait to invest becomes more valuable. If asset volatilities in an economy tend to move together, then changing market volatility may play an important role in explaining investment booms and busts in the business cycle. Some preliminary evidence (Pindyck (1986)) suggests that it does.

d) Dynamic Capital asset pricing theory: If conditional covariances of asset returns vary randomly, then CAPM models that assume constant asset covariances are incorrect. Merton (1973), Cox, Ingersoll and Ross (1985), Huang (1987), and others have developed asset pricing models that allow asset volatility to change randomly.

Recent years have also seen a surge of interest in econometric models of changing variance. Probably the most widely used (but by no means the only1) such models are the ARCH (autoregressive conditional heteroskedasticity) and GARCH (generalized ARCH) models introduced by Engle (1982) and Bollerslev (1986). The new ARCH methodology has been

1) There are many approaches in the literature to modeling changing conditional variances. Perhaps the most important alternative method is the approach found in Poterba and Summers (1986) and French, Schwert and Stambaugh (1986). For a critique of the approach used by these authors, see Nelson (1987a), which presents yet another way to model changing conditional variance.
fruitfully applied in asset pricing models: For example, Engle, Lilien and Robins (1987) extended the ARCH method of Engle to allow conditional variance to be a determinant of expected bond returns. Bollerslev, Engle and Wooldridge (1986) extended this model and Bollerslev's (1986) GARCH model further to test CAPM with time varying covariances of asset returns.

A univariate GARCH-M model for the excess return \( (ER_t) \) on the market portfolio can be formulated as follows:

1.1 \[ ER_t = f(\sigma_t^2) + \epsilon_t \]

1.2 \[ \epsilon_t | \psi_{t-1} \sim N(0, \sigma_t^2) \]

1.3 \[ \sigma_t^2 = \omega + \sum_{i=1}^{a} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{b} \alpha_j \epsilon_{t-j}^2 \]

\( \psi_t \) is the information set at time \( t \), and \( \sigma_t^2 \) is the conditional variance of the error term \( \epsilon_t \) given information at time \( t-1 \). \( f(\sigma_t^2) \) is the risk premium, the conditional mean of \( ER_t \). \( f \) is allowed to depend on the conditional variance \( \sigma_t^2 \), since economic theory may require that required rates of return are high when risk (measured by \( \sigma_t^2 \)) is high. For \( \sigma_t^2 \) to be well specified as a conditional variance, it must be non-negative a.s.. This, in turn, requires that \( \omega \) and the \( \alpha_j \) and \( \beta_j \) are non-negative. We can also write the system given in equations 1.1 to 1.3 as

1.4 \[ \{z_t\} t=-\infty, \infty \sim i.i.d. N(0,1) \]

1.5 \[ ER_t = f(\sigma_t^2) + z_t \sigma_t \]
1.6) \[ \sigma_t^2 = \omega + \sum_{i=1}^{\infty} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{m} \alpha_j z_{t-j}^2 \sigma_{t-j}^2 \]

Substituting for the $\beta_i \sigma_{t-i}^2$ terms lets us re-write 1.6 as

1.7) \[ \sigma_t^2 = \omega^* + \sum_{k=1}^{\infty} \phi_k z_{t-k}^2 \sigma_{t-k}^2 \]

When the unconditional expectation of $\sigma_t^2$ exists

1.7)\' \[ \sigma_t^2 = \omega^* + \sum_{k=1}^{\infty} \phi_k \mathbb{E}(\sigma_t^2) + \sum_{k=1}^{\infty} \phi_k \left[ z_{t-k}^2 \sigma_{t-k}^2 - \mathbb{E}(\sigma_{t-k}^2) \right] \]

1.7\' gives the moving average representation of $\sigma_t^2$.

While the GARCH-M equations above provide a simple and elegant means for modeling changing asset risk and the influence of risk on required returns, careful examination of equations 1.1 through 1.7 reveals some uncomfortable restrictions on the $\sigma_t^2$ process. Perhaps the most serious limitation of ARCH models of asset pricing is the assumption that only the size and not the sign of excess returns determines future $\sigma_t^2$. In equation 1.7, $\sigma_t^2$ is a function of lagged $\sigma_t^2$ and lagged $z_t^2$, and so it is a function only of the magnitude and not the direction of the $z_t$'s. That is, the conditional variance tomorrow does not depend on whether prices today move up or down; it only depends on whether the price movement today is small or large. This assumption contradicts a widely noted stylized fact (see Black (1976), Christie (1982), Nelson (1987a)) that when the price of a stock (or portfolio of stocks) rises, its returns volatility tends to fall, and when the price falls, its
volatility tends to rise. For example, let $\sigma^2(t, z_{t-1}, \psi_{t-2})$ denote the returns variance at time $t$, conditioned on $z_{t-1}$ and all information up to time $t-2$. The GARCH-M model outlined above would tell us that $\sigma^2(t, 1, \psi_{t-2}) = \sigma^2(t, -1, \psi_{t-2}) < \sigma^2(t, 2, \psi_{t-2}) = \sigma^2(t, -2, \psi_{t-2})$, whereas the stylized facts of stock volatility would lead one to expect that $\sigma^2(t, 2, \psi_{t-2}) < \sigma^2(t, 1, \psi_{t-2}) < \sigma^2(t, -1, \psi_{t-2}) < \sigma^2(t, -2, \psi_{t-2})$.

Another limitation of ARCH modeling is in the restrictions it places on the dynamics of the conditional variance process $\sigma_t^2$: a conditional variance process must be non-negative almost surely, and as we noted earlier, this requires that $\alpha_j$ and $\beta_j$ be non-negative for all $j$. This in turn restricts the moving average coefficients in (1.7)' to be positive, so that a positive shock to conditional variance at one time implies strictly higher conditional variance at all future periods. We therefore rule out cycling behavior in the conditional variance process. Furthermore, these restrictions on the coefficients have created considerable difficulties in estimating ARCH models, in that the estimated coefficients show a distressing

---

2) Of course, ARCH models could accommodate correlation between $z_t$ and $\text{sign}(z_t)$ by drawing $z_t$ from an asymmetric distribution, but at the cost of imposing what may be quite unnatural restrictions on the conditional distribution. For any symmetric distribution, the size and the sign of $z_t$ will be uncorrelated (so long as the correlation is well-defined.)
tendency to wander into the forbidden region. For example, Engle, Lilien, and Robins (1987), had to impose a linearly declining structure on the $\alpha_j$ ($j \geq 1$) coefficients to prevent some of them from becoming negative. In a multivariate context, Friedman and Kuttner had to impose quite severe restrictions to keep the estimated conditional covariance matrix non-negative definite.

A third limitation of ARCH models is their apparent lack of continuous time versions. This is a drawback in financial applications since many intertemporal asset pricing models (i.e. Merton (1973), Breeden (1979), Cox, Ingersoll and Ross (1985) and Huang (1987)) model asset price movements as realizations of systems of stochastic differential equations, and conditional variance in these models itself follows an Ito process. In ARCH models, prices and conditional variance are given by a system of stochastic difference equations. A natural question to ask in applications is whether these stochastic difference equations converge in some sense to stochastic differential equations as the time interval between observations shrinks to zero. If they do, then we would expect to be able to use ARCH processes to approximate the continuous time diffusion processes used in asset pricing models. Unfortunately, no diffusion version of ARCH has yet appeared in the literature, and it seems unlikely that there will ever be one. To see heuristically why ARCH processes are unlikely
to have continuous time versions, recall equation 1.7:

\[ 1.7) \quad \sigma_t^2 = \omega^* + \sum_{k=1}^{\infty} \phi_k z_{t-k}^2 \sigma_{t-k}^2 \]

A natural continuous time analog of 1.7 is

\[ 1.7)^* \quad \sigma_t^2 = \omega^* + \int_0^\infty \phi(s) \sigma_{t-s}^2 [dW_s]^2 \]

where \( W_t \) is a standard Brownian motion. But as is well known from the theory of Brownian motions (see for example Oksendal (1980))

\[ 1.10) \quad [dW]^2 = dt \]

so that

\[ 1.7)^{**} \quad \sigma_t^2 = \omega^* + \int_0^\infty \phi(s) \sigma_{t-s}^2 ds \]

In other words, while in discrete time \( \sigma_t^2 \) follows a stochastic difference equation, its continuous time counterpart follows a deterministic differential equation. We will develop this argument more formally in appendix I below.

The object of this paper is to present an alternative to ARCH that meets the objections outlined above, and so may be more suitable than ARCH for modeling conditional variance in asset returns. In section II, we describe this \( \sigma_t^2 \) process, and develop some of its properties. In section III, we discuss estimation and specification testing with the new approach, and in section IV we estimate a simple
model of excess returns on the S&P 500. Section V is a brief conclusion. In appendix I, we develop methods for showing that a sequence of stochastic difference equations converge in distribution to a diffusion process, and apply these tools to both traditional ARCH models and our alternative model. In particular, we show how to use exponential ARCH processes to approximate the sort of diffusion models of asset returns found in the finance literature. All proofs are in appendix II.
A necessary condition for a stochastic process $\sigma_t^2$ to be well defined as a conditional variance for the innovations of another process is that $\sigma_t^2$ be non-negative for all $t$ almost surely. The ARCH and related models ensure this by making $\sigma_t^2$ a positive linear combination of positive random variables. The positive random variables chosen were the products of lagged $\sigma_t^2$'s and squared standard normal random variables (the $z_t^2$.) This approach, as we saw in section I, leads to a number of important limitations. But the original ARCH formulation found in Engle (1982) equations (1)-(5) allows a much more general dependence of $\sigma_t^2$ on lagged residuals than the ARCH models we discussed in section I: this (most general ARCH formulation) is

$$
\sigma_t^2 = \sigma^2(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, b) = \sigma^2(\sigma_{t-1} z_{t-1}, \sigma_{t-2} z_{t-2}, \ldots, b)
$$

where $b$ is the vector of system parameters. In this section we develop another member of this general family that does not suffer from the drawbacks of the ARCH model we discussed in section I. We adopt another natural device for ensuring that $\sigma_t^2$ remains non-negative: we let the log of $\sigma_t^2$ be linear in some function of time and lagged $z_t$'s -- i.e. for some suitable function $g$:

$$
\ln(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k \ g(z_{t-k}) \quad \beta_1 = 1
$$
where $|\alpha_t|$, $t=-\infty, \infty$ and $|\beta_k|$, $k=1, \infty$ are real, non-stochastic scalar sequences.

To allow conditional variance to depend on both the magnitude and the sign of returns, $g(z_t)$ must be a function of both the magnitude and the direction of $z_t$. One choice (that will turn out to give $\sigma_t^2$, well-behaved moments) is to make $g(z_t)$ a linear combination of $z_t$ and $|z_t|$, i.e. we could choose

2.3) $g(z_t) = \theta z_t + \delta |z_t|$

Using the fact that $E|z_t| = (2/\pi)^{1/2}$, we set

2.3) $g(z_t) = \theta z_t + \delta [|z_t|-(2/\pi)^{1/2}]$

which makes $g(z_t)$ a zero-mean, i.i.d. innovations term. The two components of $g(z_t)$ are $\theta z_t$ and $\delta [|z_t|-(2/\pi)^{1/2}]$. Each has mean zero and the two are orthogonal, though of course they are not independent.

The term $\delta [|z_t|-(2/\pi)^{1/2}]$ represents a magnitude effect in the spirit of the ARCH models discussed in section I. Assume for the moment that the $\delta$ terms and the $\theta$ term are positive and that $\theta = 0$. This implies that conditional variance rises when the magnitude of $z_t$ is larger than expected, and falls when the magnitude of $z_t$ is smaller than expected. Suppose now that $\theta < 0$, and $\delta = 0$. This would
imply that conditional variance rises (falls) when returns innovations are negative (positive). θ therefore allows for the leverage effects investigated by Black (1976) that we discussed earlier. This exponential form for ARCH, therefore, meets the first objection that we raised to the ARCH models in section I.

Another way to interpret \( g(z_t) \) is to write 2.3 as

\[
2.4) \quad g(z_t) = \text{constant} + [\theta + \varphi \text{sign}(z_t)]z_t
\]

Over the range \( 0 < z_t < \infty \), \( g(z_t) \) is linear in \( z_t \) with slope \( \theta + \varphi \), and over the range \( -\infty < z_t < 0 \), \( g(z_t) \) is linear with slope \( \theta - \varphi \). Thus, \( g(z_t) \) allows conditional variance to respond asymmetrically to rises and falls in stock price.

The moments of the \( \sigma_t^2 \) process are easily found using the following result:

\[ \textbf{Theorem 2.1} \]

Let \( \{z_t\}, t=-\infty,\infty \) be i.i.d. \( N(0,1) \), and define the function

\[
2.5) \quad Q(\{z_t\}) = \exp[\alpha + \sum_{t=1}^{\infty} \beta_t g(z_{t-1})]
\]

where

\[
2.6) \quad g(z_t) = \theta z_t + \varphi [|z_t|-(2/n)^{1/2}]
\]

Then
2.7) \( E[Q|z_t|] = \exp[\alpha - \delta (2/\pi)^{1/2} \sum_{i=1}^{\infty} \beta_i] \)
\[ \ast \pi \left[ \phi(\beta_1 + \theta \beta_1) \exp[\beta_1^2 (\theta + \delta)^2 / 2] + \phi(\beta_1 - \theta \beta_1) \exp[\beta_1^2 (\delta - \theta)^2 / 2] \right] \]

where \( \phi(\ ) \) is the standard normal cumulative distribution function. \( E[Q|z_t|] \) is finite if \( |\beta_1| = o(j^{-b}) \) for some \( b > 1 \).

Proof: see appendix II

The expectation and higher moments of \( \sigma_t^2 \) and \( E \sigma_t \) are easily obtained using theorem 2.1. An important implication of theorem 2.1 is that if \( \Sigma \beta_j \) is absolutely convergent then \( \sigma_t^2 \) has arbitrary finite moments. The covariogram of \( \sigma_t^2 \) is also easy to compute using the theorem:

2.8) \( \text{Cov}(\sigma_t^2, \sigma_{t-k}^2) = \exp[\alpha_t + \alpha_{t-k} - 2\delta (2/\pi)^{1/2} \sum_{i=1}^{\infty} \beta_i] \)
\[ \ast \left[ \pi \left[ \exp[\beta_1^2 (\delta + \theta)^2 / 2] \phi(\beta_1 + \theta \beta_1) + \exp[\beta_1^2 (\delta - \theta)^2] \phi(\beta_1 - \theta \beta_1) \right] \right] \]
\[ \ast \pi \left[ \exp[(\beta_k + \beta_k + j)^2 (\delta + \theta)^2 / 2] \phi[(\beta_k + \beta_k + j)(\delta + \theta)] + \right. \]
\[ \left. \exp[(\beta_k + \beta_k + j)^2 (\delta - \theta)^2 / 2] \phi[(\beta_k + \beta_k + j)(\delta - \theta)] \right] \right] \]
\[ - \pi \left[ \exp[\beta_k^2 (\theta + \delta)^2 / 2] \phi[\beta_k (\delta + \theta)] + \exp[\beta_k^2 (\theta - \delta)^2 / 2] \phi[\beta_k (\delta - \theta)] \right]^2 \]

where again \( \phi(\ ) \) is the standard normal cumulative distribution function.
Examining 2.8, we see that time only enters the covariance through the $\alpha_t$ term. Since all moments are finite as long as $\sum \delta_t$ is absolutely convergent, it immediately follows that $\exp[-\alpha_t] \sigma_t^2$ will be covariance stationary when the \{\delta_t\} sequence dies out quickly enough.

The covariances in 2.8 may be positive or negative, and $\sigma_t^2$ stays non-negative for any values of the system's parameters. The exponential ARCH process therefore does not suffer from the restricted dynamics drawback that we outlined in the previous section (i.e. cycling behavior is allowed), and since there are no inequality constraints on the parameters, estimation should be free from the numerical problems that such constraints cause in standard ARCH models.

As is the case with linear processes, there is often a more convenient expression for $\ln(\sigma_t^2)$ than the infinite moving average representation in 2.2. In many applications, the most convenient form will be an ARMA representation:

$$2.9) \quad \ln(\sigma_t^2) = \alpha_1 + \sum_{i=1}^{p} a_i \ln(\sigma_{t-i}^2) + \sum_{j=1}^{q} b_j g(z_{t-j})$$

which will have arbitrary finite moments as long as the roots of

$$[1.0 - \sum_{i=1}^{p} a_i y^i]$$

lie outside the unit circle.
The final criticism that we made in section I of traditional ARCH models was that they did not seem to have well behaved continuous time versions. It turns out, however, that exponential ARCH processes do, and we conclude this section with an example to illustrate this. Specifically, we consider a relatively simple continuous time model of the market risk premium due to Merton (1973), and present a sequence of exponential ARCH processes that converge to it in distribution. In this section we confine ourselves to this example, but in appendix I we present more general results.

Merton (1973) showed that if the conditional variance of the market portfolio ($\sigma_t^2$) follows an Ito process, and if there is a representative agent with log utility, then the instantaneous risk premium on the market is $\sigma_t^2$. Suppose that $\ln(\sigma_t^2)$ follows an AR(1) in continuous time. Then the market excess returns process $S_t$ is given by:

2.10) \[ \frac{dS_t}{S_t} = \sigma_t^2 dt + \sigma_t dW_{1,t} \]

2.11) \[ d[\ln(\sigma_t^2)] = -\beta [\ln(\sigma_t^2) - \alpha] dt + dW_{2,t} \]

where $S_t$ is the cumulative excess returns on the market at time $t$, and $W_{1,t}$ and $W_{2,t}$ are Brownian motions with

2.12) \[
\begin{bmatrix}
    dw_{1,t} \\
    dw_{2,t}
\end{bmatrix}
= \begin{bmatrix}
    1 & C_{12} \\
    C_{12} & C_{22}
\end{bmatrix}
\begin{bmatrix}
    dt \\
    dt
\end{bmatrix}
\]

and where $C_{12}$ and $C_{22}$ are constrained in the obvious way to
make $C$ non-negative definite. $\sigma_t^2$ is the instantaneous variance per unit of time.

Setting $S_t^* = \ln(S_t)$ and applying Ito's lemma, we can rewrite 2.10 in a more convenient form:

$$2.10^* \quad dS_t^* = 0.5*\sigma_t^2 dt + \sigma_t dW_{t, t}$$

Our next step is to construct a sequence of stochastic difference equations that converge in distribution to the system given by 2.10*, 2.11 and 2.12 as the measurement interval $h$ goes to zero. $\sigma_t^2$ will be measured in the same (say annual) units for all $h$. We model $S_t^*$ and $\sigma_t^2$ as step functions with jumps at times $h$, $2h$, $3h$ and so on. Since $\ln(\sigma_t^2)$ follows a continuous time AR(1) process in 2.11, our discrete time models will also make $\{\ln(\sigma_k^2)\}, k=0, \infty$ an AR(1) process. For a given $h$, our discrete-time equations are

$$2.13 \quad S_{j+h} = S_{(j+1)h} + (h/2)\sigma_{jh}^2 + \sigma_{jh}Z_{jh}$$

$$2.14 \quad \ln[\sigma_{(j+1)h}^2] = \ln(\sigma_{jh}^2) - (1/2)[\ln(\sigma_{jh}^2) - \alpha]h + C_{1,2}Z_{jh}$$

$$+ \frac{\sigma}{\sqrt{h}}[(Z_{jh}^2 - (2h/\pi)^{1/2})]$$

where $\omega \equiv [(C_{1,2} - C_{1,2}^2)/(1-2/\pi)]^{1/2}$, and

$$2.15 \quad Z_{jh} \sim \text{i.i.d. } N(0, h)$$

It is easy to check that

$$2.16 \quad E \left[ Z_{jh} \begin{bmatrix} Z_{jh} \\ C_{1,2}Z_{jh} + \frac{\sigma}{\sqrt{h}}[(Z_{jh}^2 - (2h/\pi)^{1/2})] \end{bmatrix} \right] = \begin{bmatrix} Z_{jh} & C_{1,2}Z_{jh} + \frac{\sigma}{\sqrt{h}}[(Z_{jh}^2 - (2h/\pi)^{1/2})] \end{bmatrix}$$
\[
\begin{bmatrix}
1 & C_{12} \\
C_{12} & C_{22}
\end{bmatrix} h
\]

which is the discrete time analog of 3.3. \((S_{j+h}^* - S_{(j-1)+h}^*), (1n[\sigma_{j+h}^2] - 1n(\sigma_j^2)), h, Z_{j+h} \) and \([C_{12}Z_{j+h} + \delta[|Z_{j+h}| - (2h/\pi)^{1/2}]\) are the discrete time counterparts of \(dS^*\), \(d(1n(\sigma^2))\), \(dt\), \(dW_{1,t}\) and \(dW_{2,t}\) respectively. We then have

**Proposition 2.17**

As \(h \to 0\), \(\{S_{j+h}^*, \sigma_{j+h}^2\} \to \{S_{t}, \sigma_{t}^2\}\)

\(\text{d}\) where \(\to\) denotes convergence in distribution.

Proposition 2.17 is proved by applying weak convergence results that we develop in appendix I. With these results, we can develop approximations for a wide variety of diffusion models with time-varying drift and conditional variance. We can approximate many diffusions of the general form

\[dS_t = f(S_t, Y_t, t)dt + g(S_t, Y_t, t)dW_{1,t}\]

\[dY_t = F(S_t, Y_t, t)dt + G(S_t, Y_t, t)dW_{2,t}\]

\[\begin{bmatrix}
dW_{1,t} \\
dW_{2,t}
\end{bmatrix} = \Omega dt\]

where \(Y_t\) is a vector of (possibly unobserved) state variables and \(X_t\) is an observed process. We must also impose local boundedness conditions on \(f, g, F\) and \(G\), and we
require \( \text{rank}(\Omega) \leq 2 \).

Even with these restrictions, a wide variety of processes are included in 2.18 through 2.20. In 2.11 for example, \( \ln(\sigma_t^2) \) followed a continuous time AR(1), but 2.18 through 2.20 allows \( \ln(\sigma_t^2) \) to be an AR(k) or the sum of AR processes. Similarly, in 2.10 \( S_t^* \) was a random walk with time-varying drift, but we can extend this to allow more serial dependence. We can also allow seasonals in both the \( S_t^* \) and \( \sigma_t^2 \) processes. The interested reader is referred to appendix I for details.
III. ESTIMATION AND SPECIFICATION TESTING

As in the conventional ARCH model, we can write the log likelihood function as

\[ L_T = \frac{1}{T} \sum_{t=1}^{T} \ln \sigma_t^2 \]

where

\[ \ln \sigma_t^2 = \text{constant} - (1/2)[\ln(\sigma_t^2) + z_t^2] \]

\[ = \text{constant} - (1/2)[\ln(\sigma_t^2) + \sigma_t^{-2} \xi_t^2] \]

The first-order conditions for maximizing the likelihood with respect to a vector of unknown parameters \( b \) are

\[ \nabla_b \ln \sigma_t = (2\sigma_t^2)^{-1} \left( \xi_t^2 / \sigma_t^2 - 1 \right) \nabla_b \sigma_t^2 \]

\[ + \sigma_t^{-2} \xi_t \nabla \xi_t \]

The Hessian is

\[ \nabla_b^2 \ln \sigma_t = \left( \xi_t / 2\sigma_t^2 \right) \nabla_b \xi_t \nabla_b \sigma_t^2 + \left( 1/2 \sigma_t^2 \right) \nabla_b \xi_t \nabla_b \sigma_t^2 \xi_t \nabla \xi_t \]

\[ - 3\xi_t / 2\sigma_t^4 \nabla_b \xi_t \nabla_b \sigma_t^2 \xi_t \nabla \xi_t \nabla \xi_t \]

\[ - \sigma_t^{-4} \left[ 1/2 - \xi_t^2 / \sigma_t^2 \right] \nabla_b \sigma_t^2 \nabla_b \xi_t^2 \]

In the model developed in section II and discussed in this section, \( \xi_t | \sigma_t^2 \) is normally distributed. We could easily give \( \xi_t | \sigma_t^2 \) some other distribution—i.e., we could make \( \xi_t | \sigma_t^2 \) student, as has been done in a traditional ARCH model by Engle and Bollerslev (1986.)

Under sufficient regularity conditions, the maximum likelihood estimator will be consistent and asymptotically normal (CAN). Unfortunately, verifying that these conditions are satisfied has proved extremely difficult in
both traditional ARCH models and in the exponential form introduced in this paper. Weiss (1986) has developed a set of sufficient conditions for the CAN property in the simple ARCH model. These conditions are quite restrictive, and are not satisfied by the coefficient estimates obtained in most studies using the ARCH model. In the GARCH-M model, in which conditional variance appears in the conditional mean, the asymptotics are even more cloudy, and no sufficient conditions for CAN in this model are yet known. The asymptotics in models using the exponential form of ARCH introduced in this paper are equally difficult, and the asymptotic distribution of the estimated coefficients of an exponential ARCH model is as yet unknown.

I conjecture, however, that the methods developed in appendix I can be used to develop conditions for consistency and asymptotic normality. These methods provide conditions for a sequence of Markov chains to converge in distribution to an Ito process. Since the gradient and hessian of the likelihood are themselves Markov chains, it seems likely that we can show that they converge to Ito processes under regularity conditions. We could then apply conditions developed in Kutoyants (1975) for the asymptotic normality of stochastic integrals to prove consistency and asymptotic normality. This remains to be carried out, however, and in the remainder of this paper we will assume (as is the practice of researchers using ARCH methods) that the maximum
The theory of specification testing in models with
dynamic conditional variance has recently been developed by
Wooldridge (1987). Wooldridge builds on the conditional
moment tests developed in Newey (1985). These tests are
based on orthogonality conditions implied by correct model
specification. Since there are typically an infinite number
of orthogonality conditions implied by a model (i.e. a
requirement that a particular error term is uncorrelated at
all leads and lags) the conditional moment tests are not
consistent. That is, a model can be misspecified without
the error being detected by the test, even asymptotically.
These tests are not, therefore, tests against general
misspecification. We can, however, test against whatever
kind of misspecification seems potentially most important
within the context of a particular model.

In many cases, maximum or quasi maximum likelihood
estimators are obtained by minimizing a function of one step
ahead prediction errors of a random vector \( y_t \) given a random
vector \( X_t \), which may include lagged values of \( y_t \). Define
the normal quasi-maximum likelihood estimator \( \beta^* \) of a
parameter vector \( \beta \) by

\[
3.5) \quad \beta^* = \arg\max_{\beta} \left\{ \sum_{t=1}^{T} R_t \right\}
\]

where

\[
3.6) \quad R_t = -(1/2) \ln|\Sigma_{k=1}^{T} (b)| - (1/2)u_t(b)^\prime \Sigma_{k=1}^{T} (b)^{-1} u_t(b),
\]

\[
3.7) \quad u_t = y_t - m_t(X_t,b),
\]

128
$m_t$ is the conditional mean of $y_t$ given $X_t$, and $\eta(b)_t$ is the conditional covariance matrix of $y_t$ at time $t$ given $X_t$.

$\eta(b)_t$ may depend on $X_t$. Suppose that the null hypothesis $H_0$ specifies conditional means $m_t$ and the conditional variances correctly, and that this null hypothesis implies a set of $n$ conditional moment restrictions $r_j$, $j=1,\ldots,n$. That is, for all $t$, $t=1,\ldots,T$, we have under $H_0$

3.8) \[ E(r_j \mid y_t, X_t, b) \mid X_t = 0, \quad j=1,\ldots,n \]

For notational convenience we will write $r_j \mid (y_t, X_t, b) \mid X_t = r_j \mid (b)$. To test the orthogonality conditions implied by 3.8, we first define the test statistic

3.9) \[ R_T = T^{-1/2} \sum_{t=1}^{T} [r_{1t}(b^*), \ldots, r_{nt}(b^*)]' \]

\[ = T^{-1/2} \sum_{t=1}^{T} R_t(b^*) \]

and define

3.10) \[ O_T = T^{-1} \sum_{t=1}^{T} \begin{bmatrix} R_t(b^*) \end{bmatrix} [R_t(b^*), \ldots, R_t(b^*)]' \]

3.11) \[ G_T = T^{-1} \sum_{t=1}^{T} \nabla R_t(b^*) \]

3.12) \[ I_n = \text{an n by n identity matrix} \]

3.13) \[ H_T = T^{-1} \sum_{t=1}^{T} \nabla^2 R_t \]

3.14) \[ P_T = \begin{bmatrix} I_n & 0 \\ -G_T H_T^{-1} \end{bmatrix} O_T \begin{bmatrix} I_n & -H_T^{-1} G_T' \end{bmatrix} \]

129
Then, subject to regularity conditions found in Wooldridge (1987)

3.15) \( R'_t P_{t-1} R_t - x^2 c \), where \( c \) is the rank of plim \( P_t \).

The test statistic in 3.15 provides a specification test of the orthogonality conditions in 3.8, and may be used to test for serial correlation, conditional heteroskedasticity not captured in the model, and many other hypotheses of interest.

To close this section, I present a proposition that I believe to be true, but which I have not yet proven:

**Proposition 3.1:**

Suppose that the true process for excess returns an conditional variance is a diffusion process such as 2.10-2.12. Define \( b \) to be the vector of true system parameters, and define \( b^*_t \) to be the maximum likelihood estimates from the continuous record of observations of \( S \) on the time interval \([0,T]\). Define \( b_{h,t} \) to be the quasi-maximum likelihood estimator of \( b \) obtained by maximizing the likelihood function for the system 2.13-2.15. Then

3.16) \( b_{h,t} \rightarrow b^*_t \) almost surely as \( h \rightarrow 0^+ \).

As we saw in section II, the system 2.13-2.15 converges in distribution to the system 2.10-2.12 for parameter \( b \). But the maximum likelihood function is based on the densities corresponding to different values of \( b \), not on the
distribution, and convergence in distribution does not necessarily imply the convergence of the corresponding densities (Feller (1971)). This convergence of the quasi-maximum likelihood estimators has been shown to hold for linear stochastic differential equations (Bergstrom (1984)) but it is quite difficult to show a parallel result for an exponential ARCH model, since the estimator does not have a closed form.
IV: A SIMPLE MODEL OF THE MARKET RISK PREMIUM

In this section, we estimate and test a simple model of market risk, asset returns and changing conditional volatility. We will use this simple model to examine several issues that have been investigated in the economics and finance literature, namely 1) the relation between the level of risk and required rates of return, 2) the persistence of shocks to volatility, and 3) the relative importance of leverage-related movements in conditional variance (i.e. the direction of returns) vs. magnitude effects (the size of returns) in determining conditional variance.

We use the model developed in section II for the conditional variance process, assuming an ARIMA representation for the log of conditional variance. Specifically, we model the log of conditional variance as being the sum of a stochastic trend and a zero-mean, stationary (ARMA) component $S_t$, i.e.

4.1) \[ \ln(\sigma_t^2) = \text{Trend}_t + S_t \]

where

4.2) \[ \text{Trend}_t = \text{Trend}_{t-1} + \theta_1 z_{t-1} - \gamma_1 [|z_{t-1}| - E|z_{t-1}|] \]

4.3) \[ S_t = \sum_{i=1}^{p} a_i S_{t-i} + \sum_{j=1}^{q} b_j [\theta_2 z_{t-j} - \gamma_2 [|z_{t-j}| - E|z_{t-j}|]] \]
and the roots of \(1 - \sum_{i=1}^{p} \alpha_i y_i\) lie outside the unit circle.

The theory of stochastic trend modeling was developed originally in Watson (1986). In Watson's models, and in most stochastic trend models, the covariance between the innovations in the trend and stationary components is not identified and so cannot be consistently estimated. The reason for this is that there are two separate innovations which cannot be disentangled from only one observed data series. This problem is usually dealt with by assuming either zero or perfect correlation between the innovations. In our model, however, \(\delta_1, \delta_2, \theta_1,\) and \(\theta_2\) are all identified, since the second innovations component \(|z_t| - E|z_t|\) is a simple function of the first \((z_t)\) and the two can therefore be disentangled.

The purpose of introducing a stochastic trend is to try to analyze the long-run behavior of volatility. But if by "long-run" we mean serial dependence at lags of, say, five years, we must keep in mind that our data set contains only about twelve non-overlapping five-year intervals. As Cochrane (1986) has forcefully pointed out, parametric time series models impose restrictions on the relation between the system's low and high frequencies, so that much of the information about the system's low-frequency behavior actually comes from its behavior at high frequencies. If the parametric model is not correct, the predicted low
frequency behavior may be very inaccurate even when the high
frequency behavior is approximately correct. Phillips
(1987), Cochrane (1986) and Quah (1987) have recently
developed non-parametric approaches to testing for a unit
root in linear processes. It remains to be seen if any of
these approaches can be applied to a highly non-linear model
such as exponential ARCH, and for the time being our
analysis must take place within a parametric framework. We
need therefore to be quite careful in making long term
forecasts based on our estimated models.

A constant on the right-hand side of 4.2 was ruled out
on theoretical grounds, since a simple application of the
strong law of large numbers will show that the presence of a
positive drift implies that $\sigma_1^2$ goes to infinity almost
surely, and a negative drift implies that $\sigma_1^2$ goes to 0
almost surely. Perhaps a better specification in 4.3,
however, would be to allow the presence of a bounded
deterministic function of time $\alpha_t$, reflecting, for example,
seasonal patterns in variance, or accounting for non-
trading days (French and Roll (1986)). In future versions
of the paper, I hope to do this. For the present, however,
we exclude seasonals in variance and assume that non-trading
periods make no contribution to conditional variance.

To allow for the possibility that the $\{z_t\}$ sequence may
not be normal, we modelled $\{z_t\}$ as independent draws from a
generalized error distribution (Harvey, (1981)). The
generalized error distribution (GED) includes the normal as a special case, along with many other distributions, some more fat-tailed than the normal (for example the double exponential), some more thin tailed (i.e. the uniform). Although GED distributions can be more fat tailed than the normal, they still possess finite moments of all orders, so that the convergence theorems of appendix I apply.

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

$$f(z) = \frac{v \exp[-0.5^v |z/\lambda|^v]}{\lambda \Gamma(v/\lambda) \Gamma(1/v)}$$

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

$$\lambda = \left[ 2(-2/v) \Gamma(1/v) / \Gamma(1/v) \right]^{1/2}$$

$v$ is a tail-thickness parameter. When $v = 2$, $z$ is standard normal. For $v < 2$, the distribution of $z$ has thicker tails than the normal (i.e. when $v = 1$, $z$ has a double exponential distribution) and for $v > 2$, the distribution of $z$ has thinner tails than the normal (i.e. for $v = \infty$, $z$ is uniform.)

The returns data is derived from daily observations on the level of the S&P 500 index from January 1928 through October 23, 1987. This data from 1928 through 1985 was provided to the author by Kenneth French, G. William Schwert and Robert Stambaugh, who originally gathered it for their (1986) study of market volatility. Unfortunately, neither daily dividends or riskless interest rate series are available for the sample period. One possibility would be
to use the Ibbotson and Singuefield (1982) monthly numbers for t-bills and dividends and assume that these remained constant for each day within the month. Since these numbers are not available for very recent months, and since it seems likely that the events of September and October 1987 carry considerable information about movements in stock market volatility, I elected to ignore dividends in the results presented in this draft of the paper. For a riskless interest rate series, I used the discount rate of the Federal Reserve Bank of New York. Since shocks to volatility persist indefinitely in the model I estimate, it may well be that small errors in returns measurement accumulate and severely affect the estimation results. Future drafts of this paper will, I hope, make better adjustments for dividends and riskless rates. Another possibility, of course, would be to use the daily CRSP data from 1962 to 1986. But to estimate long-run behavioral characteristics of a series requires data that stretches over a very long period. The least of all evils seemed to be to use the long data series, even with its measurement problems.

Our next task is to model required excess returns given the process for conditional variance we have outlined above. As we mentioned in section II, Merton (1973) showed that in an economy with a representative agent who has logarithmic utility, and in which cumulative returns and conditional
variance evolve as Ito processes, the instantaneous risk premium on the market portfolio is equal to the instantaneous conditional variance $\sigma_t^2$. If we make the observation time intervals small, then the skewness per unit of time and all other higher moments drop out, so that a good discrete time approximation to the continuous time model for excess returns on the market should be

$$4.6) \quad e_{r_t} = (1/2)\sigma_t^2 + \sigma_t z_t$$

Where $\sigma_t^2$ is the variance per unit of time and the time unit equals the length of the observation interval. A simple alternative against which to test 4.5 is

$$4.7) \quad e_{r_t} = a_0 + a_1 \sigma_t^2 + \sigma_t z_t$$

If the dividend yield is constant, $a_0$ should pick up the effects of the missing dividends.--i.e. if $\delta$ is the (constant) dividend yield

$$4.6)' \quad e_{r_t} - \delta = -\delta + (1/2)\sigma_t^2 + \sigma_t z_t$$

Under the assumption of a constant dividend yield, therefore, we can estimate the model using only capital gains data and riskless rates by including a constant in the excess returns equation 4.7.

To complete the model, we must choose an observation interval. Efficient variance estimation requires measurement intervals as small as possible. Unfortunately, since we are modelling returns on a market index, the effects of non-trading in the underlying securities induces a spurious autocorrelation in returns (Merton, 1980.) Such
effects should be small in the heavily-traded stocks comprising the S&P 500, and several authors have argued that the measured autocorrelation in index returns is much too large to be explained by non-trading effects (see, for example, Lo and MacKinlay (1987). Lo and MacKinlay argue that measuring returns over intervals of a few days should eliminate virtually all autocorrelation induced by non-trading.

Accordingly, we used an observation interval of four days which gave us 4020 observations in the data set. We are throwing away quite a lot of variance information by doing this, however, so that it may well be better to model non-trading effects directly and use the daily observations. This will be done in a future version of this paper.

The results we report below are for a simple model in which the stationary component is an AR(1). I also estimated an ARMA(1,1), but the results were nearly identical to the AR(1) model (i.e. the MA parameter was quite small (about 0.2) and not significantly different from zero) and so I will not report it below. In future work it will certainly be necessary to estimate more general models for the stationary component. Specification tests will also have to wait for a future version of this paper.

For the AR(1) + stochastic trend model, the estimated
coefficients (with standard errors in parenthesis)\(^3\) were:

\[
\begin{align*}
4.8) & \quad \text{Trend}_0 = -7.8997 \quad \text{S}_0 = -1.8688 \\
& \quad (0.5006) \quad (0.7387)
\end{align*}
\]

\[
\begin{align*}
4.9) & \quad \text{S}_t - \text{S}_{t-1} = 0.0012 - 0.2374*\text{S}_t^2 + \text{S}_t \text{Z}_t \\
& \quad (0.0003) \quad (0.055)
\end{align*}
\]

\[
\begin{align*}
4.10) & \quad \ln(\text{S}_t^2) = \text{Trend}_t + \text{S}_t \\
4.11) & \quad \text{Trend}_t = \text{Trend}_{t-1} + 0.11027(|z_{t-1} - E|z|) - 0.00362*z_{t-1} \\
& \quad (0.0212) \quad (0.0083)
\end{align*}
\]

\[
\begin{align*}
4.12) & \quad \text{S}_t = 0.7975*\text{S}_{t-1} + 0.04458(|z_{t-1} - E|z|) - 0.18510*z_{t-1} \\
& \quad (0.0362) \quad (0.03680) \quad (0.02705)
\end{align*}
\]

\[
\begin{align*}
4.13) & \quad z_t \sim \text{i.i.d. GED with parameter v} = 1.4529 \\
& \quad (0.101)
\end{align*}
\]

Recall that \(\text{var}[\theta z + \gamma(\text{z} - E|\text{z}|)] = \theta^2 + \gamma^2[1 - E^2|\text{z}|].\)

For the model just estimated, \(E|\text{z}| = 0.7634\) and \([1 - E^2|\text{z}|] = 0.41719.\) Using this, the variance of the innovations components in the trend and stationary components of log variance are easily found to be \(0.00525\) and \(0.03509\) respectively. The random-walk component changes much more slowly, but its changes persist over time, whereas the

\[\]
shocks in the stationary component die out quite quickly— i.e. with an AR coefficient of 0.7975, shocks in the stationary component have a half-life of about two and a half weeks. One thing that is surprising is the extent to which the leverage and non-leverage effects are divided between the stationary and trend innovations; the leverage-related effects (the $\theta z_t$ term) account for almost 98% of the variance of the stationary component's innovations and about 1/5 of 1% of the innovations variance in the stochastic trend.

One surprising (and somewhat disturbing) aspect of the estimated coefficients is the negative coefficient estimate on the $\sigma^2$ term in the returns equation. As we saw earlier, this coefficient should equal 1/2 if investors all have log utility, and a value under -1/2 would mean that there is a negative reward for risk bearing. With an estimated coefficient of -0.2374, the evidence seems to point to a very weak effect of risk on expected returns. It may be that by omitting dividends we have destroyed any information the data series could give us about the relation of risk and return.

The essential properties of the series are most clearly explained by figures 1 through 6. Figure 1 plots the fitted trend and total log of four-day variance for 1986 through the end of October 1987, with forecast values through March 1988. The solid line is the trend component and the dashed
Figure 1
line is the stationary component. The enormous jump in October 1987 is precipitated by the crash of that month. The model forecasts that much of the increase in volatility will wear off by the end of 1987, but that the last two months of 1987 may well be quite volatile. There is a "permanent" component in the volatility increase, however, that leads us to forecast higher volatility for the indefinite future than would have occurred without the crash.

This "permanent" component, however, should be interpreted with extreme caution: As we noted above, we have only about twelve non-overlapping five-year intervals in our data set, so that we have little statistical power to forecast at this horizon. The model's assumption of an AR(1) + random walk imposed the existence of a "permanent" component. I also estimated a model with log variance following a stationary ARMA(2,2), and could not reject the null hypothesis that the largest AR root equalled one. This does not mean, however, that it is one. The largest estimated AR root was 0.9967 with a standard error of 0.0063. Shocks associated with this AR root would have a half life of about three and one-half years, so that although the short-run forecasting properties of the two models are the same, but the long-term properties are very different. (I have not formally presented the other model because it was done under slightly different assumptions
than the models we have reported. Clearly, a great deal of empirical work remains to be done on this model.

Figure 2 plots the random walk component of the log of four-day variance for the entire sample period, and figure 3 plots the stationary component, figure 4 the total. Figure 5 plots the trend and total from late 1971 on (again, the solid line is the trend and the dashes and dots are the total). Finally, figure 6 plots the annualized standard deviation through the entire sample period and the forecast beyond. Again, if instead of a random walk component we had an AR root of 0.9967, the long-term forecast would not be horizontal, but rather would slope gently downward, returning half of the distance to its long-term level in about 3.5 years.
V: CONCLUSION

This paper has presented a new approach to modelling the time series behavior of variance in asset markets that does not suffer from some of the drawbacks of traditional ARCH and GARCH models. A very preliminary application was presented that modelled changing returns variances and the price of risk. The extension of the exponential ARCH model to multivariate systems awaits future research.
APPENDIX I: PASSAGE TO CONTINUOUS TIME

In this appendix we present methods developed by Stroock and Varadhan (1979) for approximating stochastic differential equations with stochastic difference equations. We apply the methods we develop to both classical and exponential ARCH models. In particular, we show how to create a sequence of exponential ARCH models that converge in distribution to the sort of Ito process models that are often found in the finance literature. This section also provides a more formal argument on why classical ARCH models are unlikely to have well-behaved continuous time versions.

The plan of this appendix is as follows: we first present the basic Stroock-Varadhan approximation results, and apply these results to develop a constructive method for approximating stochastic differential equation systems using exponential ARCH models. The we reexamine the passage to continuous time in traditional ARCH models.

Convergence of Stochastic Difference Equations to Stochastic Differential Equations: General Results

In this subsection we present general conditions for a sequence of finite dimensional, discrete time Markov processes to converge to an Ito process. As we noted earlier, these are drawn largely from Stroock and Varadhan
(1979). These results have been extended to diffusion-jump processes by Kushner (1984), but it would be beyond the scope of this paper to deal with this more general case.

Our general setup is as follows: Let $D([0,\infty),\mathbb{R}^n)$ be the space of mappings from $[0,\infty)$ into $\mathbb{R}^n$ that are continuous from the right with finite left limits. $D$ is a metric space when endowed with the Skorohod metric. (Billingsley (1968)) For each $h > 0$, let $\mathcal{M}_{kh}$ be the $\sigma$-algebra generated by $x_0$, $x_h$, $x_{2h}$,...,$x_{kh}$, let $B(\mathbb{R}^n)$ denote the Borel sets on $\mathbb{R}^n$, and let $\Pi_h(x,\cdot)$ be a transition function on $\mathbb{R}^n$ -- i.e.,
a) $\Pi_h(x,\cdot)$ is a probability measure on $(\mathbb{R}^n,B(\mathbb{R}^n))$ for all $x \in \mathbb{R}^n$.
b) $\Pi_h(\cdot,\Gamma)$ is $B(\mathbb{R}^n)$ measurable for all $x \in B(\mathbb{R}^n)$.
c) For all $kh$ and for all $\Gamma \in B(\mathbb{R}^n)$, $P[x_{(k+1)h} \in \Gamma | \mathcal{M}_{kh}] = \Pi_h(x_{kh},\Gamma)$.

Given $x \in \mathbb{R}^n$, let $P_{h,x}$ be the probability measure on $D([0,\infty),\mathbb{R}^n)$ such that

A1.1) $P_{h,x}[x_0 = x_0^*] = 1$

A1.2) $P_{h,x}[x_t = x_{kh}, kh \leq t \leq (k+1)h] = 1$

A1.3) $P_{h,x}[x_{(k+1)h} \in \Gamma | \mathcal{M}_{kh}] = \Pi_h(x_{kh},\Gamma)$ a.s. $P_{h,x}$ for all $k \geq 0$ and $\Gamma \in B(\mathbb{R}^n)$.

$x$ is an $n$-dimensional Markov process with a known starting point $x_0^*$ given by A1.1 and transition probabilities given by A1.3. By A1.2, $x$ is a step function with jumps at times $h$, $2h$, $3h$ and so on. Note that the transition probabilities in A1.2 can be made to depend on
time by the simple trick of making time itself an element of x (i.e. time should be jointly Markov with almost any process of interest.)

For each \( h, \varepsilon > 0 \), define

**A1.4)** \[ a_h(x_k, h) = h^{-1} \int (x_h(k+1) - x_h) (x_h(k+1) - x_h) \, \Pi_h(x_k, dx_h(k+1)) \quad |x_h(k+1) - x_h| < 1 \]

**A1.5)** \[ b_h(x_k, h) = h^{-1} \int (x_h(k+1) - x_h(k)) \, \Pi_h(x_k, dx_h(k+1)) \quad |x_h(k+1) - x_h(k)| < 1 \]

**A1.6)** \[ \Delta_h, \varepsilon(x_k, h) = h^{-1} \int \Pi_h(x_k, dx_h(k+1)) \quad |x_h(k+1) - x_h| > \varepsilon \]

\( \Delta_h, \varepsilon(x) \) is a measure of the probability of a large jump over the interval of length \( h \). The integration in A1.4 and A1.5 is taken over \( |x_h(k+1) - x_h| < 1 \) rather than over \( \mathbb{R}^n \) because the regular conditional moments may not be finite. When they are finite, \( a_h(x) \) and \( b_h(x) \) are measures of the variance and drift respectively per unit of time. The convergence results we present below will require that \( a_h(x) \) and \( b_h(x) \) converge to finite limits, and that \( \Delta_h, \varepsilon(x) \) goes to zero. If a conditional moment greater than two exists, then \( \Delta_h, \varepsilon(x) \) will automatically go to zero as \( h \to 0 \) (by Chebychev's inequality) and the integrals in A1.4 and A1.5 may be taken over \( \mathbb{R}^n \). (see the discussion in Basawa and Rao (1980) section 9.2.)

The convergence results given below will require the
following assumptions:

ASSUMPTION 1) For all $e > 0$ and $R > 0$, there exists a continuous mapping $a(x)$ from $\mathbb{R}^n$ into the space of $n$ by $n$ non-negative definite matrices and a continuous mapping $b(x)$ from $\mathbb{R}^n$ into $\mathbb{R}^n$ such that

A1.7) \[ \lim_{h \to 0^+} \sup_{|x| < R} \left[ a_h(x) - a(x) \right] = 0_{n \times n} \]

A1.8) \[ \lim_{h \to 0^+} \sup_{|x| < R} \left[ b_h(x) - b(x) \right] = 0_n \]

A1.9) \[ \lim_{h \to 0^+} \sup_{|x| < R} \Delta_{e,w}(x) = 0 \]

where $0_{n \times n}$ is an $n$ by $n$ matrix of zeros and $0_n$ is an $n$ by one vector of zeros.

ASSUMPTION 2) There exists a continuous mapping $\sigma(x)$ from $\mathbb{R}^n$ into the space $n$ by $n$ non-negative definite matrices such that

A1.10) \[ a(x) = \sigma(x)\sigma(x)'. \]

ASSUMPTION 3) $a(x)$, $\sigma(x)$ and $b(x)$ are locally bounded (i.e. bounded for bounded $x$.)

ASSUMPTION 4) The process $X_t$ defined by the stochastic integral equation system
\begin{align}
A1.11) \quad X_t &= X_0 + \int_0^1 b(X_s) \, ds + \int_0^1 \sigma(X_s) \, dW_s, \\

\text{where } W_s \text{ is an n-dimensional standard Brownian motion}, \\
\text{remains bounded in finite time intervals with probability one -- i.e. for every } T < \infty, \\

A1.12) \quad P[ \sup_{t \in [0,T]} |X_t| < \infty ] = 1. 
\end{align}

**THEOREM A1.1**

Under assumptions 1 through 4, the $x_{K,n}$ process defined by A1.1 through A1.3 converges in distribution to the $X_t$ process defined by the stochastic integral equation system A1.11.4

**Proof:** see appendix II

**APPLICATION TO EXPONENTIAL ARCH PROCESSES**

In this subsection we will use theorem A1.1 to develop a more general method for approximating Ito processes by exponential ARCH processes. The results are not as general as one could hope, for several reasons. First, the exponential ARCH models developed thus far are strictly univariate--i.e. only one observable random series is allowed. Multivariate models are a logical next step for

\footnote{I would like to thank Daniel Stroock for answering so many questions about this theorem.}
future research. Second, we will be able to approximate only systems where the covariance matrix of the underlying Brownian motions has at most rank 2. This assumption is easy to relax, but to do so would require that we move outside the framework of exponential ARCH processes.

Finally, theorem A1.1 applies only to processes that have finite-order Markov representations. In discrete time, finite order ARIMA processes have finite-order Markov representations. As we pass to continuous time, however, the number of states required to represent a moving-average process goes to infinity, which is not allowed in theorem A1.1. AR processes, however, do not require more states in their Markov representations as time is partitioned more finely. So our results will be limited to models in which the state vector remains finite as we pass to continuous time. I believe that convergence results analogous to theorem A1.1 can be shown for more general processes, but this is a conjecture.

Consider the following stochastic differential equation system

\[ dS_t = f(S_t, Y_t, t)dt + g(S_t, Y_t, t)dW_{1,t} \]
\[ dY_t = F(S_t, Y_t, t)dt + G(S_t, Y_t, t)dW_{2,t} \]

\[ \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} = \Omega \, dt \]

where \( Y \) is an \( n \)-dimensional vector of (unobservable) state
variables, S is an (observable) scalar process, $W_1$ is a one-dimensional standard Brownian motion, $W_2$ is an n-dimensional Brownian motion, $f$ and $g$ are real-valued, locally bounded, continuous scalar functions, $F(\cdot,\cdot)$ and $G(\cdot,\cdot)$ are real, locally bounded, continuous n-dimensional functions, and $1 \leq \text{rank}(\alpha) \leq 2$. We will also assume that the process given in A1.13-A1.15 is non-explosive: i.e. for any finite time interval $[0,T]$,

$$A1.16 \quad P[\sup_{t \in [0,T]} \max |S_t|, |Y_t| < \infty] = 1.$$ 

General conditional for explosion or non-explosion can be found in Stroock and Varadhan (1979) chapter 10.

Next we propose a sequence of approximating processes that converge to A1.13-A1.15 in distribution. Then we will then explain the intuition behind this approximation. The approximating process is given by

$$A1.17 \quad S_{h_{(k+1)}} = S_{h_k} + f(S_{h_k}, Y_{h_k}, hk)h + g(S_{h_k}, Y_{h_k}, hk)Z_{h_k}$$

$$A1.18 \quad Y_{h_{(k+1)}} = Y_{h_k} + F(S_{h_k}, Y_{h_k}, hk)h + G(S_{h_k}, Y_{h_k}, hk)Z_{h_{h_k}}$$

where

$$A1.19 \quad Z_{h_k} \sim \text{i.i.d. } N(0,h)$$

$$A1.20 \quad Z_{h_{h_k}} = \begin{bmatrix}
\Theta_1 Z_{h_k} + \xi_1 \left[ |Z_{h_k}| - \left(2h/\pi\right)^{1/2} \right] \\
\vdots \\
\Theta_n Z_{h_k} + \xi_n \left[ |Z_{h_k}| - \left(2h/\pi\right)^{1/2} \right]
\end{bmatrix}$$
and \( \Theta_1, \Theta_2, \ldots, \Theta_n, \Theta_n \) are selected so that

\[
A1.20) \quad E \begin{bmatrix} Z_{h,k} \\ Z_{h,k}^* \end{bmatrix} \begin{bmatrix} Z_{h,k} & Z_{h,k}^* \end{bmatrix} = \Omega h
\]

**THEOREM A1.2**

\[
A1.21) \quad \text{As } h \to 0, (S_{h,k}, Y_{h,k}) \to (S_t, Y_t)
\]

Theorem A1.2 is proved by a straightforward application of theorem A1.1. The details are in appendix II. Intuitively, \((S_{h,(k+1)} - S_{h,k}), (Y_{h,(k+1)} - Y_{h,k})\) and \(h\) are the discrete time equivalents of \(dS, dY\) and \(dt\) respectively. It is a little bit harder to see why \(Z_{h,k}\) and \(Z_{h,k}^*\) are the discrete time equivalents of \(dW_{1,t}\) and \(dW_{2,t}\). To see that they are, the following proposition may be helpful:

**Proposition A1.22**

Let \(W_t\) be a standard Brownian motion on \([0,1]\). For each \(k = 1, 2, \ldots\) and each \(t \in [0,1]\) define \(Q_{k,t}\) by

\[
Q_{k,t} = (1 - 2/\pi)^{-1/2} \sum_{j=1}^{k-1} \frac{|W_{j+1}/k - W_j/k| - (2/\pi k)^{1/2}}{
\]

Then as \(k \to \infty\),

\[
\begin{bmatrix} Q_{k,t} \\ W_t \end{bmatrix} \overset{D}{\to} W_t^{**}
\]

where \([kt]\) is the integer part of \(kt\) and \(W_t^{**}\) is a two-
dimensional standard Brownian motion on \([0,1]\).^5

*Proof: in appendix II*

Although \(Q_{k,t}\) is a function of the path of \(W_t\), as \(k \to \infty\), it converges in distribution to a brownian motion that is independent of \(W_t\). Proposition A1.22 tells us therefore that by observing the sample path of one Brownian motion, we can manufacture a sequence of random processes that will converge in distribution to a second, independent Brownian motion. To manufacture a sequence of processes that converges to a Brownian motion that is imperfectly correlated with \(W_t\), we simply take a linear combination of \(W_t\) and \(Q_{k,t}\). Heuristically, this is why \(Z_{hk}\) and \(Z^*_{hk}\) are the discrete time counterparts of \(dW_1\) and \(dW_2\).

We could allow \(n\) to have rank greater than two by considering \(||Z_1-\mathbb{E}|Z_1|| - \mathbb{E}||Z_1-\mathbb{E}|Z_1||\) and so on, but this would take us outside the realm of exponential ARCH processes.

Often the most difficult aspect of proposition A1.22 to

---

5) Richard Dudley pointed out that \(Q_{k,t}\) does not converge in probability to any Brownian motion, since

\[
\begin{bmatrix}
Q_{k,t} \\
Q_{k\*,k,t} \\
W_t
\end{bmatrix} \overset{D}{\to} W_t^{***}
\]

where \(W_t^{***}\) is a three dimensional standard brownian motion. In other words, the \(Q\) function corresponding to a fine partition of time is asymptotically independent of the \(Q\) function corresponding to a partition that is much finer still.
apply is to show that the stochastic differential equation system A1.13-A1.15 is non-explosive. As we mentioned earlier, Stroock and Varadhan (1979) provide criterion for explosion and non-explosion. But these criteria are not always easy to apply, and sometimes the following result will be helpful:

THEOREM A1.3
Suppose that the system A1.13-A1.15 can be simplified to:

A1.13)* \[ dS_t = f(Y_t, t)dt + g(Y_t, t)dW_t, \]

A1.14)* \[ dY_t = F(Y_t, t)dt + G(Y_t, t)dW_t, \]

A1.15)* \[
\begin{bmatrix}
dW_{t1}
dW_{t2}
\end{bmatrix}
= \Omega \begin{bmatrix}
dW_{t1}
dW_{t2}
\end{bmatrix}
\]

(i.e. that f, g, F and G are no longer functions of S) and suppose further that Y_t can be shown to be non-explosive (i.e. for every T < ∞, P[ \sup_{\tau \geq 0} |Y_\tau| < ∞ ] = 1.) Then S_t is also non-explosive.

Proof: see appendix II.

In proposition 2.8, ln(\(\sigma_t^2\)) obeys Lipschitz and growth conditions and so is non-explosive (Lipster and Shiryayev (1977)). The non-explosion of S_t* follows immediately by Theorem A1.3.

It is easy to apply the results of this section to show
that exponential ARCH processes in which $\ln(\sigma^2)$ either has a finite-order AR representation or else is the sum of a finite number of finite order AR processes has a well-defined continuous time limit. For discussion of continuous time AR processes and their relation to discrete time AR processes, see Priestly (1981) chapter 3.

Finally, we note that all of the results of this section follow if, instead of $h^{-1/2}Z_{h,k}$ being i.i.d. $N(0,1)$, we allow $h^{-1/2}Z_{h,k}$ to be i.i.d. with any symmetric distribution possessing a finite absolute moment of order higher than two. By Chebychev's inequality, a finite moment higher than two will guarantee that A1.9 is satisfied -- i.e. that

$$A1.9 \quad \lim_{h \to 0^+} \sup_{|x| < R} \Delta_{e,h} (x) = 0.$$ 

**Passage to Continuous Time in Standard ARCH Models**

In this subsection we reexamine the continuous time limits of traditional ARCH processes. For the sake of simplicity we will examine only the most commonly used ARCH model, the GARCH(1,1) process developed by Bollerslev (1986). The extension to more general ARCH models is straightforward. The GARCH(1,1) process for conditional variance is

$$A1.23 \quad \sigma_{h,k}^2 = \omega + \sigma_{h,(k-1)}^2 [\epsilon_h + \alpha_h Z_{h,(k-1)}^2]$$
where \([z_{h,k}] \sim \text{i.i.d. } N(0,h)\).

\(\omega, \alpha\) and \(\delta\) have \(h\) subscripts because as we reduce \(h\) toward zero, since our object is to discover which sequences \([\omega_h, \alpha_h, \delta_h]\) make their corresponding \(\sigma_h^2\) process converge in distribution to a diffusion limit. For all \(h\), however, we require that \([\omega_h, \alpha_h, \delta_h]\) be non-negative. In addition \(\delta_h\) is bounded above by one and \(h\alpha_h\) must also be bounded above. These requirements are necessary to ensure non-explosion and non-negativity of \(\sigma_h^2\) (Nelson 1988.)

The process in A1.23 is clearly Markov, and the drift and second moment per unit of time are given by

\[
A1.24) \ E[(\sigma_h^2-k_{(h)}^2)/h] = \omega_h/h + (\delta_h-1 + h\alpha_h)\sigma_h^2_{(h)}/h
\]

\[
A1.25) \ E[(\sigma_h^2-k_{(h)}^2)^2/h] = \omega_h^2/h + 2\omega_h (\delta_h-1+h\alpha_h)\sigma_h^2_{(h)}/h + (\delta_h-1)^2\sigma_h^2_{(h)}/h + 3\omega^2_h h\sigma_h^2_{(h)}^4 + 2\alpha_h (\delta_h-1)\sigma_h^2_{(h)}^4
\]

It is straightforward though tedious to verify that the third moment also exists, is finite and disappears strictly faster than the second moment, so that for the conditions in assumption 1 we require only that A1.24 and A1.25 converge to finite limits.

For A1.24 to converge to a finite limit it is necessary that

\[
A1.26) \ \omega_h/h \to \omega^* > 0, \ \ (\delta_h-1+\omega_h)/h = O(h)
\]

161
One requirement for A1.25 to have a finite limit is that

A1.27) \( \alpha_h^2 h = O(1) \).

Together, the conditions in A1.26 and A1.27 imply that

A1.28) \( \beta_h - 1 = O(h) \)

and

A1.29) \( \alpha_h = O(1) \).

Using A1.26 through A1.29 it is easy to check that the right-hand side of A1.25 must go to zero—i.e. that the variance per unit of time disappears. Similarly, it is easy to check that any sequence \( \{ \omega_h, \alpha_h, \beta_h \} \) that makes the variance per unit of time converge to a positive limit makes the drift per unit of time go to infinity. Intuitively, the reason this happens is that the randomness of the conditional variance of an ARCH process comes from the squared normal random variables \( z_h^2 \), whose variance goes to zero much faster than their mean as \( h \to 0 \). Indeed, this is precisely why \( [dW]^2 = dt \) in Ito's lemma (Oksendal (1985)).

Unfortunately, theorem A1.1 gives sufficient rather than necessary conditions for convergence of a Markov chain to a diffusion. There is no known set of necessary and sufficient conditions for a sequence of Markov Chains to converge to a diffusion. This should not surprise us, since there is no known set of necessary and sufficient conditions for the central limit theorem to hold for an arbitrary sequence of random variables.
It is conceivable that there may be a weaker set of sufficient conditions under which ARCH processes could be shown to converge to some diffusion limit. This seems very unlikely, however, since the definition a diffusion process (i.e. Basawa and Rao, (1980)) will require that for the limit process

\begin{equation}
A.1.30 \quad a(x_t) = \lim_{h \to 0} h^{-1} \int \frac{(x_{t+h} - x_t)(x_{t+h} - x_t)'}{\Pi(x_t, dx_{t+h})} |x_{t+h} - x_t| < 1
\end{equation}

\begin{equation}
A.1.31 \quad b(x_t) = \lim_{h \to 0} h^{-1} \int \frac{(x_{t+h} - x_t) \Pi(x_t, dx_{t+h})}{|x_{t+h} - x_t| < 1}
\end{equation}

exist and are finite, and that

\begin{equation}
A.1.32 \quad \lim_{h \to 0} h^{-1} \int \frac{\Pi(x_t, dx_{t+h})}{|x_{t+h} - x_t| \geq e} = 0
\end{equation}

Since we can't get the moments in A.1.24 and A.1.25 to converge unless \( \lim a_h(x) = 0 \), it seems unlikely that one will ever find a diffusion process that is a limit of ARCH processes and satisfies A.1.30 through A.1.32 unless \( a(x) = 0 \).
APPENDIX II: PROOFS

Proof of theorem 2.1:

If z is standard normal, then straightforward integration yields:

A2.1) \[ E[\exp(\theta z + \delta |z|)] = \exp((\delta + \theta)^2/2)\phi(\delta + \theta) + \exp((\delta - \theta)^2/2)\phi(\delta - \theta) \]

2.6 follows immediately from A2.1 and 2.4. It remains only to show that the expectation in 2.6 is finite if \(|\beta_1| = o(i^{-b})\) for some \(b > 1\). An infinite product

\[
A2.2) \prod_{i=1}^{\infty} a_i
\]

converges to a finite, non-zero number iif (Gradshteyn and Ryzhik, 1980) the series

\[
A2.3) \sum_{i=1}^{\infty} [-1 + a_i]
\]

converges. In our problem, let \(a_i\) equal the \(i^{th}\) term in 2.4. Taking Taylor series expansions of \(e^x\) and \(\Phi(x)\) around zero yields

A2.4) \[ \Phi(x) = 1/2 + (2\pi)^{-1/2}x + O(x^2), \quad |x| < 1 \]

\[ e^x = 1 + x + O(x^2), \quad |x| < 1 \]

Substituting for \(\Phi\) and \(e^x\) in A.3 and carrying out the multiplication yields the following expression for the \(i^{th}\) term:

A2.5) \[ -1 + a_i = (\delta + \theta)\beta_1 (2/\pi)^{1/2} + O(\beta_1^2) \]

\[ = O(\beta_1) \]
which is $o(i^{-b})$ for some $b > 1$ if $B_t = o(i^{-c})$ for some $c > 1$. Q.E.D.

Proof of theorem A1.1:

First we need some definitions: Let $\mathfrak{n}$ denote the space of continuous mappings from $[0, \infty)$ into $\mathbb{R}^n$. Let $\omega$ be a generic element of $\mathfrak{n}$, and let $x(t, \omega) = \omega(t)$. Define the metric

$$A2.6) \quad D(\omega, \omega') = \sum_{n=1}^{\infty} 2^{-n} \left[ \frac{\sup_{0 \leq t \leq n} |x(t, \omega) - x(t, \omega')|}{1 + \sup_{0 \leq t \leq n} |x(t, \omega) - x(t, \omega')|} \right]$$

and define $\mathcal{M}$ to be the borel $\sigma$-field of subsets of $(\mathfrak{n}, D)$. Given a locally bounded, measurable mapping $a(t, x)$ from $\mathbb{R}^n$ into the space of real $n$ by $n$ non-negative-definite matrices, and a locally bounded, measurable mapping $b(t, x)$ from $\mathbb{R}^n \times \mathbb{R}^n$ into $\mathbb{R}^n$, and given $(s, x) \in \mathbb{R} \times \mathbb{R}^n$, a solution to the martingale problem for $a, b$ starting from $(s, x)$ is a probability measure $P$ on $(\mathfrak{n}, \mathcal{M})$ such that

$$A2.7) \quad P( x(t) = x, 0 \leq t \leq s ) = 1$$

and

$$A2.8) \quad \text{for any function } f \text{ from } \mathbb{R}^n \text{ into } \mathbb{R} \text{ that is bounded and possesses derivatives of all orders,}$$
\[
\begin{align*}
\int_{t}^{1} \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}(u,x(s)) \frac{\partial^2 f(x(u,\omega))}{\partial x_i \partial x_j} \right] du \\
+ \left[ \sum_{i=1}^{n} b_{i}(u,x(u,\omega)) \frac{\partial f(x(u,\omega))}{\partial x_i} \right]
\end{align*}
\]

where \( a_{i,j} \) and \( b_{i} \) are the \( i-j \)th and \( i \)th elements of the \( a \) and \( b \), is a \( P \)-martingale after time \( s \). The martingale problem for \( a \) and \( b \) is said to be well-posed if for each \( (s,x) \), there is exactly one solution to the martingale problem starting from \( (s,x) \).

Now on to the main proof: By Stroock and Varadhan (1979) corollary 10.1.5, assumptions 3 and 4 are sufficient for the martingale problem for \( a \) and \( b \) to be well-posed. For the case when condition A1.2 is replaced by

\[
\text{A1.2'} \quad P_{h,x} \{ x_t = [(k+1)h-t]x_{kh}/h + (t-kh)x_{(k+1)h}/h, \quad kh \leq t \leq (k+1)h \} = 1
\]

Stroock and Varadhan (1979) theorem 11.2.3 proved that when the martingale problem for \( a \) and \( b \) is well-posed and assumption 1 holds, then \( x_{h,t} \) converges in distribution to \( x_t \), the solution to the martingale problem for \( a \) and \( b \). That this still holds when A1.2' is replaced by A1.2 is proved in Ethier and Kurtz (1986), chapter 7, theorem 4.1 and corollary 4.2. Finally, under assumptions 2 through 4,
\( x_t \) has the representation A1.11 by Ethier and Kurtz (1986) chapter 5, theorem 3.3.

Q.E.D.

Proof of theorem A1.2:

To prove the theorem we need only check that assumptions 1 through 4 hold. We already assumed 4, so we will first check 2. It is straightforward to verify that \( h^{-1} |y-x| \) has a finite third moment for all \( h \). By Chebychev's inequality, this implies that

A2.9) \( \lim_{h \to 0} \sup_{|x| < R} \Delta_{e,h}(x) = 0 \)

It is also easy to check that

A2.10) \( b(S,Y,t) = \begin{bmatrix} f(S,Y,t) \\ F(S,Y,t) \end{bmatrix} \)

A2.11) \( a(S,Y,t) = \begin{bmatrix} g^2 \Omega_{1,1} & gG_{1} \Omega_{1,2} & \ldots & gG_n \Omega_{1,n+1} \\ gG_{1} \Omega_{2,1} & G_1^2 \Omega_{2,2} & \ldots & G_1 G_n \Omega_{2,n+1} \\ \vdots & \vdots & \ddots & \vdots \\ gG_{n} \Omega_{n+1,1} & G_n G_{1} \Omega_{n+1,2} & \ldots & G_n^2 \Omega_{n+1,n+1} \end{bmatrix} \)

where \( \Omega_{i,j} \) is the \( i,j^{th} \) element of \( \Omega \) and \( G_i \) is the \( i^{th} \) element of the vector function \( G \). Since we assumed local boundedness of \( f,g,F \) and \( G \), assumption 3 follows immediately from A2.10 and A2.11, and the representation A1.15 carries out the factorization in assumption 2.

Q.E.D.
Proof of Proposition A1.22:

Let \( \{Z_j\}, j = 1, \infty \) be an i.i.d. \( N(0,1) \) sequence, and note that

\[
A2.12) \quad Q_{k,t} \sim (1-2/\pi)^{-1/2} [kt]^{-1/2} \sum_{j=1}^{[kt]} |Z_j| - E|Z_j|
\]

where "\( \sim \)" denote equality in distribution. Using the fact that \( E|Z_j| = (2/\pi)^{1/2} \), proposition A1.22 follows immediately by Donsker's theorem (Billingsly (1968)).

Q.E.D.

Proof of Theorem A1.3:

Since \( Y \) is bounded a.s. on \([0,T]\), and since both the drift and variance terms in the \( S \) process are locally bounded functions of \( Y \), these terms are bounded a.s. on \([0,T]\), and we would therefore expect \( S \) to remain bounded on \([0,T]\). Indeed

\[
A2.13) \quad \left| \int_0^t f(Y_s,s) \, ds \right| < t \sup_{0 \leq s \leq t} |f(Y_s,s)| < \infty \quad \text{a.s.}
\]

by the a.s. boundedness of \( Y \) and the local boundedness of \( f \). Further, by Friedman (1975) chapter 4 theorem 3.3, we have for all \( \eta > 0 \)
A2.14) \[ P \left( \sup_{0 \leq s \leq t} \left| \int_0^t g(Y_s, s) \, dW_i(s) \right| > \eta \right) \]

\[ \leq \frac{\eta^{-1}}{t} + P \left( \sup_{0 \leq s \leq t} g^2(Y_s, s) > \eta \right) \]

Letting \( \eta \to \infty \), assumption 4 now follows by the a.s. boundedness of \( Y \) and the local boundedness of \( g \).

Q.E.D.
REFERENCES:


Cochrane, John (1986), "How Big is the Random Walk Component in GNP?" mimeo, University of Chicago.


**TABLE I: AR(1) plus stochastic trend**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Robust s.e.</th>
<th>Score s.e.</th>
<th>Hessian s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Trend_0$</td>
<td>-7.89972200</td>
<td>0.50062450</td>
<td>0.71244633</td>
</tr>
<tr>
<td>$S_0$</td>
<td>-1.86882200</td>
<td>0.73878595</td>
<td>2.6887706</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.00119800</td>
<td>0.00029379</td>
<td>0.00030135</td>
</tr>
<tr>
<td>$a_1$</td>
<td>-0.23738500</td>
<td>0.05503028</td>
<td>0.77811443</td>
</tr>
<tr>
<td>$\nu$</td>
<td>1.45290355</td>
<td>0.10120457</td>
<td>0.01849918</td>
</tr>
<tr>
<td>$\Theta_T$</td>
<td>-0.00361500</td>
<td>0.00826585</td>
<td>0.00769985</td>
</tr>
<tr>
<td>$\Psi_T$</td>
<td>0.11026600</td>
<td>0.02120267</td>
<td>0.01229947</td>
</tr>
<tr>
<td>$\Theta_S$</td>
<td>-0.18510200</td>
<td>0.02705013</td>
<td>0.02035595</td>
</tr>
<tr>
<td>$\Psi_S$</td>
<td>0.04458000</td>
<td>0.03679719</td>
<td>0.02964851</td>
</tr>
<tr>
<td>$AR$</td>
<td>0.79747800</td>
<td>0.03617651</td>
<td>0.03383756</td>
</tr>
</tbody>
</table>

Robust form of the covariance matrix for coefficient vector:

0.2506249    -0.2920926   -1.1131662E-005  -0.0056246
-6.913060E-005 5.498894E-005  -0.0028304  0.0023947
  0.0048317       0.0041877

-0.29209259   0.54580468   4.081552E-005   0.0092954
  0.00938162   0.00015873   0.00403846  -0.0038409
-0.00563257   -0.00493564

-1.113166E-005 4.081552E-005  8.631471E-008  3.130571E-006
-1.289235E-005 1.050064E-007  4.999653E-007  -1.572699E-006
-3.468726E-007 -3.960309E-007

-0.00562459   0.00929544   3.130566E-006   0.0030283
-0.00341407   0.00019903   0.00094510  -0.0010728
-0.00148549   -0.00028704

-6.913060E-005 -0.00938162  -1.289235E-005  -0.00341407
  0.01024237  -0.00027382   -0.00065978   0.00125465
  0.00050844   -0.00027382

5.498893E-005  0.00015873   1.050064E-007   0.00019903
-0.00027382   6.832436E-005  5.629284E-005  -5.372180E-005
-8.313781E-006 7.082141E-005

-0.00283038   0.00403846   4.999650E-007   0.00094510
-0.00065978   5.629284E-005   0.00044955  -0.00024456
-0.00048812   -0.00019347

0.00239469  -0.00384091  -1.572699E-006  -0.00107275
  0.00125465  -5.372180E-005  -0.00024456   0.00073171
  0.00049157   0.00049855
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00483167</td>
<td>-0.00563257</td>
<td>-3.468726E-007</td>
<td>-0.00148549</td>
</tr>
<tr>
<td>0.00050844</td>
<td>-8.313781E-006</td>
<td>-0.00048812</td>
<td>0.00049157</td>
</tr>
<tr>
<td>0.00135403</td>
<td>0.00050513</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00418766</td>
<td>-0.00493564</td>
<td>-3.960309E-007</td>
<td>-0.00028704</td>
</tr>
<tr>
<td>-0.00027382</td>
<td>7.082141E-005</td>
<td>-0.00019347</td>
<td>0.00049855</td>
</tr>
<tr>
<td>0.00050513</td>
<td>0.00130874</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>