International Finance Discussion Paper

Number 322

May 1988

THE SIMULTANEOUS EQUATIONS MODEL WITH GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY: THE SEM-GARCH MODEL

Richard Harmon

ABSTRACT

In this paper I generalize the standard simultaneous equations model by allowing the innovations of the structural equations to exhibit Generalized Autoregressive Conditional Heteroskedasticity (**GARCH**). I refer to this new specification as the **SEM-GARCH** model. I develop two estimation strategies: **LIM-GARCH**, a limited information estimator, and **FIM-GARCH**, a full information estimator. I show that these estimators are consistent and asymptotically normal. Following Weiss (1986) I show that when the errors in the **SEM-GARCH** process are incorrectly assumed to be conditionally normal the likelihood function is still maximized at the true parameters, given certain regularity conditions. This results in the asymptotic variance-covariance matrix being more complex than the usual inverse of the information matrix.

THE SIMULTANEOUS EQUATIONS MODEL WITH GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY: THE SEM-GARCH MODEL

Richard Harmon*

In this paper I generalize the standard simultaneous equations model (**SEM**) by allowing the innovations of the structural equations to exhibit Generalized Autoregressive Conditional Heteroskedasticity (**GARCH**). A **GARCH(p,q)** process is a process whose conditional variance at time t is a function of the information available at time t-1. Specifically, it is a function of variances of past innovations and of the squared realizations of past innovations. Unconditionally, the current innovation reverts to the standard specification: white noise with fixed variances over time. I refer to this new specification as the **SEM-GARCH** model. I develop two estimation strategies: **LIM-GARCH**, a limited information estimator, and **FIM-GARCH**, a full information estimator. I show that these estimators are consistent and asymptotically normal.

The outline of the paper is as follows. Section I presents the unrestricted and diagonal **SEM-GARCH(p,q)** models. Sections II and III derive the **LIM-GARCH** estimator and its asymptotic properties. Sections IV and V develop the **FIM-GARCH** estimator and its asymptotic properties.

---

*I am grateful to James Albrecht, Dale W. Henderson, and especially Dan Westbrook for their patience and guidance throughout the course of research embodied in this paper. Helpful suggestions are also appreciated from Frank Diebold, Neil Ericsson, David Howard, Ralph Tryon and the participants of a seminar at the Federal Reserve Board. This work was completed while I was an intern in the International Finance Division of the Federal Reserve Board. This paper represents the views of the author and should not be interpreted as reflecting the view of the Board of Governors of the Federal Reserve System. Unfortunately, any remaining errors are my responsibility.

Finally in Section VI the Information Matrix Test of White (1982) is used to ascertain the correct form of the variance-covariance matrix and test for misspecification. This is followed by some concluding remarks.

I. THE SEM-GARCH(p,q) MODEL

The Autoregressive Conditional Heteroskedastic model (**ARCH**) developed by Engle (1982) and its generalization, the **GARCH** model, developed by Bollerslev (1986) specifies the conditional variance of the current innovation as a function of the available information set; specifically, the conditional variance is a function of the squared realizations of past innovations and of variances of past innovations. In the **GARCH(p,q)** model, the conditional variance, denoted h_t^2, has the following specification

$$(1.1) \qquad h_t^2 = \omega_o + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \delta_i h_{t-i}^2 .$$

Kraft and Engle (1982) extend the **ARCH** model into a multivariate time series framework. Extensions with respect to multivariate **GARCH** models are given by Bollerslev, Engle, and Wooldridge (1985). The **SEM-GARCH** model, to be specified below, generalizes the standard **SEM** by allowing the innovations of the structural equations to exhibit **GARCH** processes. This type of specification has many potential applications. For example, in models of foreign exchange rate determination it can be used to model the joint determination of the foreign exchange rate with domestic and foreign interest rates.

The standard SEM consists of M linear equations:

$$(1.2) \qquad Y_t\Gamma + X_t\beta = \epsilon_t , \qquad\qquad t = 1,\ldots,T ,$$

where Γ is a M x M matrix of coefficients of current endogenous variables, β is a K x M matrix of coefficients of predetermined variables, ϵ_t is a row vector consisting of M innovations at time t, and Y_t and X_t are row vectors consisting of observations on M endogenous variables and K predetermined variables, respectively, at time t. It is assumed that the vector of innovations, ϵ_t, is given by

$$(1.3) \qquad \epsilon_t \sim N(0,\Sigma) ,$$

where Σ is nonsingular. The likelihood function for the sample $Y_1,\ldots,Y_T$ conditional on X is given by

$$(1.4) \qquad L(Y_1,\ldots,Y_T|X) = (2\pi)^{-TM/2}||\Gamma||^T|\Sigma|^{-T/2}$$

$$\exp\left[-\frac{1}{2}\sum_{i=1}^{T}(Y_t\Gamma + X_t\beta)\Sigma^{-1}(Y_t\Gamma + X_t\beta)'\right]$$

Leaving aside the variance-covariance matrix Σ, which has M(M+1)/2 distinct elements, the likelihood function defined by (1.4) has M^2 + MK parameters. Clearly, without a priori restrictions on Γ, β, or Σ, none of the parameters of the structural model are identified.

For now I restrict my attention to a model whose predetermined variables are exogenous; dynamic linear **SEM** remain a topic for future research. Following Kraft and Engle's multivariate **ARCH** specification, the **SEM-GARCH(p,q)** model is given by

$$(1.5) \qquad Y_t\Gamma + X_t\beta = \epsilon_t ,$$

$$(1.6) \qquad (\epsilon_t | I_{t-1}) \sim N(0,H_t) ,$$

where

$$(1.7) \qquad H_t = \begin{bmatrix} H_{11,t} & H_{12,t} & \cdots & H_{1M,t} \\ H_{21,t} & H_{22,t} & \cdots & H_{2M,t} \\ \vdots & & \ddots & \vdots \\ H_{M1,t} & H_{M2,t} & \cdots & H_{MM,t} \end{bmatrix}$$

$$= \mathbf{W} + \left[I_M \otimes \epsilon_{t-1}\right]\mathbf{C}_1\left[I_M \otimes \epsilon_{t-1}\right]' + \cdots + \left[I_M \otimes \epsilon_{t-q}\right]\mathbf{C}_q\left[I_M \otimes \epsilon_{t-q}\right]'$$

$$+ \left[I_M \otimes \mathbf{h}_{t-1}\right]\mathbf{D}_1\left[I_M \otimes \mathbf{h}_{t-1}\right]' + \cdots + \left[I_M \otimes \mathbf{h}_{t-p}\right]\mathbf{D}_p\left[I_M \otimes \mathbf{h}_{t-p}\right]' ,$$

where ϵ_{t-i} and $\mathbf{h}_{t-i}$ are M-element row vectors consisting of ith lagged innovations and ith lagged conditional standard deviations. Y_t and X_t are row vectors consisting of M endogenous variables and K exogenous variables, respectively. $H_{ij,t}$ $(i,j=1,\ldots,M)$ is a scalar and represents the conditional variance of the ith equation when $i=j$ and the conditional covariance between the ith and jth equations when $i\neq j$. The primary restriction imposed on H_t is that it must be bounded and positive-definite. **C** and **D** are $M^2 \times M^2$ symmetric matrices with $M \times M$ symmetric blocks of C_{ij} and D_{ij}, respectively. **W**, a $M \times M$ scalar matrix, nests the hypothesis of homoskedastic ϵ_t within the SEM-GARCH specification and allows the conditional variance-covariance matrix to exist when all **C** and **D** are zero.

While I explicitly assume that the conditional distribution of the innovations is normal, in many circumstances this may not be the case. If the conditional probability model is non-normal, then the estimators proposed here are quasi-maximum likelihood estimators (QMLE). White (1982) notes that the asymptotic variance-covariance matrix of the QMLE no longer equals the inverse of Fisher's Information matrix, but can be consistently estimated by a more complex form. For expository purposes I will continue to assume that the conditional probability model is correctly specified and explore the consequences of it being incorrectly specified when the asymptotic properties of the estimates are derived.

The following assumptions will be made throughout:

(i) $\quad \text{plim}\left[\dfrac{X'X}{T}\right] = Q$, where Q is a finite and nonsingular matrix,

(ii) $\quad \text{plim}\left[\dfrac{X'\epsilon}{T}\right] = 0$,

(iii) $\quad \text{plim}\left[\dfrac{\epsilon'\epsilon}{T}\right] = \Sigma$, where Σ is a bounded and positive-definite unconditional variance-covariance matrix.[1]

A alternative parameterization of the **SEM-GARCH(p,q)** model facilitates estimation and highlights the alternative variance-covariance specifications available. Let $\mathbf{h}_t^*$ be the $M(M+1)/2$ column vector whose components are the unique elements in H_t, taken by vectorizing the lower triangle of H_t. For expository purposes I will focus on the **SEM-GARCH(1,1)** model in a two equation system (M=2). Here the conditional variance-covariance matrix is defined to be

$$(1.8) \qquad H_t = \begin{bmatrix} H_{11,t} & H_{12,t} \\ H_{21,t} & H_{22,t} \end{bmatrix},$$

$$= \mathbf{W} + \left[I_M \otimes \epsilon_{t-1}\right] C_1 \left[I_M \otimes \epsilon_{t-1}\right]'$$

$$+ \left[I_M \otimes \mathbf{h}_{t-1}\right] D_1 \left[I_M \otimes \mathbf{h}_{t-1}\right]' ,$$

where $\mathbf{C}_1$ and $\mathbf{D}_1$ are 4 x 4 parameter matrices given by

[1] An excellent discussion of Central Limit Theory for the non-i.i.d. case is provided in White (1984) Ch.5.

$$(1.9) \qquad C_1 = \left[\begin{array}{cc:cc} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ \hdashline C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{array}\right]$$

$$(1.10) \qquad D_1 = \left[\begin{array}{cc:cc} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ \hdashline D_{31} & D_{32} & D_{33} & D_{34} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{array}\right]$$

Note that the symmetry of C and D require that $C_{21} = C_{12}$, $C_{23} = C_{14}$, $C_{41} = C_{32}$, $C_{43} = C_{34}$, $D_{21} = D_{12}$, $D_{23} = D_{14}$, $D_{41} = D_{32}$ and $D_{43} = D_{34}$. The alternative parameterization consists of vectorizing the lower triangle of H_t, denoted $\mathbf{h}_t^*$,

$$(1.11) \qquad \mathbf{h}_t^* = \begin{bmatrix} H_{11,t} \\ H_{21,t} \\ H_{22,t} \end{bmatrix}$$

$$= \begin{bmatrix} \omega_{11} \\ \omega_{21} \\ \omega_{22} \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}\epsilon_{1,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix}$$

$$+ \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \begin{bmatrix} H_{11,t-1} \\ H_{21,t-1} \\ H_{22,t-1} \end{bmatrix} ,$$

which in matrix notation can be written as

$$(1.12) \qquad \mathbf{h}_t^* = \boldsymbol{\omega} + A\boldsymbol{\eta}_{t-1} + B\mathbf{h}_{t-1}^* .$$

where $\boldsymbol{\eta}_{t-1} = \mathrm{vec}(\epsilon'_{t-1}\epsilon_{t-1})$ and $\epsilon_{t-1} = (\epsilon_{1,t-1}, \epsilon_{2,t-1})$. This setup will be referred to as the "unrestricted" **SEM-GARCH(1,1)** model since no restrictions are imposed on matrices A and B except those required to ensure the positive definiteness of H_t. Expansion of (1.11) highlights why this specification is referred to as "unrestricted",

$$(1.13) \qquad H_{11,t} = \omega_{11} + A_{11}\epsilon_{1,t-1}^2 + A_{12}(\epsilon_{2,t-1}\epsilon_{1,t-1}) + A_{13}\epsilon_{2,t-1}^2 + B_{11}H_{11,t-1} + B_{12}H_{21,t-1} + B_{13}H_{22,t-1} ,$$

$$(1.14) \qquad H_{21,t} = \omega_{21} + A_{21}\epsilon_{1,t-1}^2 + A_{22}(\epsilon_{2,t-1}\epsilon_{1,t-1}) + A_{23}\epsilon_{2,t-1}^2 + B_{21}H_{11,t-1} + B_{22}H_{21,t-1} + B_{23}H_{22,t-1} ,$$

$$(1.15) \quad H_{22,t} = \omega_{22} + A_{31}\epsilon^2_{1,t-1} + A_{32}(\epsilon_{2,t-1}\epsilon_{1,t-1}) + A_{33}\epsilon^2_{2,t-1} + B_{31}H_{11,t-1} + B_{32}H_{21,t-1} + B_{33}H_{22,t-1} .$$

As shown in Table 1, the conditional variance-covariance matrix of a two equation **SEM-GARCH(1,1)** model has 21 free parameters. While there should be no problem estimating the parameters of this small system, the number of parameters rapidly increases as the system gets larger. Table 1 shows that a five equation **SEM-GARCH(1,1)** model has 465 parameters. Many time series data sets do not have enough observations for estimation; clearly a more parsimonious specification is required.

One such specification models the conditional variance of each equation as a function of its own lagged squared innovations and lagged conditional variances. Similarly, the conditional covariances can be modelled as functions of the lagged cross-innovations and lagged conditional covariances. Imposing this structure on the **SEM-GARCH(1,1)** model yields (1.8) with $\mathbf{C}_1$ and $\mathbf{D}_1$ given by

$$(1.16) \quad \mathbf{C}_1 = \left[\begin{array}{cc:cc} C_{11} & 0 & 0 & C_{14} \\ 0 & 0 & C_{23} & 0 \\ \hdashline 0 & C_{32} & 0 & 0 \\ C_{41} & 0 & 0 & C_{44} \end{array}\right]$$

Table 1

THE SEM-GARCH(p,q) MODEL

The Conditional Variance-Covariance Matrix

GARCH Model	Number Of Equations	Number Of Parameters To Be Estimated: Unrestricted*	Diagonal**
(1,1)	2	21	9
	3	78	18
	4	210	30
	5	465	45
(2,2)	2	39	15
	3	150	30
	4	410	50
	5	915	75
(3,3)	2	57	21
	3	222	42
	4	610	70
	5	1365	105

* Unrestricted Var-Cov: # = $[M(M+1)/2]^2(p+q) + [M(M+1)/2]$

** Diagonal Var-Cov: # = $(1+p+q)\sum_{i=1}^{M} i$ where M is the number of equations in the system.

$$D_1 = \left[\begin{array}{cc:cc} D_{11} & 0 & 0 & D_{14} \\ 0 & 0 & D_{23} & 0 \\ \hdashline 0 & D_{32} & 0 & 0 \\ D_{41} & 0 & 0 & D_{44} \end{array}\right] \tag{1.17}$$

where symmetry implies $C_{23} = C_{14}$, $C_{41} = C_{32}$, $D_{23} = D_{14}$, and $D_{41} = D_{32}$. Under the more parsimonious parameterization, h_t^* becomes

$$h_t^* = \begin{bmatrix} H_{11,t} \\ H_{21,t} \\ H_{22,t} \end{bmatrix} \tag{1.18}$$

$$= \begin{bmatrix} \omega_{11} \\ \omega_{21} \\ \omega_{22} \end{bmatrix} + \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}\epsilon_{1,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix}$$

$$+ \begin{bmatrix} B_{11} & 0 & 0 \\ 0 & B_{22} & 0 \\ 0 & 0 & B_{33} \end{bmatrix} \begin{bmatrix} H_{11,t-1} \\ H_{21,t-1} \\ H_{22,t-1} \end{bmatrix} .$$

Expanding (1.18) yields

$$H_{11,t} = \omega_{11} + A_{11}\epsilon_{1,t-1}^2 + B_{11}H_{11,t-1} , \tag{1.19}$$

$$(1.20) \qquad H_{21,t} = \omega_{21} + A_{22}(\epsilon_{2,t-1}\epsilon_{1,t-1}) + B_{22}H_{21,t-1} ,$$

$$(1.21) \qquad H_{22,t} = \omega_{22} + A_{33}\epsilon^2_{2,t-1} + B_{33}H_{22,t-1} .$$

Specification (1.18) is referred to as the "diagonal" **SEM-GARCH(1,1)** model. Table 1 summarizes the parametric requirements of the full and diagonal **SEM-GARCH(1,1)** model. For the **SEM-GARCH(1,1)** model in a two equation system the number of parameters is reduced from 21 to 9. For a five equation system the number of parameters is reduced from 465 to only 45.

As stated earlier, H_t is required to be bounded and positive definite. For the **SEM-GARCH(1,1)** model this requires that

$$(1.22) \qquad H_{11,t} > 0 , \qquad\qquad H_{22,t} > 0 ,$$

and

$$(1.23) \qquad H_{11,t}H_{22,t} - H^2_{21,t} > 0 .$$

Clearly, a sufficient condition for H_t to be positive definite is having every term in **ω**, A, and B be positive definite. A less stringent sufficient condition is established by setting **A=B=0** which from (1.22) and (1.23) implies

$$(1.24) \qquad \omega_{11} > 0 \,, \quad \omega_{22} > 0 \quad \text{and} \quad \omega_{11}\omega_{22} - \omega_{21}^2 > 0 \,.$$

Similarly, setting ω=B=0 yields

$$(1.25) \qquad A_{11} > 0 \,, \quad A_{22} > 0 \quad \text{and} \quad A_{11}A_{22} - A_{21}^2 > 0 \,.$$

Alternatively, setting ω=A=0 yields

$$(1.26) \qquad B_{11} > 0 \,, \quad B_{22} > 0 \quad \text{and} \quad B_{11}B_{22} - B_{21}^2 > 0 \,.$$

For higher order **SEM-GARCH** models the parameter constraints to ensure the positive definiteness of H_t are extremely complicated. In those cases, one might impose penalty functions to ensure that H_t is positive definite.

II. THE LIM-GARCH ESTIMATOR

Since the reduced form innovations involve linear combinations of all the structural innovations, my focus is on structural form estimation to facilitate the conditional heteroskedasticity specification. I develop two estimation strategies for the **SEM-GARCH(p,q)** model. The first is a limited information approach that concentrates upon a single equation of the simultaneous equations system while disregarding the parametric restrictions that bind the system as a whole. The second is a full system estimation that makes efficient use of all available information. The limited information approach is useful when the full model is too complex to be estimated by a full information technique or when one suspects specification errors in equations other than the equation of primary interest.[2]

As described in Section I, the structural model consists of M equations with a single equation, say the first, given by

$$(2.1) \qquad Y_t\Gamma_{\cdot 1} + X_t\beta_{\cdot 1} = \epsilon_{1t}, \qquad t = 1,\ldots,T$$

where $\Gamma_{\cdot 1}$, $\beta_{\cdot 1}$, and ϵ_{1t} are the first columns of Γ, β, and ϵ, respectively. The variables are arranged so that the usual identifying restrictions may be shown by the following partitions:

$$(2.2) \qquad \Gamma_{\cdot 1} = \begin{bmatrix} a_o \\ \cdots \\ 0 \end{bmatrix} \qquad \beta_{\cdot 1} = \begin{bmatrix} b_o \\ \cdots \\ 0 \end{bmatrix}$$

and

$$(2.3) \qquad Y = \left[Y_1 \ \vdots \ Y_1^* \right] \qquad X = \left[X_1 \ \vdots \ X_1^* \right]$$

The number of included and excluded endogenous variables in the first

[2]The LIM-GARCH estimator can be viewed as a particular case of FIM-GARCH where the other M-1 equations are just identified and have non-ARCH innovations.

equation are denoted by m_1 and m_1^* ($= M - m_1$), respectively. The numbers of included and excluded exogenous variables are k_1 and k_1^* ($= K - k_1$). Hence, a_o and b_o are column vectors with m_1 and k_1 elements, respectively. The matrices of endogeneous and exogenous variables are partitioned in (2.3) to correspond to the partitioning of the coefficient vectors in (2.2). The usual order and rank conditions for identification of the first equation are given by

(Order Condition): $\qquad k_1^* \geq m_1 - 1$

(Rank Condition): $\qquad \mathrm{Rank}(\phi_1 \Delta_1) = M - 1$

where ϕ_1 is a $R \times (M+K)$ selection matrix composed of zeros and ones, and $\Delta_1 = (\Gamma'_{.1}, \beta'_{.1}) = (a_o' \vdots 0' \vdots b_o' \vdots 0')'$ is a (M+K)—element column vector composed of all the parameters of the system.

The limited information approach to estimation of the **SEM-GARCH(p,q)** model is referred to as the **LIM-GARCH** estimator. Following the standard derivation of the LIML estimator, [c.f. Koopmans, Rubin, and Leipnik (1950), Dhrymes (1970):328-357 and Schmidt (1976):184-195], I begin with the conditional log-likelihood function of the **LIM-GARCH** estimator for the structural equation defined in (2.1):

$$(2.4) \qquad L(a_o, b_o, w_o, \alpha, \delta | I_{t-1}) = \sum_{t=1}^{T} \ell_t$$

$$(2.5) \qquad \ell_t = C_1 - \frac{1}{2}\log(h_{1t}^2) + \frac{1}{2}\log(a_o'W_{11,t}a_o) - \frac{1}{2}\left[(Y_{1t}a_o + X_{1t}b_o)(Y_{1t}a_o + X_{1t}b_o)h_{1t}^{-2}\right]$$

where

$$(2.6) \qquad W_{11,t} = (Y_{1t}'P_tY_{1t}) ,$$

$$(2.7) \qquad P_t = (1 - X_t(X_t'X_t)^{-1}X_t') ,$$

$$(2.8) \qquad C_1 = -\frac{m_1}{2}\left[\ln(2\pi) + 1\right] + \frac{1}{2}\left[1 - \ln|W_t|\right] ,$$

$$(2.9) \qquad W_t = (Y_t'P_tY_t) ,$$

$$(2.10) \qquad (\epsilon_{1t}|I_{t-1}) \sim N(0, h_{1t}^2) ,$$

$$(2.11) \qquad h_{1t}^2 = \omega_o + \sum_{i=1}^{q}\alpha_i\epsilon_{1,t-i}^2 + \sum_{i=1}^{p}\delta_i h_{1,t-i}^2 ,$$

$$(2.12) \qquad \epsilon_{1t} = Y_{1t}a_o + X_{1t}b_o .$$

Note that I_{t-1} denotes information available at time t-1, including Y_t and X_t. The innovations are assumed to be conditionally normal distributed with the conditional variance, h_{1t}^2, following the **GARCH(p,q)** specification of Bollerslev (1986). P_t is a symmetric and idempotent projection matrix,

W_t is the second moment matrix of residuals of the least squares estimate of the reduced form of the entire system and $W_{11,t}$ is the sub-matrix of W_t pertaining to the included endogenous variables, Y_{1t}.

It should be noted that the standard **LIML** and **2SLS** estimators are consistent due to the conditions (i) – (iii) provided in Section I. The primary benefit of the **LIM-GARCH** estimator, as with the **ARCH** and **GARCH** models, is in terms of efficiency.

Since the conditional log-likelihood function L depends on the parameters a_o, b_o, ω_o, α and δ in a nonlinear fashion, maximization of $L(a_o,b_o,\omega_o,\alpha,\delta|I_{t-1})$ requires an iterative technique. Maximum Likelihood (ML) estimates of the parameters a_o, b_o, ω_o, α and δ are derived from the first order conditions of (2.4). These derivatives, given in the Appendix, have a complex recursive structure making it extremely difficult to derive compact analytic expressions. While analytic expressions are in general the desired path to pursue, they are very inflexible to changes in specification and computationally extremely burdensome. Therefore, I rely on numerical derivatives in the actual estimation procedure.

III. ASYMPTOTIC PROPERTIES OF THE LIM-GARCH ESTIMATOR

This section explores the behavior of the **LIM-GARCH** estimator in large samples.[3] In the previous sections conditional normality was explicity assumed. But now, following White (1982) and Weiss (1986), this assumption is relaxed. As with most **QMLE**, the likelihood function is

[3]This section is based on Weiss (1986) who derives the asymptotic properties of an extended **ARCH** model in the context of a dynamic linear regression model with moving average errors.

derived as though the innovations are, in fact, conditionally normal. In Theorem I it is shown that in the limit the conditional log-likelihood function is maximized at the true parameters even though the assumption of normality may not be valid. Theorems II and III verify the consistency and asymptotic normality properties of the **LIM-GARCH** estimates, respectively.

As described in Section II, the **LIM-GARCH** specification involves maximizing the conditional log-likelihood function given by equation (2.4). Throughout this section let θ to be an s-element column vector ($s = m_1+k_1+p+q+1$) containing all the parameters of the **LIM-GARCH** specification, that is,

$$(3.1) \qquad \theta' = (\mathbf{m}', \mathbf{v}') = (a_o', b_o', \omega_o, \alpha', \delta') .$$

where θ is partitioned such that $\mathbf{m}$ is the parameter vector corresponding to the structural equation under investigation, and $\mathbf{v}$ is the parameter vector of the conditional variance specification, given by equation (2.11). Furthermore, I assume $\theta \in \Xi$, where Ξ is a compact subspace of Euclidean space, and let θ_o represent the true parameter vector.

First, a set of lemmas is required to provide the foundation for the ensuing theorems. Lemmas I and II require that certain matrices of partial derivatives be well defined and positive definite.

LEMMA I:

For all $\theta \in \Xi$, there exists a constant $M_1 < \infty$, not depending on θ such that

$$(3.2) \qquad E\left[\frac{\partial \epsilon_{1t}}{\partial m} \frac{\partial \epsilon_{1t}}{\partial m'}\right] < M_1$$

and

$$(3.3) \qquad \mathrm{Det}\ E\left[\frac{\partial \epsilon_{1t}}{\partial m} \frac{\partial \epsilon_{1t}}{\partial m'}\right] > 0 .$$

Proof: See Appendix.

An equivalent requirement for the conditional variance h_{1t}^2 is given by Lemma II.

LEMMA II:

Assume that the fourth moment of ϵ_{1t} exists and is bounded. Then for all $\theta \in \Xi$, there exists a constant $M_2 < \infty$, not depending on θ, such that

$$(3.4) \qquad E\left[\frac{\partial \epsilon_{1t}}{\partial v} \frac{\partial \epsilon_{1t}}{\partial v'}\right] < M_2$$

and

(3.5) $$\text{Det } E\left[\frac{\partial \epsilon_{1t}}{\partial v} \frac{\partial \epsilon_{1t}}{\partial v'}\right] > 0 .$$

Proof: See Appendix.

Lemmas I and II imply that the negative expected value of the matrix of second derivatives of the conditional log-likelihood function is positive definite:

LEMMA III:

Under the same conditions as Lemma II, there exists a constant $M_3 < \infty$ *not depending on* θ *such that*

(3.6) $$A = -E\left[\frac{\partial^2 L}{\partial \theta \partial \theta'}\right] < M_3$$

and

(3.7) $$\text{Det } A > 0 .$$

Proof: See Appendix.

These results provide the basis for the following theorems.

<u>THEOREM I</u>:

For the LIM-GARCH specification given by equations (2.4) to (2.9) and under the same conditions as **Lemma** II

$$(3.8) \qquad L = \lim_{T\to\infty} L(\theta) \qquad \textit{(exists a.s. for all } \theta \in \Xi\textit{)}$$

and the **lim** L(θ) *is uniquely maximized at* θ_o.

Proof: See Appendix.

Thus, in the limit the conditional log-likelihood function is maximized at the true parameters even though the assumption of normality may not be valid.

<u>THEOREM II</u>: **(Consistency)**

For the LIM-GARCH specification given by equations (2.4) to (2.9), the maximum likelihood estimate $\hat{\theta}$ *is consistent for* θ_o *provided* θ_o *is interior to* Ξ.

Proof: See Appendix.

Note that the condition that θ_o is interior to Ξ ensures that, for T large enough, the first derivatives of L(θ) are "well-behaved" at θ_o.

To consider the asymptotic distribution of the LIM-GARCH estimates, first, define A_o and B_o to be the information matrix in Hessian and outer product form, respectively.

$$(3.9) \qquad A_o = -E\left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right],$$

$$(3.10) \qquad B_o = E\left[T\frac{\partial L}{\partial\theta}\frac{\partial L}{\partial\theta'}\right].$$

As will be shown, both A_o and B_o appear in the variance-covariance matrix of the asymptotic distribution of the **LIM-GARCH** estimates and are therefore required to be invertible.

<u>THEOREM III</u>: **(Asymptotic Normality)**

For the LIM-GARCH specification under the same conditions as Theorem I, with the requirement that $\det(B_o) > 0$, *then*

$$(3.11) \qquad B_o^{-1/2} A_o T^{1/2}(\hat{\theta} - \theta_o) \sim N(0,I)$$

Furthermore, consistent estimates of A_o *and* B_o *are given by*

$$(3.12) \qquad \hat{A} = (2T)^{-1}\sum_{t=1}^{T} h_{1t}^{-4}\frac{\partial h_{1t}^2}{\partial\theta}\frac{\partial h_{1t}^2}{\partial\theta'} + T^{-1}\sum_{t=1}^{T} h_{1t}^{-2}\frac{\partial \epsilon_{1t}^o}{\partial\theta}\frac{\partial \epsilon_{1t}^o}{\partial\theta'}$$

and

$$(3.13) \qquad \hat{B} = T^{-1}\sum_{t=1}^{T}\frac{\partial \ell_t}{\partial\theta}\frac{\partial \ell_t}{\partial\theta'}$$

with all derivatives evaluated at $\theta = \theta_o$

Proof: See Appendix.

Thus the asymptotic variance-covariance matrix of $\hat{\theta}$ has the following general form

$$(3.14) \qquad \text{Asymptotic Var-Cov}(\hat{\theta}) = A_o^{-1} B_o A_o^{-1} ,$$

since the conditional log-likelihood function may not be correctly specified. When the conditional distribution of ϵ_{1t}^{o} is normal then $A_o = B_o$ and the standard form for the asymptotic variance-covariance matrix is appropriate, that is

$$(3.15) \qquad \text{Asymptotic Var-Cov}(\hat{\theta}) = A_o^{-1}$$

To determine which form of the variance-covariance matrix is appropriate I rely on White's (1982) information matrix test which is derived in Section VI.

IV. THE FIM-GARCH ESTIMATOR

An estimator for the complete **SEM-GARCH(p,q)** model requires a more complex specification than the **LIM-GARCH** estimator. The full system estimator will be referred to as the Full Information Model with Generalized Autoregressive Conditional Heteroskedastic error processes, or the **FIM-GARCH** estimator. The derivation of the **FIM-GARCH** estimator will focus on the two equation diagonal **SEM-GARCH(1,1)** model, defined in Section I, since this is the most tractable model and can easily be generalized.

The alternative parameterization of the diagonal **SEM-GARCH(1,1)** model, given by (1.18), defines a vector $\mathbf{h}_t^*$ consisting of all the unique elements of the conditional variance-covariance matrix. In matrix notation, this can be written as

$$\text{(4.1)} \qquad \mathbf{h}_t^* = \omega + A\eta_{t-1} + B\mathbf{h}_{t-1}^* .$$

where

$$\text{(4.2)} \qquad \eta_{t-1} = \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}\epsilon_{1,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} ,$$

$$\text{(4.3)} \qquad \mathbf{h}_{t-1}^* = \begin{bmatrix} H_{11,t-1} \\ H_{21,t-1} \\ H_{22,t-1} \end{bmatrix} .$$

FIM-GARCH estimation of the diagonal **SEM-GARCH(1,1)** model requires maximizing the conditional log-likelihood function

$$L = \sum_{t=1}^{T} \ell_t \tag{4.4}$$

where

$$\ell_t(Y_t|I_{t-1}) = -\frac{M}{2}\log(2\pi) + \log||\Gamma|| - \frac{1}{2}\log|H_t| - \frac{1}{2}\mathrm{Tr}\left[(Y_t\Gamma + X_t\beta)'(Y_t\Gamma + X_t\beta)H_t^{-1}\right] \tag{4.5}$$

where $|H_t|$ denotes the determinate of H_t, and $||\Gamma||$ is the absolute value of the determinate of Γ. For estimation purposes, H_t is constructed from its unique elements, defined by h_t^* in equation (1.18).

Since the conditional log-likelihood function L depends on the parameters Γ, β, ω, A, and B in a highly nonlinear fashion, maximization of L requires iterative techniques. The **FIM-GARCH** estimator is derived from the first order conditions of (4.4).

As with the **LIM-GARCH** estimator, the derivatives of H_t with respect to Γ and β are a function of past derivatives of ϵ_t and H_t. As a result, the analytic derivatives of the **FIM-GARCH** estimator have a complex recursive structure which are difficult to calculate. Therefore, I will rely on numerical derivatives for actual estimation.

The difficulties involved in deriving the analytical derivatives can be highlighted by examining the partial derivative of the conditional log-likelihood function with respect to the unrestricted structural parameters, Γ^{μ}. The superscript μ denotes a selection operator, as defined by Hendry (1976), to choose only the unrestricted elements of a matrix. This is necessary since only the derivatives with respect to the unknown elements are equated to zero.

$$(4.6) \qquad \frac{\partial L}{\partial \Gamma^{\mu}} = T\frac{\partial \log||\Gamma||}{\partial(\Gamma^{\mu})} - \frac{T}{2}\left[\frac{\partial \log|H|}{\partial|H|}\right]\left[\frac{\partial|H|}{\partial H}\right]\left[\frac{\partial H}{\partial \Gamma^{\mu}}\right]$$

$$- \frac{1}{2}\left[\frac{H\frac{\partial(\epsilon'\epsilon)}{\partial \Gamma^{\mu}} - \epsilon'\epsilon\frac{\partial H}{\partial \Gamma^{\mu}}}{H'H}\right] = 0 .$$

This can be simplified to yield

$$(4.7) \qquad \frac{\partial L}{\partial \Gamma^{\mu}} = (\Gamma^{-1})' - \frac{1}{2}(H^{-1})'\left[\frac{\partial H}{\partial \Gamma^{\mu}}\right]\left[T - \frac{(Y\Gamma+X\beta)'(Y\Gamma+X\beta)}{H}\right] - \frac{\epsilon' Y}{H} = 0.$$

But the partial derivative of the conditional variance-covariance matrix, H, with respect to Γ is also a function of derivatives of lagged residual variances and covariances as well as derivatives of lagged conditional variances and covariances. Similar recursive structures arise with respect to the other parameters of the system.

In estimation, as with the **LIM-GARCH** estimator, numerical derivatives are used. As explained in Section III, when the model is correctly specified and the conditional distribution of ϵ_{1t} is correctly assumed to be normal, then the information matrix can be equivalently expressed in either Hessian or outer product form. In this case, the Berndt, Hall, Hall, Hausman (**BHHH**) (1974) algorithm, based on the outer product form of the information matrix, can be used to maximize the conditional log-likelihood functions for the **LIM-GARCH** and **FIM-GARCH** estimators. The **BHHH** method is an iterative method for calculating the optimal parameters, θ. Let θ^{i} denote the parameter estimates after the ith iteration. θ^{i+1} is then calculated from,

$$(4.8) \qquad \theta^{i+1} = \theta^{i} + \lambda_{i}\left[\sum_{t=1}^{T} \frac{\partial \ell_t'}{\partial \theta^{i}} \frac{\partial \ell_t}{\partial \theta^{i}}\right]^{-1} \sum_{t=1}^{T} \frac{\partial \ell_t'}{\partial \theta^{i}} ,$$

where λ_i is a variable step length chosen to maximize the likelihood function in the given direction.

An alternative algorithm is the Newton-Raphson method which is based on the Hessian form of the information matrix. As mentioned earlier, both methods are equivalent when the conditional log-likelihood function is correctly specified. The asymptotic variance-covariance matrix is given by (3.15) where $A_o = B_o$. Alternatively, if the conditional distribution is incorrectly specified to be normal, then under certain regularity conditions the asymptotic variance-covariance matrix has the form given in (3.14).

V. <u>ASYMPTOTIC PROPERTIES OF THE FIM-GARCH ESTIMATOR</u>

The derivation of the asymptotic properties of the **FIM-GARCH** estimator closely follows that of the **LIM-GARCH** estimator. As with the LIM-GARCH estimator, if the conditional distribution is non-normal then the **FIM-GARCH** estimator is a **QMLE**. For the **FIM-GARCH** estimator of the diagonal **SEM-GARCH(p,q)** model let θ now be a S—element column vector

$$(5.1) \qquad \theta' = (\mathbf{m}', \mathbf{v}') \qquad \text{with } S = M + K + (1+p+q)\sum_{i=1}^{M} i \ .$$

where θ, as in Section III, is partitioned such that $\mathbf{m}$ is the parameter vector corresponding to the unrestricted parameters of all the structural equations of the system, that is

$$(5.2) \qquad \mathbf{m} = \left[\mathbf{vec}(\Gamma, \beta)\right]^{\mu} ,$$

where **vec** denotes that the matrices are vectorized by column stacking operations and μ denotes a selection operator that chooses only the a priori unrestricted structural parameters. Similarly, $\mathbf{v}$ is the parameter vector of the conditional variance specification, given by equation (4.5),

$$(5.3) \qquad \mathbf{v} = \left[\mathrm{vec}(\omega, A, B)\right]^{\mu} .$$

As before, I assume $\theta \in \Xi$, where Ξ is a compact subspace of Euclidean space, and let θ_o represent the true parameter vector. Theorems IV and V are the **FIM-GARCH** equivelent to Theorems II and III which verify the consistency and asymptotic normality properties of the estimator.

__THEOREM IV__: **(Consistency)**

For the FIM-GARCH specification given by equations (4.1) to (4.5), the maximum likelihood estimate $\hat{\theta}$ is consistent for θ_o provided θ_o is interior to Ξ.

Proof: See Appendix.

<u>THEOREM V</u>: **(Asymptotic Normality)**

Assuming that $det(B_o)>0$*, the FIM-GARCH specification is distributed asymptotically normal,*

$$(5.4) \qquad B_o^{-1/2} A_o T^{1/2}(\hat{\theta} - \theta_o) \sim N(0,I)$$

where consistent estimates of A_o *and* B_o *are given by*

$$(5.5) \qquad \hat{A} = (2T)^{-1} \sum_{t=1}^{T} (H^{-1})'(H^{-1}) \frac{\partial H_t}{\partial \theta} \frac{\partial H_t}{\partial \theta'} + T^{-1} \sum_{t=1}^{T} (H^{-1})'(H^{-1}) \frac{\partial \epsilon_t}{\partial \theta} \frac{\partial \epsilon_t}{\partial \theta'}$$

and

$$(5.6) \qquad \hat{B} = T^{-1} \sum_{t=1}^{T} \frac{\partial \ell_t}{\partial \theta} \frac{\partial \ell_t}{\partial \theta'}$$

with all derivatives evaluated at $\theta = \theta_o$

Proof: See Appendix.

Thus the asymptotic variance-covariance matrix of $\hat{\theta}$ has the following general form

$$(5.7) \qquad \text{Asymptotic Var-Cov}(\hat{\theta}) = A_o^{-1} B_o A_o^{-1} ,$$

since the conditional log-likelihood function may not be correctly specified. When the conditional distribution of ϵ_t is truly normal then $A_o = B_o$ and the standard form for the asymptotic variance-covariance matrix is appropriate, that is

$$(5.8) \qquad \text{Asymptotic Var-Cov}(\hat{\theta}) = A_o^{-1}$$

As with the **LIM-GARCH** estimator, White's Information Matrix Test can be used to determine the correct form for the asymptotic variance-covariance matrix.

VI. <u>THE INFORMATION MATRIX TEST</u>

A well known test for misspecification associated with maximum likelihood estimation is the information matrix test of White (1982). The test is based on his information matrix equivalence theorem. This theorem essentially says that when the model is correctly specified and the conditional distribution of ϵ_{1t}^o is correctly assumed to be normal, then the information matrix can be expressed either in Hessian form, $A(\theta_o)$, or in outer product form, $B(\theta_o)$. White shows that under these conditions

$$(6.1) \qquad A(\theta_o) - B(\theta_o) = 0 .$$

Thus a test for the null hypothesis of conditional normality and correct model specification is based on the difference between consistent estimates of A_o and B_o, denoted $\hat{A}_T$ and $\hat{B}_T$, respectively. For the **LIM-GARCH** estimator θ is a s-element parameter vector, following White (1982) and Weiss (1986), I define q [= s(s+1)/2] vectors $d_t(\theta)$ such that $d_t(\theta)$ has kth element

$$(6.2) \qquad d_{tk}(\theta) = \left[\frac{\partial \ell_t}{\partial \theta_i}\frac{\partial \ell_t}{\partial \theta_j} + \frac{\partial^2 \ell_t}{\partial \theta_i \partial \theta_j}\right] \qquad \begin{array}{l} k = 1,\dots,q \\ i,j = 1,\dots,s \\ i \le j \end{array}$$

The test is based on what White refers to as the "indicators"

$$(6.3) \qquad D_T(\theta) = T^{-1} \sum_{t=1}^{T} d_t(\theta) \, ,$$

which are the elements of $\hat{A}_T - \hat{B}_T$. Next, define $V(\theta)$ as

$$(6.4) \qquad V(\theta) = E\left[d_t(\theta)\, d_t(\theta)'\right] .$$

$V(\theta)$ turns out to be the asymptotic variance-covariance matrix of $T^{1/2} D_T(\hat{\theta})$. Two additional assumptions are required to meet the preconditions for the test to be valid. First, $V(\theta)$ must be nonsingular. Secondly, for cases when ϵ^o_{1t} is not normally distributed it is necessary to assume that

$$(6.5) \qquad E\left[\frac{\epsilon_{1t}^{8}}{(h_{1t}^{2})^{4}}\middle| I_{t-1}\right] < \infty \qquad \text{for all } \theta \in \Xi.$$

Based on these conditions, White proves that

$$(6.6) \qquad T^{1/2} D_T(\hat{\theta}) \overset{A}{\sim} N[0, V(\theta)]$$

and

$$(6.7) \qquad V_T(\hat{\theta}) \xrightarrow{a.s.} V(\theta_o) ,$$

where $V_T(\hat{\theta})$ is the estimate of $V(\theta_o)$. Then it follows that the information test statistic, T_T, is given by

$$(6.8) \qquad T_T = T D_T(\hat{\theta}) \left[V_T(\hat{\theta})\right]^{-1} D_T(\hat{\theta}) \overset{A}{\sim} \chi_q^2 .$$

To carry out the Information Matrix Test, one computes T_T and compares it to the critical value of the χ_q^2 distribution for a given size of test. If T_T does not exceed this value, then one can not reject the null hypothesis that the model is correctly specified and A_o^{-1} may be used as the variance-covariance matrix of the **LIM-GARCH** estimates. This applies equally well to the **FIM-GARCH** estimator

VII. CONCLUSION:

In this paper I have extended the **GARCH(p,q)** model of Bollerslev (1986) into a simultaneous equations framework and derived two estimation strategies: **LIM-GARCH** and **FIM-GARCH**. Furthermore, it has been shown that these two estimation strategies have the desireable asymptotic properties, consistency and asymptotic normality, of Maximum Likelihood estimators. Alternative specifications of conditional variances and covariances that are heteroskedastic remain to be explored. My approach has exclusively focused on the **ARCH** specification originally developed by Engle (1982) and generalized by Bollerslev (1986). Future research will focus on extending the **SEM-GARCH(p,q)** model to a dynamic framework by including lagged endogenous variables. In Harmon (1988) I extend the **SEM-GARCH** model by incorporating the conditional variances and covariances as variables in the structural equations themselves. This is referred to as the **SEM-GARCH-M** model. It is a logical extension of the **ARCH(q)-in-Mean** model (c.f., Engle, Lilien, and Robins (1987)) and the **GARCH(p,q)-in-Mean** model (c.f., Bollerslev, Engle, and Wooldridge (1985)).

BIBLIOGRAPHY

Anderson, G.J. "The Structure Of Simultaneous Equations Estimators: A Comment." Journal of Econometrics. Vol. 14 (1980): 271-276.

Basawa, I.V., P.D. Feigin, and C.C. Heyde. "Asymptotic Properties Of Maximum Likelihood Estimators For Stochastic Processes." Sankhya: The Indian Journal of Statistics, Vol. 38, Series A, Part 3 (1976): 259-270.

Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman. "Estimation and Inference in Nonlinear Strutural Models." Annals of Economic and Social Measurement, Vol. 4 (1974): 653-665.

Bollerslev, Tim. "Generalized Autoregressive Conditional Heteroskedasticity." Journal of Econometrics, Vol. 31 (1986): 307-327.

__________. "A Conditionally Heteroskedastic Time Series Model For Security Price and Rates of Return Data." Discussion Paper #85-32, Department of Economics, University of California, San Diego. (1985).

__________, R. Engle, and J. Wooldridge. "A Capital Asset Pricing Model With Time Varying Covariances." Discussion Paper #82-4, Department of Economics, University of California, San Diego. (1982).

Crowder, Martin. "Maximum Likelihood Estimation For Dependent Observations." Journal of the Royal Statistical Society, Series B38, (1976): 45-53.

Dhrymes, P. J. Econometrics: Statistical Foundations and Applications. New York: Springer-Verlag. (1974).

Diebold, Francis X. and M. Nerlove. "ARCH Models of Exchange Rate Fluctuations." Unpublished manuscript, Department of Economics, University of Pennsylvania, (1986).

________. "Modelling The Persistence Of Conditional Variance: A Comment." Econometric Reviews. Vol.5, No.1 (1986): 51-56.

__________. "Temporal Aggregation of ARCH Processes and Distribution of Asset Returns.," Special Studies Paper #200, Board of Governors of the Federal Reserve System, (1986).

__________. "Testing For Serial Correlation In The Presence Of ARCH." Proceedings of the American Statistical Association, Business and Economic Statistics Section. Washington, D.C.: The American Statistical Society, (1986).

__________ and P. Pauly. "Endogenous Risk In A Portfolio-Balance Rational-Expectations Model Of The Deutschemark-Dollar Rate." Forthcoming, European Economic Review (1988).

Domowitz, Ian and Craig Hakkio. "Conditional Variance And The Risk Premium In The Foreign Exchange Market." Journal of International Economics, Vol. 19, (1985): 47-66.

Domowitz, Ian and Craig Hakkio. "Testing For Serial Correlation In The Preseence Of Heteroscedasticity With Applications To Exchange Rate Models." Research Paper #83-11, Federal Reserve Bank of Kansas City. (October, 1983).

Engel, Robert F. "Autoregressive Conditional Heteroscedasticity With Estimates Of The Variance Of United Kingdom Inflation." Econometrica, Vol. 50, No. 4 (July, 1982): 987-1007.

_________. "Wald, Likelihood Ratio, And Lagrange Multiplier Tests In Econometrics." Handbook of Econometrics, Volume II. Edited by Z. Griliches and M.D. Intriligator. (1984): 776-826.

_________. "A General Approach To Lagrange Multiplier Model Diagnostics." Journal of Econometrics, Vol. 20 (1982): 83-104.

_________ and T. Bollerslev. "Modelling The Persistence Of Conditional Variances." Econometric Reviews, Vol.5, No.1 (1986): 1-50.

_________,C.W.J. Granger and D. Kraft. "Combining Competing Forecasts Of Inflation Using A Bivariate ARCH Model." Journal of Economic Dynamics and Control, Vol. 6 (1984): 151-165.

_________ and D. Kraft. "Multiperiod Forecast Error Variances Of Inflation Estimated From ARCH Models." In Applied Time Series Analysis of Economic Data, edited by A. Zellner, Bureau of the Census, (1983): 293-202.

_________, D. Lilien, and R. Robins. "Uncertainty And The Term Structure." Discussion Paper #82-4, Department of Economics, University of California, San Diego. (1982).

_________, D. Lilien, and R. Robins. "Estimating Time Varying Risk Premia In The Term Structure: The ARCH-M Model." Discussion Paper #85-17, Department of Economics, University of California, San Diego. (1984).

_________, D. Lilien, and R. Robins. "Estimating Time Varying Risk Premia In The Term Structure: The ARCH-M Model." Econometrica. Vol. 55, No. 2 (1987): 391-407.

Harmon, Richard. "The Simultaneous Equations Model With Generalized Autoregressive Conditional Heteroskedasticity In The Mean: The SEM-GARCH-M Model. Unpublished Manuscript, Georgetown University. (1988).

Hendry, David F. "The Structure Of Simultaneous Equations Estimators." Journal of Econometrics. Vol. 4 (1976): 51-88.

Koopmans, T.C., H. Rubin and R.B. Leipnik. "Measuring the Equation System of

Dynamic Economics." In Ch.2 of Statistical Inference in Dynamic Economic Models, Edited by T.C. Koopmans. New York: Wiley (1950).

Kraft, Dennis and R. Engle. "Autoregressive Conditional Heteroskedasticity In Multiple Time Series Models." Discussion Paper #82-23, Department of Economics, University of California, San Diego. (1982).

Pollock, D.S.G. The Algebra of Econometrics. New York: John Wiley & Sons. (1979).

_________. "Varieties Of The LIML Estimator." Australian Economic Papers. (December, 1983): 499-506.

Schmidt, P. Econometrics. New York: Marcel Dekker. (1976).

Spanos, Aris. Statistical Foundations Of Econometric Modelling. Cambridge: Cambridge University Press. 1986.

Weiss, Andrew A. "On The Stability Of A Heteroskedastic Process." Journal Of Time Series Analysis. Vol. 7, No. 4, (1985): 303-310.

_________. "ARMA Models With ARCH Errors." Journal Of Time Series Analysis. Vol. 5, (1984): 129-143.

_________. "ARCH And Bilinear Time-Series Models: Comparison And Combination." Journal of Business and Economic Statistics, Vol. 4 (1986): 59-70.

_________. "Asymptotic Theory For ARCH Models: Estimation And Testing." Econometric Theory, Vol. 2, No. 1 (1986): 107-128.

White, Halbert. Asymptotic Theory for Econometricians. Academic Press, Inc. (1984).

_________. "Maximum Likelihood Estimation Of Misspecified Models." Econometrica, Vol. 50, No. 1 (January, 1982): 1-25.

_________. "Corrigendum." Econometrica, Vol. 51, No. 2 (March, 1983): 513.

APPENDIX

Note: All expectations are conditional on I_{t-1} unless otherwise stated.

First Order Conditions For The LIM-GARCH Estimator:

Differentiating the conditional log-likelihood function, given by equation (2.4), with respect to a_o yields

$$\text{(A.1)} \quad \left.\frac{\partial L}{\partial a_o}\right|_{a_o=\hat{a}_o} = \sum_{t=1}^{T}\left[-\left[\frac{a_o'W_{11,t}}{a_o'W_{11,t}a_o}\right] - \frac{1}{2h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial a_o}\right] - \frac{1}{2}\left[\frac{h_{1t}^2\frac{\partial(\epsilon_{1t}'\epsilon_{1t})}{\partial a_o} - \epsilon_{1t}'\epsilon_{1t}\left[\frac{\partial h_{1t}^2}{\partial a_o}\right]}{\left[h_{1t}^2\right]^2}\right]\right] = 0.$$

When simplified this can be written as

$$\text{(A.2)} \quad \left.\frac{\partial L}{\partial a_o}\right|_{a_o=\hat{a}_o} = \sum_{t=1}^{T}\left[-\left[\frac{a_o'W_{11,t}}{a_o'W_{11,t}a_o}\right] - \frac{\epsilon_{1t}'Y_{1t}}{2h_{1t}^2} + \frac{1}{2h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial a_o}\right]\left[\frac{\epsilon_{1t}'\epsilon_{1t}}{h_{1t}^2} - 1\right]\right] = 0,$$

where

$$(A.3)\qquad \frac{\partial h_{1t}^2}{\partial a_o} = \sum_{i=1}^{q} \alpha_i \frac{\partial \epsilon_{1,t-i}^2}{\partial a_o} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{1,t-i}^2}{\partial a_o}$$

$$= 2 \sum_{i=1}^{q} \alpha_i \epsilon_{1,t-i}' Y_{1,t-i} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{1,t-i}^2}{\partial a_o} .$$

Similarly, differentiating the conditional log-likelihood function with respect to b_o yields

$$(A.4)\qquad \left.\frac{\partial L}{\partial b_o}\right|_{b_o=\hat{b}_o} = \sum_{t=1}^{T} \left[- \frac{\epsilon_{1t}' X_{1t}}{2h_{1t}^2} + \frac{1}{2h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial b_o}\right]\left[\frac{\epsilon_{1t}'\epsilon_{1t}}{h_{1t}^2} - 1\right]\right] = 0$$

where

$$(A.5)\qquad \frac{\partial h_{1t}^2}{\partial b_o} = \sum_{i=1}^{q} \alpha_i \frac{\partial \epsilon_{1,t-i}^2}{\partial b_o} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{1,t-i}^2}{\partial b_o}$$

$$= 2 \sum_{i=1}^{q} \alpha_i \epsilon_{1,t-i}' X_{1,t-i} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{1,t-i}^2}{\partial b_o} .$$

Now, differentiating with respect to the parameters ω_o, α_j, and δ_j that comprise the conditional variance, h_{1t}^2, yields

$$(A.6)\qquad \left.\frac{\partial \ell}{\partial \omega_o}\right|_{\omega_o=\hat{\omega}_o} = \sum_{t=1}^{T} \left[- \frac{1}{2h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial \omega_o}\right]\left[\frac{\epsilon_{1t}'\epsilon_{1t}}{h_{1t}^2} - 1\right]\right] = 0 .$$

but notice that

$$(A.7) \qquad \frac{\partial h_{1t}^2}{\partial \omega_o} = \left[1 + \sum_{i=1}^{p} \delta_i \frac{\partial h_{1,t-i}^2}{\partial \omega_o} \right]$$

It is clear from (A.7) that a complex recursive structure exists, which requires expansions in order to derive an analytic expression for (A.6).

Differentiating with respect to α_j yields

$$(A.8) \qquad \frac{\partial L}{\partial \alpha_j} = \sum_{t=1}^{T} \left[\frac{1}{2h_{1t}^2} \left[\frac{\partial h_{1t}^2}{\partial \alpha_j} \right] \left[\frac{\epsilon_{1t}' \epsilon_{1t}}{h_{1t}^2} - 1 \right] \right] = 0 ,$$

where

$$(A.9) \qquad \frac{\partial h_{1t}^2}{\partial \alpha_j} = \left[\epsilon_{1,t-j}^2 + \sum_{i=1}^{p} \delta_i \frac{\partial h_{1,t-j}^2}{\partial \alpha_j} \right] .$$

Then, differentiating with respect to δ_j yields the following first order condition

$$(A.10) \qquad \frac{\partial L}{\partial \delta_j} = \sum_{t=1}^{T} \left[\frac{1}{2h_{1t}^2} \left[\frac{\partial h_{1t}^2}{\partial \delta_j} \right] \left[\frac{\epsilon_{1t}' \epsilon_{1t}}{h_{1t}^2} - 1 \right] \right] = 0 ,$$

where in this case

$$\text{(A.11)}\qquad \frac{\partial h_{1t}^2}{\partial \delta_j} = \left[h_{1,t-j}^2 + \sum_{i=1}^{q} \alpha_i \frac{\partial \epsilon_{1,t-i}^2}{\partial \delta_j} \right] .$$

The expressions for the first order conditions derived in equations (A.6) through (A.11) clearly show their complex recursive nature. The remaining first order conditions require differentiating L with respect to the conditional variance, h_{1t}^2.

$$\text{(A.12)}\qquad \frac{\partial L}{\partial h_{1t}^2} = \sum_{t=1}^{T} \left[\frac{1}{2h_{1t}^2} \left[\frac{\epsilon_{1t}'\epsilon_{1t}}{h_{1t}^2} - 1 \right] \right] = 0 .$$

It is not possible to derive general compact analytic expressions due to the recursive structure of these first order conditions .

<u>Proof of LEMMA I</u>:

Expressions for $\partial\epsilon_{1t}/\partial m$, where $m' = (a_o', b_o')$, are given by

$$\text{(A.13)}\qquad \frac{\partial \epsilon_{1t}}{\partial a_o} = Y_{1t} ,$$

$$\text{(A.14)}\qquad \frac{\partial \epsilon_{1t}}{\partial b_o} = X_{1t} .$$

Then for bounded constant vectors $\lambda < \infty$, one has

$$\text{(A.15)} \qquad \lambda' \frac{\partial \epsilon_{1t}}{\partial m} = \lambda' Z_{1t} ,$$

where $Z'_{1t} = [Y_{1t}, X_{1t}]$. Now, one can write

$$\text{(A.16)} \qquad E\left[\lambda' \frac{\partial \epsilon_{1t}}{\partial m} \frac{\partial \epsilon_{1t}}{\partial m'} \lambda\right] = E\left[\lambda' Z_{1t} Z'_{1t} \lambda\right]$$

where $Z_{1t}Z'_{1t}$ is the cross product matrix and the right hand side is thus a scalar. Hence there exists a constant $M_1 < \infty$ such that

$$\text{(A.17)} \qquad E\left[\lambda' \frac{\partial \epsilon_{1t}}{\partial m} \frac{\partial \epsilon_{1t}}{\partial m'} \lambda\right] < M_1 \qquad \text{for all } \theta \in \Xi .$$

Next, to show that

$$\text{(A.18)} \qquad \det E\left[\frac{\partial \epsilon_{1t}}{\partial m} \frac{\partial \epsilon_{1t}}{\partial m'}\right] > 0$$

is equivalent to showing that

$$\text{(A.19)} \qquad E\left[\lambda' \frac{\partial \epsilon_{1t}}{\partial m} \frac{\partial \epsilon_{1t}}{\partial m'} \lambda\right] > 0 \qquad \text{for all } \lambda \neq 0.$$

Following Weiss (1986), proof is by contradiction. First, assume there exists $\lambda \neq 0$ such that (A.19) equals zero. Then it must true that

$$\text{(A.20)} \qquad \lambda' \frac{\partial \epsilon_{1t}}{\partial m} = 0 \qquad \text{a.s. for all t.}$$

This implies from (A.15) that $\lambda' Z_{1t} = 0$ a.s. for all t. Now let $\hat{Y}_{1t}$ represent the forecast of Y_{1t} given information available at time t-1, that is

$$\text{(A.21)} \qquad \hat{Y}_{1t} = E\left[Y_{1t} \middle| I_{t-1}\right] .$$

Then specify Y_{1t} to be

$$\text{(A.22)} \qquad Y_{1t} = \epsilon^{o}_{1t} + \hat{Y}_{1t} ,$$

where ϵ^{o}_{1t} denotes the true errors. Furthermore, using a mean value expansion one can specify ϵ_{1t} to be

$$\text{(A.23)} \qquad \epsilon_{1t} = \epsilon^{o}_{1t} + \frac{\partial \epsilon_{1t}}{\partial m'}(m - m_o)$$

where the derivative is evaluated at m^*, which lies between m and m_o. Equations (A.22) and (A.23) imply that ϵ^{o}_{1t} is a function of I_{t-1} and Z_{1t}, since $\partial \epsilon_{1t}/\partial m' = Z_{1t}$. If that is the case then

$$\text{(A.24)} \qquad E\left[\epsilon^{o}_{1t} \middle| I_{t-1}, Z_{1t}\right] = \epsilon^{o}_{1t} .$$

But the **LIM-GARCH** specification assumes that

$$\text{(A.25)} \qquad E\left[\epsilon_{1t} \middle| I_{t-1}\right] = 0$$

and therefore it must be the case that

(A.26) $$E\left[\epsilon^{o}_{1t} \middle| I_{t-1}, Z_{1t}\right] = 0 .$$

This implies that $\epsilon^{o}_{1t} = 0$ a.s. for all t, which contradicts the fact that $E\left[(\epsilon^{o}_{1t})^2\right] = \sigma^2_{\epsilon} > 0$ (i.e., the unconditional variance is greater than zero). Therefore, no such λ exists and

(A.27) $$E\left[\lambda' \frac{\partial \epsilon_{1t}}{\partial m} \frac{\partial \epsilon_{1t}}{\partial m'} \lambda\right] > 0 \qquad \text{a.s.}$$

as required. □

<u>Proof of LEMMA II</u>:

Expressions for $\partial h^2_{1t}/\partial v$, where $v' = (\omega_o, \alpha', \delta')$, are given by

(A.28) $$\frac{\partial h^2_{1t}}{\partial \omega_o} = 1 ,$$

(A.29) $$\frac{\partial h^2_{1t}}{\partial \alpha_i} = \epsilon^2_{1,t-i} ,$$

(A.30) $$\frac{\partial h^2_{1t}}{\partial \delta_i} = \frac{\partial h^2_{1,t-i}}{\partial \delta_i} .$$

For constant vectors $\lambda < \infty$, write (A.28), (A.29), and (A.30) as

$$\text{(A.31)} \qquad \lambda' \frac{\partial h_{1t}^2}{\partial v} = \lambda' W_t$$

where

$$\text{(A.32)} \qquad W_t' = \left[1 \; , \; \epsilon_{1,t-1}^2 \; , \ldots , \; \epsilon_{1,t-q}^2 \; , \; \frac{\partial h_{1,1-1}^2}{\partial \delta_1} \; , \ldots , \; \frac{\partial h_{1,t-p}^2}{\partial \delta_p} \right]$$

Since by assumption $E\left[(\epsilon_{1t}^{o})^4\right] < \infty$ then $E\left[\epsilon_{1t}^4\right] < M_2 < \infty$, for all $\theta \in \Xi$. Thus, the first part of the lemma is straightforward.

For the second part of Lemma II, apply the same method of proof as in Lemma I. As before, to show that

$$\text{(A.33)} \qquad \det E\left[\frac{\partial h_{1t}^2}{\partial v} \frac{\partial h_{1t}^2}{\partial v'}\right] > 0$$

is equivalent to showing that

$$\text{(A.34)} \qquad E\left[\lambda' \frac{\partial h_{1t}^2}{\partial v} \frac{\partial h_{1t}^2}{\partial v'} \lambda\right] > 0 \qquad\qquad \text{for all } \lambda \neq 0.$$

Assume there exists a $\lambda \neq 0$ such that (A.34) equals zero. Then, as in Lemma I, this implies that $\lambda' W_t = 0$ a.s. for all t. Using a similar expression to (A.23), which is derived from the mean value theorem, yields an expression for ϵ_{1t}^2

$$\text{(A.35)} \quad \epsilon_{1t}^2 = (\epsilon_{1t}^{o})^2 + 2\epsilon_{1t}^{o} \frac{\partial \epsilon_{1t}}{\partial v}(v - v_o) + (v - v_o)' \left[\frac{\partial \epsilon_{1t}}{\partial v} \frac{\partial \epsilon_{1t}}{\partial v'}\right](v - v_o)$$

This is a quadratic function in ϵ_{1t}^{o} which yields two solutions,

$$\text{(A.36)} \qquad \epsilon_{1t}^{o} = f_1(t) \quad \text{or} \quad f_2(t) \ .$$

These solutions are functions of I_{t-1} and X_{1t}. But ϵ_{1t}^{o} having two values ocnditional on I_{t-1} and X_{1t} is not permitted. Therefore, no such λ exists and (A.34) holds. □

<u>Proof of LEMMA III</u>:

Differentiating the conditional log-likelihood function with respect to θ yields

$$\text{(A.37)} \qquad \left.\frac{\partial L}{\partial \theta_i}\right|_{\theta_i=\theta_o} = \sum_{t=1}^{T}\left[-\frac{\epsilon_{1t}}{h_{1t}^2}\left[\frac{\partial \epsilon_{1t}}{\partial \theta_i}\right] + \frac{1}{2h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial \theta_i}\right]\left[\frac{\epsilon_{1t}^2}{h_{1t}^2} - 1\right]\right] = 0$$

where the derivatives with respect to θ_i are evaluated at their true values θ_o. Second order derivatives can be shown to have the following general form

$$\text{(A.38)} \qquad \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = \sum_{t=1}^{T}\left[-\frac{\epsilon_{1t}}{h_{1t}^2}\left[\frac{\partial^2 \epsilon_{1t}}{\partial \theta_i \partial \theta_j}\right] - \frac{1}{h_{1t}^2}\left[\frac{\partial \epsilon_{1t}}{\partial \theta_i}\right]\left[\frac{\partial \epsilon_{1t}}{\partial \theta_j}\right]\right.$$

$$+ \frac{\epsilon_{1t}}{h_{1t}^4}\left[\frac{\partial \epsilon_{1t}}{\partial \theta_i}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta_j}\right] + \frac{1}{2h_{1t}^4}\left[\frac{\epsilon_{1t}^2}{h_{1t}^2} - 1\right]\left[\frac{\partial^2 h_{1t}^2}{\partial \theta_i \partial \theta_j}\right]$$

$$+ \frac{\epsilon_{1t}}{h_{1t}^4}\left[\frac{\partial h_{1t}^2}{\partial \theta_i}\right]\left[\frac{\partial \epsilon_{1t}}{\partial \theta_j}\right] - \frac{1}{h_{1t}^4}\left[\frac{\epsilon_{1t}^2}{h_{1t}^2}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta_i}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta_j}\right]$$

$$\left. + \frac{1}{2h_{1t}^4}\left[\frac{\partial h_{1t}^2}{\partial \theta_i}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta_j}\right]\right]$$

Expressions for $\partial \epsilon_{1t}/\partial m_i$ and $\partial h_{1t}^2/\partial v_i$ are given by (A.15) and (A.31), respectively. For $\partial h_{1t}^2/\partial m_j$ one has

$$\text{(A.39)} \qquad \frac{\partial h_{1t}^2}{\partial m_j} = 2 \sum_{i=1}^{q} \alpha_i \epsilon_{1,t-i} \frac{\partial \epsilon_{1,t-i}}{\partial m_j} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{1,t-i}^2}{\partial m_j}$$

where m_j is the jth element of m. From Lemma II and (A.39) it is easy to see that $\partial h_{1t}^2/\partial \theta_i$ have bounded second moments.

Since every term in $\partial h_{1t}^2/\partial \theta_i$ and $\partial^2 h_{1t}^2/\partial \theta_i \partial \theta_j$ also appears in h_{1t}^2 itself, the expressions

$$\text{(A.40)} \qquad \frac{1}{h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial \theta_i}\right] \quad \text{and} \quad \frac{1}{h_{1t}^2}\left[\frac{\partial^2 h_{1t}^2}{\partial \theta_i \partial \theta_j}\right]$$

are uniformily bounded from above. Hence, evaluating equation (A.29) at the true parameter values θ_o implies that the first, third, fourth, and fifth terms are zero, with the sixth term being

$$\text{(A.41)}\quad E\left[\frac{1}{h_{1t}^4}\left[\frac{\epsilon_{1t}^2}{h_{1t}^2}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta_i}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta_j}\right]\middle| I_{t-1}\right] = E\left[\frac{1}{h_{1t}^4}\left[\frac{\partial h_{1t}^2}{\partial \theta_i}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta_j}\right]\middle| I_{t-1}\right] < \infty$$

since $E(\epsilon_{1t}^2|I_{t-1}) = h_{1t}^2$. Similarly, the second term of (A.38), evaluated at θ_o, is bounded

$$\text{(A.42)}\quad E\left[\frac{1}{h_{1t}^2}\left[\frac{\partial \epsilon_{1t}}{\partial \theta_i}\right]\left[\frac{\partial \epsilon_{1t}}{\partial \theta_j}\right]\middle| I_{t-1}\right] < \infty .$$

Therefore, the matrix **A** is defined to be

$$\text{(A.43)}\quad \mathbf{A} = -E\left[\frac{\partial^2 \ell_t}{\partial \theta_o \partial \theta_o'}\middle| I_{t-1}\right] ,$$

$$= \frac{1}{2}E\left[\frac{1}{h_{1t}^4}\left[\frac{\partial h_{1t}^2}{\partial \theta}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta'}\right]\middle| I_{t-1}\right] + E\left[\frac{1}{h_{1t}^2}\left[\frac{\partial \epsilon_{1t}}{\partial \theta}\right]\left[\frac{\partial \epsilon_{1t}}{\partial \theta'}\right]\middle| I_{t-1}\right] < \infty$$

where the derivatives are evaluated at θ_o.

To show the Det **A** > 0, I follow the method utilized in Lemmas I and II. For any $\lambda \neq 0$, one can transform **A** such that

$$\text{(A.44)}\quad \lambda'\mathbf{A}\lambda = \frac{1}{2}E\left[\lambda'\frac{1}{h_{1t}^4}\left[\frac{\partial h_{1t}^2}{\partial \theta}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta'}\right]\lambda \;\middle| I_{t-1}\right]$$

$$+ E\left[\lambda'\frac{1}{h_{1t}^2}\left[\frac{\partial \epsilon_{1t}}{\partial \theta}\right]\left[\frac{\partial \epsilon_{1t}}{\partial \theta'}\right]\lambda \;\middle| I_{t-1}\right]$$

Both terms on the right hand side are nonnegative. What remains to be shown is that these are both greater than zero. First, partition λ' $= (\lambda_1', \lambda_2')$ to conform with $\theta' = (m', v')$. Now, since $\partial\epsilon_{1t}/\partial v = 0$ the second term in (A.44) is equal to

$$\text{(A.45)} \qquad E\left[\lambda' \frac{1}{h_{1t}^2}\left[\frac{\partial\epsilon_{1t}}{\partial\theta}\right]\left[\frac{\partial\epsilon_{1t}}{\partial\theta'}\right]\lambda \,\Big|\, I_{t-1}\right] = E\left[\lambda_2' \frac{1}{h_{1t}^2}\left[\frac{\partial\epsilon_{1t}}{\partial\theta}\right]\left[\frac{\partial\epsilon_{1t}}{\partial\theta'}\right]\lambda_2 \,\Big|\, I_{t-1}\right]$$

From Lemma I this is clearly positive unless $\lambda_2 = 0$. If $\lambda_2 = 0$, then $\lambda_1 \neq 0$ and the first term of (A.44) becomes

$$\text{(A.46)} \qquad E\left[\lambda' \frac{1}{h_{1t}^4}\left[\frac{\partial h_{1t}^2}{\partial\theta}\right]\left[\frac{\partial h_{1t}^2}{\partial\theta'}\right]\lambda \,\Big|\, I_{t-1}\right] = E\left[\lambda_1' \frac{1}{h_{1t}^4}\left[\frac{\partial h_{1t}^2}{\partial\theta}\right]\left[\frac{\partial h_{1t}^2}{\partial\theta'}\right]\lambda_1 \,\Big|\, I_{t-1}\right]$$

which is positive because of Lemma II. □

<u>Proof of THEOREM I</u>:

From the ergodic theorem, for any $\theta \in \Xi$,

$$\text{(A.47)} \qquad L(\theta) = \lim_{t\to\infty} \ell_t$$

$$= E(C_1) - \tfrac{1}{2}E\left[\log h_{1t}^2\right] - \tfrac{1}{2}E\left[\log(a_o' W_{11,t} a_o)\right]$$

$$- \tfrac{1}{2}E\left[\epsilon_{1t}^2/h_{1t}^2\right] \qquad \text{a.s.}$$

if the expectations exist. From Jensen's Inequality it follows that

$$\text{(A.48)} \qquad \log E(X) \geq E\Big[\log(x)\Big] ,$$

for all positive random variables X, with equality only when X is a constant a.s. This implies that

$$\text{(A.49)} \qquad \log E\Big[h_{1t}^2\Big] \geq E\Big[\log h_{1t}^2\Big] .$$

Since, by definition, $E\Big[h_{1t}^2\Big] < \infty$ then from (A.47) it must be true that $E\Big[\log(h_{1t}^2)\Big] < \infty$. From Lemma I it was shown that $E(\epsilon_{1t}^2) < \infty$, for all $\theta \in \Xi$. Then since

$$\text{(A.50)} \qquad E\Big[\epsilon_{1t}^2/h_{1t}^2\Big] \leq E\Big[\epsilon_{1t}^2\Big] ,$$

then clearly

$$\text{(A.51)} \qquad E\Big[\epsilon_{1t}^2/h_{1t}^2\Big] < \infty .$$

Since $W_{11,t}$ is the second moment matrix of residuals from the regression of Y_{1t} on X_{1t}, the vector of all exogenous variables, it is equal to zero when evaluated at the true parameter values.

Next, following Weiss (1986), make the following transformations,

$$\text{(A.52)} \qquad E\Big[\epsilon_{1t}^2/h_{1t}^2\Big] = E\Big[h_{1t}^{-2}(\epsilon_{1t} + \epsilon_{1t}^{\circ} - \epsilon_{1t}^{\circ})^2\Big]$$

$$= E\Big[h_{1t}^{-2}(\epsilon_{1t}^{\circ})^2\Big] + E\Big[h_{1t}^{-2}(\epsilon_{1t} - \epsilon_{1t}^{\circ})^2\Big] .$$

Since $h_{1t}^{-2}(\epsilon_{1t} - \epsilon_{1t}^{o})^2$ depends only on information available at time t-1 and $E(\epsilon_{1t}^{o}|I_{t-1}) = 0$. From (A.52) it is clear that

$$\text{(A.53)} \qquad E\left[\epsilon_{1t}^2/h_{1t}^2\right] > E\left[(\epsilon_{1t}^{o})^2/h_{1t}^2\right] .$$

This holds with equality when $\epsilon_{1t} = \epsilon_{1t}^{o}$, for all t a.s. Taking expectations of an expression for ϵ_{1t}^2 similar to (A.55) and based on Lemma I, yields

$$\text{(A.54)} \quad E\left[\epsilon_{1t}^2\right] = E\left[(\epsilon_{1t}^{o})^2\right] + (m - m_o)'E\left[\frac{\partial\epsilon_{1t}}{\partial m}\frac{\partial\epsilon_{1t}}{\partial m'}\right](m - m_o) \geq E\left[(\epsilon_{1t}^{o})^2\right]$$

with equality only at $m = m_o$. Thus from (A.54) one can see that $\epsilon_{1t} = \epsilon_{1t}^{o}$, for all t a.s., only when $m = m_o$.

$$\begin{aligned} \text{(A.55)} \qquad L(\theta) &= E(C_1) - \tfrac{1}{2}E\left[\log h_{1t}^2\right] - \tfrac{1}{2}E\left[\log(a_o'W_{11,t}a_o)\right] \\ &\quad - \tfrac{1}{2}\log\left[E\left[\epsilon_{1t}^2\, h_{1t}^{-2}\right]\right] , \\ &\leq E(C_1) - \tfrac{1}{2}E\left[\log h_{1t}^2\right] - \tfrac{1}{2}E\left[\log(a_o'W_{11,t}a_o)\right] \\ &\quad - \tfrac{1}{2}\log\left[E\left[(\epsilon_{1t}^{o})^2\, h_{1t}^{-2}\right]\right] , \end{aligned}$$

$$\left[\text{since } E\left[(\epsilon_{1t}^{o})^2\, h_{1t}^{-2}\right] = E\left[(\epsilon_{1t}^{o})^2\, h_{1t}^{-2}\, (h_{1t}^{o})^2\, (h_{1t}^{o})^{-2}\right]\right]$$

$$= E(C_1) - \frac{1}{2}E\left[\log h_{1t}^2\right] - \frac{1}{2}E\left[\log(a_o' W_{11,t} a_o)\right]$$

$$- \frac{1}{2}\log\left|E\left[(\epsilon_{1t}^o)^2 \; (h_{1t}^o)^{-2}\right]\right| - \frac{1}{2}\log\left|E\left[h_{1t}^{-2} \; (h_{1t}^o)^2\right]\right| ,$$

$$\left[\text{since } \log\left|E\left[h_{1t}^{-2} \; (h_{1t}^o)^2\right]\right| \geq E\left|\log\left[h_{1t}^{-2} \; (h_{1t}^o)^2\right]\right| \right]$$

$$\leq E(C_1) - \frac{1}{2}E\left[\log h_{1t}^2\right] - \frac{1}{2}E\left[\log(a_o' W_{11,t} a_o)\right]$$

$$- \frac{1}{2}E\left|\log\left[(\epsilon_{1t}^o)^2 \; (h_{1t}^o)^{-2}\right]\right| - \frac{1}{2}E\left|\log\left[h_{1t}^{-2} \; (h_{1t}^o)^2\right]\right| ,$$

$$= E(C_1) - \frac{1}{2}E\left[\log(h_{1t}^o)^2\right] - \frac{1}{2}E\left[\log(a_o' W_{11,t} a_o)\right]$$

$$- \frac{1}{2}E\left|\log\left[(\epsilon_{1t}^o)^2 \; (h_{1t}^o)^{-2}\right]\right| ,$$

$$\leq E(C_1) - \frac{1}{2}E\left[\log(h_{1t}^o)^2\right] - \frac{1}{2}E\left|\log\left[(\epsilon_{1t}^o)^2 \; (h_{1t}^o)^{-2}\right]\right| ,$$

$$= L(\theta_o)$$

with equality only at $\theta = \theta_o$. Note that

$$\text{(A.56)} \qquad \log\left|E\left[h_{1t}^{-2} \; (h_{1t}^o)^2\right]\right| = E\left|\log\left[h_{1t}^{-2} \; (h_{1t}^o)^2\right]\right|$$

only when $(h_{1t}^o)^2 = h_{1t}^2$. But based on Lemma II, this can only occur at $\theta = \theta_o$. □

Proof of THEOREM II:

This proof follows closely the proofs of Weiss (1986) and Basawa, Feigin, and Heyde (1976). I begin by first applying a Taylor series expansion to the first order derivatives for the conditional log-likelihood function. This expansion can be written as

$$\text{(A.57)} \qquad \frac{\partial L}{\partial \hat{\theta}} = \frac{\partial L}{\partial \theta_o} + \frac{\partial^2 L}{\partial \theta_o \partial \theta_o'}(\hat{\theta} - \theta_o) + \frac{1}{2}\frac{\partial^3 L}{\partial \tilde{\theta}^3}(\hat{\theta} - \theta_o)$$

where $|\hat{\theta} - \theta_o| > |\tilde{\theta} - \theta_o|$, which requires that $\tilde{\theta}$ lie between θ_o and $\hat{\theta}$. The third term can be considered as a remainder term on the Taylor series expansion. Equation (A.57) can be rewritten, based on (2.4), as

$$\text{(A.58)} \qquad \frac{\partial L}{\partial \hat{\theta}} = \sum_{t=1}^{T} \frac{\partial \ell_t}{\partial \theta_o} + (\hat{\theta} - \theta_o) \sum_{t=1}^{T} \frac{\partial^2 \ell_t}{\partial \theta_o \partial \theta_o'} + (\hat{\theta} - \theta_o) \sum_{t=1}^{T} \frac{\partial^3 \ell_t}{\partial \tilde{\theta}^3}$$

Basawa, Feigin, and Heyde derive a set of sufficient conditions for the existence of a consistent root of the log-likelihood equation

$$\text{(A.59)} \qquad \frac{\partial L}{\partial \hat{\theta}} = 0 .$$

These sufficient conditions consists of

$$\text{(I)} \qquad \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \ell_t}{\partial \theta_o} \xrightarrow{P} 0 .$$

(II) There exists a nonrandom matrix $M(\theta_o) > 0$ such that for all $\epsilon > 0$,

$$\Pr\left[-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ell_t}{\partial\theta_o \partial\theta_o'} \geq M(\theta_o)\right] > 1 - \epsilon$$

for all $T > T_1(\epsilon)$.

(III) There exists a constant $M < \infty$ such that

$$E\left|\frac{\partial^3 \ell_t}{\partial\theta_i \partial\theta_j \partial\theta_k}\right| < M \quad \text{for all } \theta \in \Xi \; .$$

Then it follows that $\hat{\theta} \xrightarrow{p} \theta_o$.

The method of proof is to show that these three conditions are satisfied. I begin by focusing on the parameters of the current endogenous variables a_o.

Condition (I): From equation (A.1), one has

$$\text{(A.50)} \quad \left.\frac{\partial L}{\partial a_o}\right|_{a_o=\hat{a}_o} = \sum_{t=1}^{T}\left[-\left[\frac{a_o' W_{11,t}}{a_o' W_{11,t} a_o}\right] - \frac{\epsilon_{1t}' Y_{1t}}{2h_{1t}^2} + \frac{1}{2h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial a_o}\right]\left[\frac{\epsilon_{1t}'\epsilon_{1t}}{h_{1t}^2} - 1\right]\right] = 0 \; .$$

When this derivative is evaluated at the true a_o, the first term is zero since $W_{11,t}$ is the second moment matrix of residuals from the regression of Y_{1t} on X_t. Then, following Weiss (1986), it can be shown that

$$(A.61) \qquad E\left[\frac{\partial L}{\partial a_o} \middle| I_{t-1}\right] = 0$$

since $E(\epsilon_{1t}|I_{t-1}) = 0$ and $E(\epsilon_{1t}^2|I_{t-1}) = h_{1t}^2$. The ergodic theorem, see White (1984), then implies that

$$(A.62) \qquad \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \ell_t}{\partial a_o} \xrightarrow{p} 0 .$$

Condition (II): By the ergodic theorem, for any constant vector $\lambda \neq 0$,

$$(A.63) \qquad \frac{1}{T}\sum_{t=1}^{T} \lambda' \left[\frac{\partial^2 \ell_t}{\partial a_o \partial a_o'}\right]\lambda \longrightarrow E\left[\lambda' \frac{\partial^2 \ell_t}{\partial a_o \partial a_o'} \lambda\right] \quad \text{a.s.}$$

$$= -\lambda' A_o \lambda$$

where A_o is defined to be

$$(A.64) \qquad A_o = - E\left[\frac{\partial^2 L}{\partial a_o \partial a_o'}\right]$$

This second order derivative, after some simplification, has the following form,

$$
\begin{aligned}
\text{(A.65)} \quad \frac{\partial^2 L}{\partial a_o \partial a_o'} = \sum_{t=1}^{T} \Bigg[& \frac{W_{11,t}}{a_o W_{11,t} a_o} - 2\frac{a_o' W_{11,t}^2 a_o}{(a_o' W_{11,t} a_o)^2} - \frac{Y_{1t}' Y_{1t}}{2h_{1t}^2} \\
& + \alpha_1 \left[\frac{\epsilon_{1t}' Y_{1t}}{2h_{1t}^2} \right] \sum_{i=1}^{T-1} \delta_1^{i-1} \, Y_{1t-i}' \epsilon_{1t-i} \\
& + \frac{\alpha_1}{h_{1t}^2} \left[\frac{\epsilon_{1t}' \epsilon_{1t}}{h_{1t}^2} - 1 \right] \sum_{i=1}^{T-1} \delta_1^{i-1} \, Y_{1t-i}' Y_{1t-i} \\
& + \sum_{i=1}^{T-1} \delta_1^{i-1} \, \epsilon_{1t-i}' Y_{1t-i} \Bigg[\frac{\alpha_1}{h_{1t}^4} \Bigg[Y_{1t}' \epsilon_{1t} + 2 \left[1 - 2\frac{\epsilon_{1t}' \epsilon_{1t}}{h_{1t}^2} \right] \\
& \sum_{i=1}^{T-1} \delta_1^{i-1} \, Y_{1t-i}' \epsilon_{1t-i} \Bigg] \Bigg] \Bigg]
\end{aligned}
$$

Taking negative expectations of (A.65) yields

$$
\text{(A.66)} \quad - E\left[\frac{\partial^2 L}{\partial a_o \partial a_o'} \right] = \sum_{t=1}^{T} \frac{Y_{1t}' Y_{1t}}{2h_{1t}^2} > 0 ,
$$

since $E(\epsilon_{1t} | I_{t-1}) = E(\epsilon_{1t-i} | I_{t-1}) = 0$ and $E(\epsilon_{1t}^2 | I_{t-1}) = h_{1t}^2$. Now, following Weiss (1986), let

$$\text{(A.67)} \qquad 0 < \delta(\lambda) < \frac{1}{2}\lambda' E\left[\frac{\partial^2 \ell_t}{\partial a_o \partial a_o'}\right]\lambda$$

for a given λ. Then, for all $\epsilon > 0$, there exists $T_1 = T_1(\epsilon)$ such that

$$\text{(A.68)} \qquad \Pr\left[\left\| \frac{1}{T}\sum_{t=1}^{T} \lambda' \frac{\partial^2 \ell_t}{\partial a_o \partial a_o'}\lambda - E\left[\lambda' \frac{\partial^2 \ell_t}{\partial a_o \partial a_o'}\lambda\right]\right\| < \delta\right] > 1 - \epsilon$$

for all $T > T_1$. Thus, condition (II) is verified by defining $M(a_o)$ to be

$$\text{(A.69)} \qquad M(a_o) = -\frac{1}{2} E\left[\frac{\partial^2 \ell_t}{\partial a_o \partial a_o'}\right].$$

Then

$$\text{(A.70)} \qquad \Pr\left[\frac{1}{T}\sum_{t=1}^{T} \lambda' \frac{\partial^2 \ell_t}{\partial a_o \partial a_o'}\lambda > \lambda' M(a_o)\lambda\right] > 1 - \epsilon$$

for all $T > T_1$.

Condition (III): The third condition to be verified requires that the third order derivatives, which can be interpreted to be the remainder term of the Taylor series expansion, be bounded in absolute value. The third order derivative is derived by differentiating equation (A.65) with respect to a_o. It is clear that the only terms that cannot be dealt with as above are those terms which contain third order derivatives of h_{1t}^2 and ϵ_{1t}. Following Weiss (1986), one is able to show that an extension of the methods applied to the first and second derivatives implies that terms containing $\partial^3 h_{1t}^2/\partial a_o^3$ are bounded and that terms containing $\partial^3 \epsilon_{1t}/\partial a_o^3$ have bounded second moments. □

The proof of consistency for the remaining parameters follows as above.

<u>Proof of THEOREM III</u>:

Basawa, Feigin, and Heyde (1976) again provide a set of sufficient conditions for asymptotic normality. These verifiable conditions are

$$\text{(I)} \qquad T^{-1/2} \sum_{t=1}^{T} \frac{\partial \ell_t}{\partial \theta_o} \overset{A}{\sim} N(0, \mathbf{B}_o) \qquad \text{for nonrandom } \mathbf{B}_o > 0.$$

$$\text{(II)} \qquad T^{-1} \sum_{t=1}^{T} \frac{\partial^2 \ell_t}{\partial \theta_o \partial \theta_o'} \xrightarrow{p} -\mathbf{A}_o \qquad \text{for nonrandom } \mathbf{A}_o > 0.$$

(III) Condition (III) of Theorem II.

The method of proof, as before, is to show that these three conditions are satisfied.

Condition (I): Following Weiss (1986), assume that $\mathbf{B}_o > 0$. Then from Condition (I) of the proof of Theorem I, it follows that

$$\text{(A.71)} \qquad E\left[\frac{\partial \ell_t}{\partial \theta_o} \middle| I_{t-1} \right] = 0 .$$

Weiss shows that if

$$\text{(A.72)} \qquad E\left[\frac{\partial \ell_t}{\partial \theta_o}\frac{\partial \ell_t}{\partial \theta_o'}\right] < \infty$$

and

$$\text{(A.73)} \qquad E\left[T\,\frac{\partial L}{\partial \theta_o}\frac{\partial L}{\partial \theta_o'}\right] = B_o < \infty ,$$

then a Martingale central limit theorem is applicable. Furthermore, Weiss shows that when (A.71) is true then (A.72) and (A.73) are equivalent. Thus, all that is required is to show that (A.72) holds.

To show that (A.72) is true I follow the method in Lemma III used to show **A** $< \infty$. From (A.37) one has

$$\text{(A.74)} \qquad E\left[\frac{\partial \ell_t}{\partial \theta}\frac{\partial \ell_t}{\partial \theta'}\right] = E\left[-\frac{\epsilon_{1t}}{h_{1t}^2}\left[\frac{\partial \epsilon_{1t}}{\partial \theta}\right] + \frac{1}{2h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial \theta}\right]\left[\frac{\epsilon_{1t}^2}{h_{1t}^2} - 1\right]\right]'$$

$$\left[-\frac{\epsilon_{1t}}{h_{1t}^2}\left[\frac{\partial \epsilon_{1t}}{\partial \theta}\right] + \frac{1}{2h_{1t}^2}\left[\frac{\partial h_{1t}^2}{\partial \theta}\right]\left[\frac{\epsilon_{1t}^2}{h_{1t}^2} - 1\right]\right]$$

$$= E\left[\frac{\epsilon_{1t}^2}{h_{1t}^4}\left[\frac{\partial \epsilon_{1t}}{\partial \theta}\right]\left[\frac{\partial \epsilon_{1t}}{\partial \theta'}\right] + \frac{\epsilon_{1t}}{h_{1t}^4}\left[\frac{\partial \epsilon_{1t}}{\partial \theta}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta'}\right]\right.$$

$$\left.\left[\frac{\epsilon_{1t}^2}{h_{1t}^2} - 1\right] + \frac{1}{4h_{1t}^4}\left[\frac{\partial h_{1t}^2}{\partial \theta}\right]\left[\frac{\partial h_{1t}^2}{\partial \theta'}\right]\left[\frac{\epsilon_{1t}^2}{h_{1t}^2} - 1\right]^2\right]$$

where the derivatives are evaluated at θ_o. Taking expectations yields

$$(A.75) \qquad E\left[\frac{\partial \ell_t}{\partial \theta}\frac{\partial \ell_t}{\partial \theta'}\right] = E\left[\frac{1}{h_{1t}^4}\left[\frac{\partial \epsilon_{1t}}{\partial \theta}\right]\left[\frac{\partial \epsilon_{1t}}{\partial \theta'}\right]\right],$$

which from Lemma III is clearly bounded.

Condition (II): Since by definition

$$(A.76) \qquad A_o = -E\left[\frac{\partial^2 \ell_t}{\partial \theta_o \partial \theta_o'}\middle| I_{t-1}\right]$$

then Lemma III verifies condition II. □

<u>Proof of THEOREM IV</u>:

This proof is nearly identical to Theorem II. Again, one must verify the three sufficient conditions derived by Basawa, Feigin, and Heyde (1976) and given in the proof of Theorem II. I begin by focusing on the parameters of the endogenous variables, Γ.

Condition I: From equation (4.8), the first order condition is given by

$$(A.77) \qquad \frac{\partial L}{\partial \Gamma^{\mu}} = \sum_{t=1}^{T}\left[(\Gamma^{-1})' + \frac{1}{2}\left[\frac{\partial H_t}{\partial \Gamma^{\mu}}\right]\left[\frac{\epsilon_t'\epsilon_t}{H_t} - 1\right] - \frac{\epsilon_t'Y_t}{H_t}\right].$$

Clearly, since $E(\epsilon_{1t}|I_{t-1}) = 0$ and $E(\epsilon_{1t}^2|I_{t-1}) = h_{1t}^2$,

$$(A.78) \qquad E\left[\frac{\partial L}{\partial \Gamma^{\mu}} \middle| I_{t-1}\right] = 0$$

Then, the ergodic theorem implies that

$$(A.79) \qquad \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \ell_t}{\partial \Gamma^{\mu}} \xrightarrow{p} 0 \, .$$

Condition (II): By the ergodic theorem, for any constant vector $\lambda \neq 0$,

$$(A.80) \qquad \frac{1}{T} \sum_{t=1}^{T} \lambda' \left[\frac{\partial^2 \ell_t}{\partial \Gamma^{\mu} \partial \Gamma^{\mu'}}\right] \lambda \longrightarrow E\left[\lambda' \frac{\partial^2 \ell_t}{\partial \Gamma^{\mu} \partial \Gamma^{\mu'}} \lambda\right] = -\lambda' A_o \lambda$$

a.s.

where A_o is defined to be

$$(A.81) \qquad A_o = -E\left[\frac{\partial^2 L}{\partial \Gamma^{\mu} \partial \Gamma^{\mu'}}\right]$$

This second order derivative for the conditional log-likelihood function is given by

$$(A.82) \qquad \frac{\partial^2 L}{\partial \Gamma^{\mu} \partial \Gamma^{\mu'}} = \sum_{t=1}^{T} \Bigg[-(\Gamma'\Gamma)^{-1} + \frac{Y_t'\epsilon_t}{H_t}\left[\frac{\partial H_t}{\partial \Gamma^{\mu}}\right] - \frac{1}{2}\left[\frac{\epsilon_t'\epsilon_t}{H_t}\right]\left[\frac{1}{H_t}\right]\left[\frac{\partial H_t}{\partial \Gamma^{\mu}}\right]\left[\frac{\partial H_t}{\partial \Gamma^{\mu'}}\right]$$

$$- \frac{1}{2}\left[1 - \frac{\epsilon_t'\epsilon_t}{H_t}\right]\left[\frac{\partial^2 H_t}{\partial \Gamma^{\mu} \partial \Gamma^{\mu'}}\right] - \frac{Y_t'Y_t}{H_t} - \frac{\epsilon_t'Y_t}{H_t}\left[\frac{1}{H_t}\right]\left[\frac{\partial H_t}{\partial \Gamma^{\mu'}}\right]\Bigg] \, .$$

Taking negative expectations conditional on information available at time t-1 of (A.82) yields

$$\text{(A.83)} \quad -E\left[\frac{\partial^2 L}{\partial\Gamma^{\mu}\partial\Gamma^{\mu'}}\middle| I_{t-1}\right] = \sum_{t=1}^{T}\left[-(\Gamma'\Gamma)^{-1} + \frac{1}{2}\left[\frac{1}{H_t}\right]\left[\frac{\partial H_t}{\partial\Gamma^{\mu}}\right]\left[\frac{\partial H_t}{\partial\Gamma^{\mu'}}\right] + \frac{Y_t'Y_t}{H_t}\right] > 0$$

As in Theorem II, for a given λ, let δ be bounded by

$$\text{(A.84)} \qquad 0 < \delta(\lambda) < \frac{1}{2}\lambda' E\left[\frac{\partial^2 \ell_t}{\partial\Gamma^{\mu}\partial\Gamma^{\mu'}}\right]\lambda .$$

Then, for all $\epsilon > 0$, there exists $T_1 = T_1(\epsilon)$ such that

$$\text{(A.85)} \qquad \Pr\left[\left\| \frac{1}{T}\sum_{t=1}^{T}\lambda'\frac{\partial^2 \ell_t}{\partial\Gamma^{\mu}\partial\Gamma^{\mu'}}\lambda - E\left[\lambda'\frac{\partial^2 \ell_t}{\partial\Gamma^{\mu}\partial\Gamma^{\mu'}}\lambda\right]\right\| < \delta\right] > 1 - \epsilon$$

for all $T > T_1$.

Thus, condition (II) is verified by defining $M(\Gamma)$ to be

$$\text{(A.86)} \qquad M(\Gamma) = -\frac{1}{2} E\left[\frac{\partial^2 \ell_t}{\partial\Gamma^{\mu}\partial\Gamma^{\mu'}}\right] .$$

Then

$$\text{(A.87)} \qquad \Pr\left[\frac{1}{T}\sum_{t=1}^{T}\lambda'\frac{\partial^2 \ell_t}{\partial\Gamma^{\mu}\partial\Gamma^{\mu'}}\lambda > \lambda' M(\Gamma)\lambda\right] > 1 - \epsilon$$

for all $T > T_1$.

Condition (III): See Theorem II, condition III. □

The proof of consistency for the remaining parameters follows as above.

Proof of THEOREM V:

As with Theorem IV, this proof is nearly identical to its single equation counterpart, Theorem III. The proof requires verifying the three sufficient conditions for asymptotic normality outlines by Basawa, Feigin, and Heyde (1976) and given in the proof of Theorem III. As before, I will focus on the parameters of the endogenous variables, Γ.

Condition I: Since it is assumed that $\mathbf{B_o}>0$, then from Condition I from the proof of Theorem IV, it follows that

$$\text{(A.90)} \qquad E\left[\frac{\partial L}{\partial \Gamma^{\mu}}\middle| I_{t-1}\right] = 0$$

As explained in Theorem III, given the (A.90) is true, all that is required to verify Condition I is to show that

$$\text{(A.91)} \qquad E\left[\frac{\partial \ell_t}{\partial \Gamma^{\mu}}\frac{\partial \ell_t}{\partial \Gamma^{\mu\prime}}\right] < \infty$$

From (A.77) condition (A.91) can be written as

$$\text{(A.92)} \qquad E\left[\frac{\partial \ell_t}{\partial \Gamma^{\mu}}\frac{\partial \ell_t}{\partial \Gamma^{\mu\prime}}\right] = E\left[\left[(\Gamma^{-1})' + \left[\frac{\partial H_t}{\partial \Gamma^{\mu}}\right]\left[\frac{\epsilon_t'\epsilon_t}{H_t} - 1\right] - \frac{\epsilon_t'Y_t}{H_t}\right]'\right.$$

$$\left.\cdot\left[(\Gamma^{-1})' + \left[\frac{\partial H_t}{\partial \Gamma^{\mu}}\right]\left[\frac{\epsilon_t'\epsilon_t}{H_t} - 1\right] - \frac{\epsilon_t'Y_t}{H_t}\right]\right]$$

$$= E\left[(\Gamma^{-1})'(\Gamma^{-1}) + \frac{Y_t'\epsilon_t\epsilon_t'Y_t}{H_t'H_t} - \frac{1}{4}\left[\frac{\partial H_t}{\partial \Gamma^{\mu}}\right]\left[\frac{\epsilon_t'\epsilon_t}{H_t} - 1\right]\right.$$

$$\left[\frac{\epsilon_t'\epsilon_t}{H_t} - 1\right]'\left[\frac{\partial H_t}{\partial \Gamma^{\mu}}\right] + (\Gamma^{-1})\left[\frac{\partial H_t}{\partial \Gamma^{\mu}}\right]\left[\frac{\epsilon_t'\epsilon_t}{H_t} - 1\right]$$

$$\left. - 2(\Gamma^{-1})\frac{\epsilon_t'Y_t}{H_t}\right]$$

where the derivatives are evaluated at the true Γ^{μ}. Taking expectations yields

$$\text{(A.93)} \qquad E\left[\frac{\partial \ell_t}{\partial \Gamma^{\mu}}\frac{\partial \ell_t}{\partial \Gamma^{\mu\prime}}\right] = E\left[(\Gamma^{-1})'(\Gamma^{-1}) + \frac{Y_t'\epsilon_t\epsilon_t'Y_t}{H_t'H_t}\right] < \infty$$

Condition II: By definition

$$\text{(A.94)} \qquad \mathbf{A_o} = -E\left[\frac{\partial^2 \ell_t}{\partial \Gamma^{\mu}\partial \Gamma^{\mu\prime}}\middle| I_{t-1}\right]$$

An argument identical to Lemma III verifies Condition II.

Condition III: Similar to Theorem II, Condition III. □

International Finance Discussion Papers

IFDP NUMBER	TITLES	AUTHOR(s)
	1988	
322	The Simultaneous Equations Model with Generalized Autoregressive Conditional Heteroskedasticity: The SEM-GARCH Model	Richard Harmon
321	Adjustment Costs and International Trade Dynamics	Joseph E. Gagnon
320	The Capital Flight "Problem"	David B. Gordon Ross Levine
	1987	
319	Modeling Investment Income and Other Services in the U.S. International Transactions Accounts	William Helkie Lois Stekler
318	Improving the Forecast Accuracy of Provisional Data: An Application of the Kalman Filter to Retail Sales Estimates	B. Dianne Pauls
317	Monte Carlo Methodology and the Finite Sample Properties of Statistics for Testing Nested and Non-Nested Hypotheses	Neil R. Ericsson
316	The U.S. External Deficit: Its Causes and Persistence	Peter Hooper Catherine L. Mann
315	Debt Conversions: Economic Issues for Heavily Indebted Developing Countries	Lewis S. Alexander
314	Exchange Rate Regimes and Macroeconomic Stabilization in a Developing Country	David H. Howard
313	Monetary Policy in Taiwan, China	Robert F. Emery
312	The Pricing of Forward Exchange Rates	Ross Levine
311	Realignment of the Yen-Dollar Exchange Rate: Aspects of the Adjustment Process in Japan	Bonnie E. Loopesko Robert E. Johnson
310	The Effect of Multilateral Trade Clearinghouses on the Demand for International Reserves	Ellen E. Meade
309	Protection and Retaliation: Changing the Rules of the Game	Catherine L. Mann

---

Please address requests for copies to International Finance Discussion Papers, Division of International Finance, Stop 24, Board of Governors of the Federal Reserve System, Washington, D.C. 20551.