International Finance Discussion Paper

Number 322

May 1988

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

Richard Harmon

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment by a writer that he has had access to unpublished material) should be cleared with the author or authors.
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 HETEROSEDASTICITY:
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

\[(1.1) \quad h_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}^2.\]

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:

\[
(1.2) \quad Y_t \Gamma + X_t \beta = \epsilon_t, \quad t = 1, \ldots, T,
\]

where \( \Gamma \) is a \( M \times M \) matrix of coefficients of current endogenous variables, \( \beta \) is a \( K \times M \) matrix of coefficients of predetermined variables, \( \epsilon_t \) 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, \( \epsilon_t \), is given by

\[
(1.3) \quad \epsilon_t \sim N(0, \Sigma),
\]

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

\[
(1.4) \quad L(Y_1, \ldots, Y_T | X) = (2\pi)^{-TM/2} ||\Gamma||^{-T} \Sigma^{-T/2} \exp\left[-\frac{1}{2} \sum_{t=1}^{T} (Y_t \Gamma + X_t \beta) \Sigma^{-1}(Y_t \Gamma + X_t \beta)'\right]
\]

Leaving aside the variance-covariance matrix \( \Sigma \), 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 \( \Gamma \), \( \beta \), or \( \Sigma \), 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

\[(1.5) \quad Y_t \Gamma + X_t \beta = \epsilon_t, \]

\[(1.6) \quad (\epsilon_t | I_{t-1}) \sim N(0, H_t), \]

where

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

\[- W + \left[ I_M \otimes \epsilon_{t-1} \right] \mathcal{C}_1 \left[ I_M \otimes \epsilon_{t-1} \right] + \cdots + \left[ I_M \otimes \epsilon_{t-q} \right] \mathcal{C}_q \left[ I_M \otimes \epsilon_{t-q} \right]
+ \left[ I_M \otimes h_{t-1} \right] \mathcal{D}_1 \left[ I_M \otimes h_{t-1} \right] + \cdots + \left[ I_M \otimes h_{t-p} \right] \mathcal{D}_p \left[ I_M \otimes h_{t-p} \right],
\]

where \( \epsilon_{t-1} \) and \( h_{t-1} \) are \( M \)-element row vectors consisting of \( i \)th lagged innovations and \( i \)th 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 \( i \)th equation when \( i=j \) and the conditional covariance between the \( i \)th and \( j \)th 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 $\epsilon_t$ 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) $\operatorname{plim} \frac{XX'}{T} = Q$, where $Q$ is a finite and nonsingular matrix,

(ii) $\operatorname{plim} \frac{X'\epsilon}{T} = 0$,

(iii) $\operatorname{plim} \frac{\epsilon'\epsilon}{T} = \Sigma$,
where $\Sigma$ is a bounded and positive-definite unconditional variance-covariance matrix.

A alternative parameterization of the SEM-GARCH(p,q) model facilitates estimation and highlights the alternative variance-covariance specifications available. Let $H^*_t$ 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

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

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

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

where $C_1$ and $D_1$ are $4 \times 4$ parameter matrices given by

---

1An excellent discussion of Central Limit Theory for the non-i.i.d. case is provided in White (1984) Ch.5.
(1.9) \[ C_1 = \begin{bmatrix} \begin{array}{cc|cc} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ \hline C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{array} \end{bmatrix} \]

(1.10) \[ D_1 = \begin{bmatrix} \begin{array}{cc|cc} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ \hline D_{31} & D_{32} & D_{33} & D_{34} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{array} \end{bmatrix} \]

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 \( h_t^* \),

\[
(1.11) \quad h_t^* = \begin{bmatrix} H_{11,t} \\ H_{21,t} \\ H_{22,t} \end{bmatrix}
\]

\[
= \begin{bmatrix} w_{11} \\ w_{21} \\ w_{22} \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}^2 \end{bmatrix}
\]
which in matrix notation can be written as

\[(1.12) \ h_t^* = \omega + A_{\eta_{t-1}} + B_{h_{t-1}}. \]

where \(\eta_{t-1} = \text{vec}(\epsilon_{t-1}' \epsilon_{t-1})\) 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",

\[(1.13) \ H_{11,t} = \omega_{11} + A_{11} \epsilon_{1,t-1}^2 + A_{12} (\epsilon_{2,t-1}' \epsilon_{1,t-1}) \]

\[+ A_{13} \epsilon_{2,t-1}^2 + B_{11} H_{11,t-1} + B_{12} H_{21,t-1} \]

\[+ B_{13} H_{22,t-1}, \]

\[(1.14) \ H_{21,t} = \omega_{21} + A_{21} \epsilon_{1,t-1}^2 + A_{22} (\epsilon_{2,t-1}' \epsilon_{1,t-1}) \]

\[+ A_{23} \epsilon_{2,t-1}^2 + B_{21} H_{11,t-1} + B_{22} H_{21,t-1} \]

\[+ B_{23} H_{22,t-1}, \]
\begin{equation}
H_{22,t} = \omega_{22} + A_{31} \varepsilon_{1,t-1}^2 + A_{32} (\varepsilon_{2,t-1} \varepsilon_{1,t-1}) + A_{33} \varepsilon_{2,t-1}^2 + B_{31} H_{11,t-1} + B_{32} H_{21,t-1} + B_{33} H_{22,t-1}.
\end{equation}

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 $C_1$ and $D_1$ given by

\begin{equation}
C_1 = \begin{bmatrix}
C_{11} & 0 & \cdots & 0 & C_{14} \\
0 & 0 & \cdots & C_{23} & 0 \\
0 & \cdots & \cdots & 0 & 0 \\
C_{32} & 0 & \cdots & 0 & 0 \\
C_{41} & 0 & \cdots & 0 & C_{44}
\end{bmatrix}
\end{equation}
Table 1

**THE SEM-GARCH(p,q) MODEL**

The Conditional Variance-Covariance Matrix

<table>
<thead>
<tr>
<th>GARCH Model</th>
<th>Number Of Equations</th>
<th>Number Of Parameters To Be Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Unrestricted</strong></td>
</tr>
<tr>
<td>(1,1)</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>465</td>
</tr>
<tr>
<td>(2,2)</td>
<td>2</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>915</td>
</tr>
<tr>
<td>(3,3)</td>
<td>2</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>222</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>610</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1365</td>
</tr>
</tbody>
</table>

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

** Diagonal Var-Cov: \( \# = (1+p+q) \sum_{i=1}^{M} i \) where \( M \) is the number of equations in the system.
(1.17) \[ D_1 = \begin{bmatrix} D_{11} & 0 & 0 & D_{14} \\ 0 & D_{23} & 0 & 0 \\ 0 & 0 & D_{32} & 0 \\ 0 & 0 & 0 & D_{44} \end{bmatrix} \]

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

(1.18) \[ h^*_t = \begin{bmatrix} H_{11,t} \\ H_{21,t} \\ H_{22,t} \end{bmatrix} \]

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

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

Expanding (1.18) yields

(1.19) \[ H_{11,t} = \omega_{11} + A_{11}\epsilon_{1,t-1}^2 + B_{11}H_{11,t-1}, \]
(1.20) \[ H_{21,t} = \omega_{21} + A_{22}(\epsilon_{2,t-1}\epsilon_{1,t-1}) + B_{22}H_{21,t-1} \]

(1.21) \[ H_{22,t} = \omega_{22} + A_{33}\epsilon_{2,t-1}^2 + B_{33}H_{22,t-1} \]

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) \[ H_{11,t} > 0, \quad H_{22,t} > 0, \]

and

(1.23) \[ H_{11,t}H_{22,t} - H_{21,t}^2 > 0. \]

Clearly, a sufficient condition for \( H_t \) to be positive definite is having every term in \( \omega, A, \) 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) \[ \omega_{11} > 0, \quad \omega_{22} > 0 \quad \text{and} \quad \omega_{11} \omega_{22} - \omega_{21}^2 > 0. \]

Similarly, setting \( \omega = B = 0 \) yields

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

Alternatively, setting \( \omega = A = 0 \) yields

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

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. \textbf{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

\[(2.1) \quad Y_t \Gamma_{1} + X_t \beta_{1} = \epsilon_{1t}, \quad t = 1, \ldots, T\]

where \( \Gamma_{1}, \beta_{1}, \) and \( \epsilon_{1t} \) are the first columns of \( \Gamma, \beta, \) and \( \epsilon, \) respectively. The variables are arranged so that the usual identifying restrictions may be shown by the following partitions:

\[(2.2) \quad \Gamma_{1} = \begin{bmatrix} a_o \\ \vdots \\ 0 \end{bmatrix} \quad \beta_{1} = \begin{bmatrix} b_o \\ \vdots \\ 0 \end{bmatrix}\]

and

\[(2.3) \quad Y = \begin{bmatrix} Y_1 : Y_1^* \end{bmatrix} \quad X = \begin{bmatrix} X_1 : X_1^* \end{bmatrix}\]

The number of included and excluded endogenous variables in the first

²The 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

\[
\text{(Order Condition): } \quad k_1^* \geq m_1 - 1
\]

\[
\text{(Rank Condition): } \quad \text{Rank}(\phi_1 \Lambda_1) = M - 1
\]

where \( \phi_1 \) is a \( R \times (M+K) \) selection matrix composed of zeros and ones, and \( \Lambda_1 = (\Gamma'_1, \beta'_1) = (a'_o \cdots 0' \cdots b'_o \cdots 0')' \) 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 (1953), 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):

\[
(2.4) \quad L(a_o, b_o, \omega_o, \alpha, \delta | I_{t-1}) = \sum_{t=1}^{T} l_t
\]
(2.5) \[ \ell_t = c_1 - \frac{1}{2} \log(h_{1t}^2) + \frac{1}{2} \log(a_o^t w_{11, t}^t a_o) \]

\[ - \frac{1}{2} \left[ (Y_{1t}^t a_o + X_{1t}^t b_o)(Y_{1t}^t a_o + X_{1t}^t b_o) h_{1t}^{-2} \right] \]

where

(2.6) \[ w_{11, t} = (Y_{1t}^t P Y_{1t}), \]

(2.7) \[ p_t = (1 - X_t (X_t' X_t)^{-1} X_t'), \]

(2.8) \[ c_1 = -\frac{m_t}{2} \left[ \ln(2\pi) + 1 \right] + \frac{1}{2} \left[ 1 - \ln|w_t|^t \right], \]

(2.9) \[ w_t = (Y_t^t P Y_t), \]

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

(2.11) \[ h_{1t}^2 = \omega + \sum_{i=1}^{q} a_i \epsilon_{1, t-i}^2 + \sum_{i=1}^{p} \delta_i h_{1, t-i}^2, \]

(2.12) \[ \epsilon_{1t} = Y_{1t}^t a_o + X_{1t}^t b_o. \]

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 \( W_{\text{ll},t} \) 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 \( a_0, b_0, \omega_0, \alpha \) and \( \delta \) in a nonlinear fashion, maximization of \( L(a_0, b_0, \omega_0, \alpha, \delta | I_{t-1}) \) requires an iterative technique. Maximum Likelihood (ML) estimates of the parameters \( a_0, b_0, \omega_0, \alpha \) and \( \delta \) 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 \textit{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 \( \theta \) 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,

\[
(3.1) \quad \theta' = (m', v') = (a_0', b_0', \nu_0, a', \delta') .
\]

where \( \theta \) is partitioned such that \( m \) is the parameter vector corresponding to the structural equation under investigation, and \( v \) is the parameter vector of the conditional variance specification, given by equation (2.11). Furthermore, I assume \( \theta \in \Xi \), where \( \Xi \) is a compact subspace of Euclidean space, and let \( \theta_o \) 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.
LEMMMA I:

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

$$E \left[ \frac{\partial \epsilon_{lt}}{\partial m} \frac{\partial \epsilon_{lt}}{\partial m'} \right] < M_1$$

and

$$\det E \left[ \frac{\partial \epsilon_{lt}}{\partial m} \frac{\partial \epsilon_{lt}}{\partial m'} \right] > 0.$$ 

Proof: See Appendix.

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

LEMMMA II:

Assume that the fourth moment of $\epsilon_{lt}$ exists and is bounded.

Then for all $\theta \in \Theta$, there exists a constant $M_2 < \infty$, not depending on $\theta$, such that

$$E \left[ \frac{\partial \epsilon_{lt}}{\partial \nu} \frac{\partial \epsilon_{lt}}{\partial \nu'} \right] < M_2$$

and
\[(3.5) \quad \text{Det } E \left[ \frac{\partial^2 \epsilon_{1t}}{\partial \nu \partial \nu'} \right] > 0. \]

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* \( \theta \) *such that*

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

and

\[(3.7) \quad \text{Det } A > 0. \]

Proof: See Appendix.

These results provide the basis for the following theorems.
THEOREM I:

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

\begin{equation}
L = \lim_{T \to \infty} L(\theta) \quad (\text{exists a.s. for all } \theta \in \Theta)
\end{equation}

and the \( \lim_{T \to \infty} L(\theta) \) is uniquely maximized at \( \theta_0 \).

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.

THEOREM II: (Consistency)

For the LIM-GARCH specification given by equations (2.4) to (2.9), the maximum likelihood estimate \( \hat{\theta} \) is consistent for \( \theta_0 \) provided \( \theta_0 \) is interior to \( \Theta \).

Proof: See Appendix.

Note that the condition that \( \theta_0 \) is interior to \( \Theta \) ensures that, for \( T \) large enough, the first derivatives of \( L(\theta) \) are "well-behaved" at \( \theta_0 \).

To consider the asymptotic distribution of the LIM-GARCH estimates, first, define \( A_0 \) and \( B_0 \) to be the information matrix in Hessian and outer product form, respectively.
(3.9) \[ A_\circ = - E \left[ \frac{\partial^2 L}{\partial \theta \partial \theta'} \right] , \]

(3.10) \[ B_\circ = E \left[ T \frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \theta'} \right] . \]

As will be shown, both \( A_\circ \) and \( B_\circ \) appear in the variance-covariance matrix of the asymptotic distribution of the LIM-GARCH estimates and are therefore required to be invertible.

**THEOREM III: (Asymptotic Normality)**

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

(3.11) \[ B_\circ^{-1/2} A_\circ T^{1/2} (\hat{\theta} - \theta_\circ) \sim N(0, I) \]

Furthermore, consistent estimates of \( A_\circ \) and \( B_\circ \) are given by

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

and

(3.13) \[ \hat{B} = T^{-1} \sum_{t=1}^{T} \frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \theta'} \]

with all derivatives evaluated at \( \theta = \theta_\circ \).
Proof: See Appendix.

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

\[
(3.14) \quad \text{Asymptotic Var-Cov}(\hat{\theta}) = A_0^{-1} B_0 A_0^{-1},
\]

since the conditional log-likelihood function may not be correctly specified. When the conditional distribution of \( \epsilon_1^0 \) is normal then \( A_0 = B_0 \) and the standard form for the asymptotic variance-covariance matrix is appropriate, that is

\[
(3.15) \quad \text{Asymptotic Var-Cov}(\hat{\theta}) = A_0^{-1}
\]

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 $h^*_t$ consisting of all the unique elements of the conditional variance-covariance matrix. In matrix notation, this can be written as

$$
(4.1) \quad h^*_t = \omega + \Lambda \eta_{t-1} + \mathbf{B} h^*_{t-1}
$$

where

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

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

FIM-GARCH estimation of the diagonal SEM-GARCH(1,1) model requires maximizing the conditional log-likelihood function
\[(4.4) \quad L - \sum_{t-1}^{T} \mathbb{I}_t\]

where

\[(4.5) \quad \mathbb{I}_t(Y_t | I_{t-1}) = -\frac{M}{2} \log(2\pi) + \log ||\Gamma|| - \frac{1}{2} \log |H_t|\]

\[-\frac{1}{2} \text{Tr} \left[ (Y_t \Gamma + X_t \beta)' (Y_t \Gamma + X_t \beta) H_t^{-1} \right]\]

where $|H_t|$ denotes the determinant of $H_t$, and $||\Gamma||$ is the absolute value of the determinant of $\Gamma$. 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 $\Gamma$, $\beta$, $\omega$, $A$, 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 $\Gamma$ and $\beta$ are a function of past derivatives of $\epsilon_t$ 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, $\Gamma^\mu$. The superscript $\mu$ 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.

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

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

This can be simplified to yield

\[
\frac{\partial L}{\partial \Gamma^\mu} = (\Gamma^{-1})' - \frac{1}{2} (H^{-1}) \left[ \frac{\partial H}{\partial \Gamma^\mu} T \frac{\partial (YT+X\beta)'(YT+X\beta)}{H} \right] - \frac{\varepsilon' \varepsilon}{H} = 0.
\]

But the partial derivative of the conditional variance-covariance matrix, \( H \), with respect to \( \Gamma \) 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 \( \varepsilon_{1t} \) 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, \( \theta \). Let \( \theta^i \) denote the parameter estimates after the \( i \)th iteration. \( \theta^{i+1} \) is then calculated from,

\[
\theta^{i+1} = \theta^i + \lambda_i \left[ \sum_{t=1}^T \frac{\partial \ell}{\partial \theta^i} \frac{\partial \ell}{\partial \theta^i} \right]^{-1} \sum_{t=1}^T \frac{\partial \ell}{\partial \theta^i},
\]

where \( \lambda_i \) 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. **ASYMPTOTIC PROPERTIES OF THE FIM-GARCH ESTIMATOR**

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 \( \theta \) now be a \( S \)-element column vector.
\[(5.1) \quad \theta' = (m', v') \quad \text{with} \quad S = M + K + \sum_{i=1}^{M} (1+p+q) \Sigma_i \]

where \( \theta \), as in Section III, is partitioned such that \( m \) is the parameter vector corresponding to the unrestricted parameters of all the structural equations of the system, that is

\[(5.2) \quad m = \left[ \text{vec}(\Gamma, \beta) \right]^{\mu}, \]

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

\[(5.3) \quad v = \left[ \text{vec}(\omega, A, B) \right]^{\mu}. \]

As before, I assume \( \theta \in \Xi \), where \( \Xi \) is a compact subspace of Euclidean space, and let \( \hat{\theta}_o \) represent the true parameter vector. Theorems IV and V are the FIM-GARCH equivalent 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 \( \theta_o \) provided \( \theta_o \) is interior to \( \Xi \).
Proof: See Appendix.

**THEOREM V:** (Asymptotic Normality)

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

\[
B_o^{-1/2} \hat{A}_o T^{1/2} (\hat{\theta} - \theta_o) \sim N(0, I)
\]

where consistent estimates of \( \hat{A}_o \) and \( \hat{B}_o \) are given by

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

and

\[
\hat{B} = T^{-1} \sum_{t=1}^{T} \frac{\partial \epsilon_t}{\partial \theta} \frac{\partial \epsilon_t}{\partial \theta'}
\]

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

\[
(5.7) \quad \text{Asymptotic Var-Cov}(\hat{\theta}) = \hat{A}_o^{-1} \hat{B}_o \hat{A}_o^{-1}
\]
since the conditional log-likelihood function may not be correctly specified. When the conditional distribution of $\epsilon_t$ is truly normal then $A_o = B_o$ and the standard form for the asymptotic variance-covariance matrix is appropriate, that is

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

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. THE INFORMATION MATRIX TEST

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 $\epsilon_t^o$ 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

$$\text{(6.1) \quad A(\theta_o) - B(\theta_o) = 0}$$

Thus a test for the null hypothesis of conditional normality and
correct model specification is based on the difference between consistent estimates of $\mathbf{A}_0$ and $\mathbf{B}_0$, denoted $\hat{\mathbf{A}}_T$ and $\hat{\mathbf{B}}_T$, respectively. For the LIM-GARCH estimator $\theta$ 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 $k$th element

$$d_{tk}(\theta) = \left[ \frac{\partial \ell_t}{\partial \theta_i} \frac{\partial \ell_t}{\partial \theta_j} + \frac{\partial^2 \ell_t}{\partial \theta_i \partial \theta_j} \right]_{i,j = 1, \ldots, s}^{k = 1, \ldots, q}$$

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

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

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

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

$V(\theta)$ turns out to be the asymptotic variance-covariance matrix of $T^{1/2}D_T(\theta)$. 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 $\epsilon_{1t}$ is not normally distributed it is necessary to assume that
(6.5) \[ E \left[ \frac{8}{\sigma_1^2} \left( \frac{h_{1t}^2}{h_{1t}^2} \right) I_{t-1} \right] < \infty \] for all \( \theta \in \mathbb{E} \).

Based on these conditions, White proves that

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

and

(6.7) \[ V_T(\hat{\theta}) \overset{a.s.}{\Rightarrow} V(\theta_0), \]

where \( V_T(\hat{\theta}) \) is the estimate of \( V(\theta_0) \). Then it follows that the information test statistic, \( T_T \), is given by

(6.8) \[ T_T = T \frac{T}{D_T(\hat{\theta})} \left[ V_T(\hat{\theta}) \right]^{-1} D_T(\hat{\theta}) \overset{A}{\Rightarrow} \chi^2_q. \]

To carry out the Information Matrix Test, one computes \( T_T \) and compares it to the critical value of the \( \chi^2_q \) 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 \( A_0^{-1} \) 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 desirable 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 \textsc{Sem-Garch}(p,q) model to a dynamic framework by including lagged endogenous variables. In Harmon (1988) I extend the \textsc{Sem-Garch} model by incorporating the conditional variances and covariances as variables in the structural equations themselves. This is referred to as the \textsc{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


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


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

\[
(A.1) \quad \frac{\partial L}{\partial a_o} \bigg|_{a_o = a_o^*} = \sum_{t=1}^{T} \left[ -\frac{a_o^*W_{11,t}}{a_o^*W_{11,t}^2} - \frac{1}{2h_{1t}^2} \frac{\partial h_{1t}^2}{\partial a_o} \right] - \frac{1}{2} h_{1t}^2 \left( \frac{\partial \epsilon_{1t} \epsilon_{1t}^*}{\partial a_o} - \frac{\epsilon_{1t} \epsilon_{1t}^* h_{1t}^2}{\epsilon_{1t} \epsilon_{1t}^* h_{1t}^2} \right) = 0.
\]

When simplified this can be written as

\[
(A.2) \quad \frac{\partial L}{\partial a_o} \bigg|_{a_o = a_o^*} = \sum_{t=1}^{T} \left[ -\frac{a_o^*W_{11,t}}{a_o^*W_{11,t}^2} - \frac{\epsilon_{1t} Y_{1t}}{2h_{1t}^2} \right]
+ \frac{1}{2} \frac{\partial h_{1t}^2}{\partial a_o} \left( \frac{\epsilon_{1t} \epsilon_{1t}^*}{h_{1t}^2} - 1 \right) = 0,
\]

where
\[ \frac{\partial^2 h_{lt}}{\partial a_o^2} = - \sum_{i=1}^{q} \alpha_i \frac{\partial \epsilon_{i-1,t-1}}{\partial a_o} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{i-1,t-1}}{\partial a_o} \]

\[ = - 2 \sum_{i=1}^{q} \alpha_i \epsilon_{i-1,t-1} x_{i-1,t-1} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{i-1,t-1}}{\partial a_o} \]

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

\[ \frac{\partial L}{\partial b_o} \bigg|_{b_o = \hat{b}_o} = - \frac{T}{2} \sum_{t=1}^{T} - \frac{\epsilon_{i-1,t-1}}{2 h_{i-1,t}} + \frac{1}{2 h_{i-1,t}} \frac{\partial h_{i-1,t}}{\partial b_o} \left[ \frac{\epsilon_{i-1,t-1}}{h_{i-1,t}} - 1 \right] \]

where

\[ \frac{\partial h_{i-1,t}}{\partial b_o} = - \sum_{i=1}^{q} \alpha_i \frac{\partial \epsilon_{i-1,t-1}}{\partial b_o} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{i-1,t-1}}{\partial b_o} \]

\[ = - 2 \sum_{i=1}^{q} \alpha_i \epsilon_{i-1,t-1} x_{i-1,t-1} + \sum_{i=1}^{p} \delta_i \frac{\partial h_{i-1,t-1}}{\partial b_o} \]

Now, differentiating with respect to the parameters \( \omega_o, \alpha_j, \) and \( \delta_j \) that comprise the conditional variance, \( h_{i-1,t} \), yields

\[ \frac{\partial l}{\partial \omega_o} \bigg|_{\omega_o = \hat{\omega}_o} = - \frac{T}{2} \sum_{t=1}^{T} - \frac{1}{2 h_{i-1,t}} \frac{\partial h_{i-1,t}}{\partial \omega_o} \left[ \frac{\epsilon_{i-1,t-1}}{h_{i-1,t}} - 1 \right] \]

\[ = 0. \]
but notice that

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

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 \(\alpha_j\) yields

\[(A.8) \quad \frac{\partial L}{\partial \alpha_j} = \sum_{t=1}^{T} \left[ \frac{1}{2h_{1t}^2} \left[ \frac{\partial h_{1t}^2}{\partial \alpha_j} \right] \left[ e_{1t}^2 e_{1t}^2 - 1 \right] \right] = 0, \]

where

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

Then, differentiating with respect to \(\delta_j\) yields the following first order condition

\[(A.10) \quad \frac{\partial L}{\partial \delta_j} = \sum_{t=1}^{T} \left[ \frac{1}{2h_{1t}^2} \left[ \frac{\partial h_{1t}^2}{\partial \delta_j} \right] \left[ e_{1t}^2 e_{1t}^2 - 1 \right] \right] = 0, \]

where in this case
\[
\frac{\partial h_{lt}^2}{\partial \delta_{j}} = \left[ h_{l1,t-j}^2 + \sum_{i=1}^{q} \alpha_{i} \frac{\partial \epsilon_{1,t-i}}{\partial \delta_{j}} \right].
\]

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, \( h_{lt}^2 \).

\[
\frac{\partial L}{\partial h_{lt}^2} = \frac{T}{\Sigma} \left[ \frac{1}{2h_{lt}^2} \left( \epsilon_{lt}' \epsilon_{lt} - 1 \right) \right] = 0.
\]

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

**Proof of Lemma I:**

Expressions for \( \frac{\partial \epsilon_{lt}}{\partial m}, \) where \( m' = (a'_o, b'_o) \), are given by

\[
\frac{\partial \epsilon_{lt}}{\partial a'_o} = Y_{1t},
\]

\[
\frac{\partial \epsilon_{lt}}{\partial b'_o} = X_{1t}.
\]

Then for bounded constant vectors \( \lambda < \infty \), one has
(A.15) \[ \lambda' \frac{\partial \epsilon_{lt}}{\partial m} = \lambda' Z_{lt}, \]

where \( Z_{lt} = [Y_{lt}, X_{lt}] \). Now, one can write

(A.16) \[ E \left[ \lambda' \frac{\partial \epsilon_{lt}}{\partial m} \frac{\partial \epsilon_{lt}}{\partial m'} \lambda \right] = E \left[ \lambda' Z_{lt} Z_{lt}' \lambda \right] \]

where \( Z_{lt} Z_{lt}' \) is the cross product matrix and the right hand side is thus a scalar. Hence there exists a constant \( M_1 < \infty \) such that

(A.17) \[ E \left[ \lambda' \frac{\partial \epsilon_{lt}}{\partial m} \frac{\partial \epsilon_{lt}}{\partial m'} \lambda \right] < M_1 \quad \text{for all } \theta \in \Xi. \]

Next, to show that

(A.18) \[ \det E \left[ \frac{\partial \epsilon_{lt}}{\partial m} \frac{\partial \epsilon_{lt}}{\partial m'} \right] > 0 \]

is equivalent to showing that

(A.19) \[ E \left[ \lambda' \frac{\partial \epsilon_{lt}}{\partial m} \frac{\partial \epsilon_{lt}}{\partial m'} \lambda \right] > 0 \quad \text{for all } \lambda \neq 0. \]

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

(A.20) \[ \lambda' \frac{\partial \epsilon_{lt}}{\partial m} = 0 \quad \text{a.s. for all } t. \]
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

\[
(A.21) \quad \hat{Y}_{1t} = \mathbb{E}[Y_{1t}|I_{t-1}].
\]

Then specify $Y_{1t}$ to be

\[
(A.22) \quad Y_{1t} = \epsilon^0_{1t} + \hat{Y}_{1t},
\]

where $\epsilon^0_{1t}$ denotes the true errors. Furthermore, using a mean value expansion one can specify $\epsilon_{1t}$ to be

\[
(A.23) \quad \epsilon_{1t} = \epsilon^0_{1t} + \frac{\partial \epsilon_{1t}}{\partial m'}(m - m_o)
\]

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

\[
(A.24) \quad \mathbb{E}[\epsilon^0_{1t}|I_{t-1}, Z_{1t}] = \epsilon^0_{1t}.
\]

But the LIM-GARCH specification assumes that

\[
(A.25) \quad \mathbb{E}[\epsilon_{1t}|I_{t-1}] = 0
\]

and therefore it must be the case that
(A.26) \[ \mathbb{E}[\epsilon^0_{1t} I_{t-1}, Z_{1t}] = 0. \]

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

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

as required. \( \Box \)

**Proof of Lemma II:**

Expressions for \( \frac{\partial^2 h_{1t}}{\partial \nu} \), where \( \nu' = (\omega_0, \alpha', \delta') \), are given by

(A.28) \[ \frac{\partial^2 h_{1t}}{\partial \omega_0} = 1, \]

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

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

For constant vectors \( \lambda < \infty \), write (A.28), (A.29), and (A.30) as
(A.31) \[ \lambda' \frac{\partial h^2_{lt}}{\partial v} = \lambda' W_t \]

where

(A.32) \[ W'_t = \begin{bmatrix} 1, & \epsilon^2_{1,t-1}, & \ldots, & \epsilon^2_{1,t-q}, & \frac{\partial h^2_{1,1,t}}{\partial \delta_1}, & \ldots, & \frac{\partial h^2_{1,t-p}}{\partial \delta_p} \end{bmatrix} \]

Since by assumption \( E[\epsilon^4_{1t}] < \infty \) then \( E[\epsilon^4_{1t}] < M_2 < \infty \), 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

(A.33) \[ \text{det } E \begin{bmatrix} \frac{\partial h^2_{lt}}{\partial v} & \frac{\partial h^2_{lt}}{\partial v'} \\ \frac{\partial h^2_{lt}}{\partial v} & \frac{\partial h^2_{lt}}{\partial v'} \end{bmatrix} > 0 \]

is equivalent to showing that

(A.34) \[ E \begin{bmatrix} \lambda' \frac{\partial h^2_{lt}}{\partial v} & \frac{\partial h^2_{lt}}{\partial v'} \end{bmatrix} > 0 \text{ for all } \lambda \neq 0. \]

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 \( \epsilon^2_{1t} \)

(A.35) \[ \epsilon^2_{1t} = (\epsilon^o_{1t})^2 + 2 \epsilon^o_{1t} \frac{\partial \epsilon_{1t}}{\partial v} (v - v_o) + (v - v_o)' \begin{bmatrix} \frac{\partial \epsilon_{1t}}{\partial v} & \frac{\partial \epsilon_{1t}}{\partial v'} \end{bmatrix} (v - v_o) \]
This is a quadratic function in $\epsilon^o_{lt}$ which yields two solutions,

$$\text{(A.36)} \quad \epsilon^o_{lt} = f_1(t) \text{ or } f_2(t).$$

These solutions are functions of $I_{t-1}$ and $X_{lt}$. But $\epsilon^o_{lt}$ having two values conditional on $I_{t-1}$ and $X_{lt}$ is not permitted. Therefore, no such $\lambda$ exists and (A.34) holds. □

**Proof of LEMMA III:**

Differentiating the conditional log-likelihood function with respect to $\theta$ yields

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

where the derivatives with respect to $\theta_i$ are evaluated at their true values $\theta_o$. Second order derivatives can be shown to have the following general form

$$\text{(A.38)} \quad \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = \frac{T}{\Sigma_{t=1}} \left[ - \frac{\epsilon_{lt}}{h_{lt}^2} \left[ \frac{\partial^2 \epsilon_{lt}}{\partial \theta_i \partial \theta_j} \right] - \frac{1}{h_{lt}^2} \frac{\partial \epsilon_{lt}}{\partial \theta_i} \left[ \frac{\partial \epsilon_{lt}}{\partial \theta_j} \right] \right]$$
\[
\epsilon_{lt} \left[ \frac{\partial \epsilon_{lt}}{\partial \theta_i} \right] \left[ \frac{\partial h^2_{lt}}{\partial \theta_j} \right] + \frac{1}{2h^4_{lt}} \left[ \frac{\epsilon_{lt}^2}{h^2_{lt}} - 1 \right] \left[ \frac{\partial h^2_{lt}}{\partial \theta_i \partial \theta_j} \right] \\
+ \epsilon_{lt} \left[ \frac{\partial h^2_{lt}}{\partial \theta_i} \right] \left[ \frac{\partial \epsilon_{lt}}{\partial \theta_j} \right] - \frac{1}{2h^4_{lt}} \left[ \frac{\epsilon_{lt}}{h^2_{lt}} \right] \left[ \frac{\partial h^2_{lt}}{\partial \theta_i \partial \theta_j} \right] \\
+ \frac{1}{2h^4_{lt}} \left[ \frac{\partial h^2_{lt}}{\partial \theta_i} \right] \left[ \frac{\partial h^2_{lt}}{\partial \theta_j} \right]
\]

Expressions for \( \frac{\partial \epsilon_{lt}}{\partial \theta_i} \) and \( \frac{\partial h^2_{lt}}{\partial \nu_i} \) are given by (A.15) and (A.31), respectively. For \( \frac{\partial h^2_{lt}}{\partial \theta_j} \) one has

\[
(A.39) \quad \frac{\partial h^2_{lt}}{\partial \theta_j} = 2 \sum_{i=1}^{q} a_i \epsilon_{1,i,t-i} \frac{\partial \epsilon_{1,t-i}}{\partial \theta_j} + \sum_{i=1}^{p} \delta_i \frac{\partial h^2_{1,i,t-i}}{\partial \theta_j}
\]

where \( m_j \) is the \( j \)th element of \( m \). From Lemma II and (A.39) it is easy to see that \( \frac{\partial h^2_{lt}}{\partial \theta_i} \) have bounded second moments.

Since every term in \( \frac{\partial h^2_{lt}}{\partial \theta_i} \) and \( \frac{\partial h^2_{lt}}{\partial \theta_i \partial \theta_j} \) also appears in \( h^2_{lt} \) itself, the expressions

\[
(A.40) \quad \frac{1}{h^2_{lt}} \left[ \frac{\partial h^2_{lt}}{\partial \theta_i} \right] \quad \text{and} \quad \frac{1}{h^2_{lt}} \left[ \frac{\partial h^2_{lt}}{\partial \theta_i \partial \theta_j} \right]
\]

are uniformly bounded from above. Hence, evaluating equation (A.29) at the true parameter values \( \theta_o \) implies that the first, third, fourth, and fifth terms are zero, with the sixth term being
\begin{align}
(A.41) \quad & E \left[ \frac{1}{h_{1t}} \left[ \frac{\partial^2 \epsilon_{1t}}{\partial \theta_i \partial \theta_j} \right] \right] I_{t-1} - E \left[ \frac{\partial h_{1t}}{\partial \theta_i} \left[ \frac{\partial h_{1t}}{\partial \theta_j} \right] I_{t-1} \right] < \infty \\
& \text{since } E(\epsilon_{1t}^2 I_{t-1}) = h_{1t}^2. \text{ Similarly, the second term of (A.38), evaluated at } \theta_o, \text{ is bounded} \\
(A.42) \quad & E \left[ \frac{1}{h_{1t}^2} \left[ \frac{\partial \epsilon_{1t}}{\partial \theta_i} \right] \right] I_{t-1} < \infty.
\end{align}

Therefore, the matrix A is defined to be

\begin{align}
(A.43) \quad & A = -E \left[ \frac{\partial^2 h_{1t}}{\partial \theta_o \partial \theta_o'} I_{t-1} \right], \\
& - \frac{1}{2} E \left[ \frac{1}{h_{1t}^4} \left[ \frac{\partial h_{1t}^2}{\partial \theta} \right] \right] I_{t-1} + E \left[ \frac{1}{h_{1t}^2} \left[ \frac{\partial \epsilon_{1t}}{\partial \theta} \right] \right] I_{t-1} < \infty
\end{align}

where the derivatives are evaluated at \( \theta_o \).

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

\begin{align}
(A.44) \quad & \lambda' A \lambda = \frac{1}{2} E \left[ \lambda' \frac{1}{h_{1t}^4} \left[ \frac{\partial h_{1t}^2}{\partial \theta} \right] \right] I_{t-1} \\
& + E \left[ \lambda' \frac{1}{h_{1t}^2} \left[ \frac{\partial \epsilon_{1t}}{\partial \theta} \right] \right] I_{t-1}
\end{align}
Both terms on the right hand side are nonnegative. What remains to be shown is that these are both greater than zero. First, partition \( \lambda' = (\lambda_1', \lambda_2') \) to conform with \( \theta' = (m', v') \). Now, since \( \partial \varepsilon_{1t} / \partial \theta = 0 \) the second term in (A.44) is equal to

\[
E \left[ \left( \lambda' \frac{1}{2} \left[ \frac{\partial \varepsilon_{1t}}{\partial \theta} \right] \right) \left[ \frac{\partial \varepsilon_{1t}}{\partial \theta'} \right] \lambda \right]_{I_{t-1}} = E \left[ \left( \lambda' \frac{1}{2} \left[ \frac{\partial \varepsilon_{1t}}{\partial \theta} \right] \right) \left[ \frac{\partial \varepsilon_{1t}}{\partial \theta'} \right] \lambda_2 \right]_{I_{t-1}}
\]

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

\[
E \left[ \left( \lambda' \frac{1}{4} \left[ \frac{\partial \theta^2}{\partial \theta} \right] \right) \left[ \frac{\partial \theta^2}{\partial \theta'} \right] \lambda \right]_{I_{t-1}} = E \left[ \left( \lambda' \frac{1}{4} \left[ \frac{\partial \theta^2}{\partial \theta} \right] \right) \left[ \frac{\partial \theta^2}{\partial \theta'} \right] \lambda_1 \right]_{I_{t-1}}
\]

which is positive because of Lemma II.

proof of Theorem I:

From the ergodic theorem, for any \( \theta \in \Xi \),

\[
L(\theta) = \lim_{t \to \infty} \ell_t
\]

\[
= E(C_1) - \frac{1}{2} E \left[ \log h_{1t} \right] - \frac{1}{2} E \left[ \log(a_0 \tilde{W}_{11}, t^a) \right]
\]

\[
- \frac{1}{2} E \left[ \varepsilon_{1t}^2 / h_{1t}^2 \right] \quad \text{a.s.}
\]
if the expectations exist. From Jensen's Inequality it follows that

\[(A.48) \quad \log \mathbb{E}(X) \geq \mathbb{E}[\log(x)] , \]

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

\[(A.49) \quad \log \mathbb{E}[h_{1t}^2] \geq \mathbb{E}[\log h_{1t}^2] . \]

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

\[(A.50) \quad \mathbb{E}[\varepsilon_{1t}^2/h_{1t}^2] \leq \mathbb{E}[\varepsilon_{1t}^2] , \]

then clearly

\[(A.51) \quad \mathbb{E}[\varepsilon_{1t}^2/h_{1t}^2] < \infty . \]

Since \(W_{11t}\) 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,

\[(A.52) \quad \mathbb{E}[\varepsilon_{1t}^2/h_{1t}^2] = \mathbb{E}[h_{1t}^{-2}(\varepsilon_{1t} + \varepsilon_{1t}^0 - \varepsilon_{1t}^0)^2] \]

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

\[
(A.53) \quad \operatorname{E} \left[ \frac{\epsilon_{1t}^2}{h_{1t}^2} \right] > \operatorname{E} \left[ \frac{(\epsilon_{1t}^o)^2}{h_{1t}^2} \right].
\]

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

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

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 \( m = m_o \).

\[
(A.55) \quad L(\theta) = \operatorname{E}(C_1) - \frac{1}{2} \operatorname{E} \left[ \log h_{1t}^2 \right] - \frac{1}{2} \operatorname{E} \left[ \log (a_o' W_{11}, t a_o) \right]
\]

\[
- \frac{1}{2} \log \left[ \operatorname{E} \left[ \frac{\epsilon_{1t}^2}{h_{1t}^2} \right] \right],
\]

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

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

\[
\left[ \text{since } \operatorname{E} \left[ (\epsilon_{1t}^o)^2 h_{1t}^{-2} \right] = \operatorname{E} \left[ (\epsilon_{1t}^o)^2 h_{1t}^{-2} (h_{1t}^o)^2 (h_{1t}^o)^{-2} \right] \right]
\]
\[ - \text{E}(C_1) - \frac{1}{2} \text{E} \left[ \log h_{1t}^2 \right] - \frac{1}{2} \text{E} \left[ \log(a_o^o W_{11}, t a_o) \right] \\
- \frac{1}{2} \log \left[ \text{E} \left[ (\epsilon_{1t}^o)^2 (h_{1t}^o)^{-2} \right] \right] - \frac{1}{2} \log \left[ \text{E} \left[ h_{1t}^{-2} (h_{1t}^o)^2 \right] \right], \]

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

\[ \leq \text{E}(C_1) - \frac{1}{2} \text{E} \left[ \log h_{1t}^2 \right] - \frac{1}{2} \text{E} \left[ \log(a_o^o W_{11}, t a_o) \right] \\
- \frac{1}{2} \text{E} \left[ \log \left[ (\epsilon_{1t}^o)^2 (h_{1t}^o)^{-2} \right] \right] - \frac{1}{2} \text{E} \left[ \log \left[ h_{1t}^{-2} (h_{1t}^o)^2 \right] \right], \]

\[ = \text{E}(C_1) - \frac{1}{2} \text{E} \left[ \log(h_{1t}^o)^2 \right] - \frac{1}{2} \text{E} \left[ \log(a_o^o W_{11}, t a_o) \right] \\
- \frac{1}{2} \text{E} \left[ \log \left[ (\epsilon_{1t}^o)^2 (h_{1t}^o)^{-2} \right] \right], \]

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

\[ = L(\theta_o) \]

with equality only at \( \theta = \theta_o \). Note that

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

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

\[
\frac{dL}{\theta} = \frac{dL}{\theta_0} + \frac{d^2L}{\theta_0 \theta_0'} (\hat{\theta} - \theta_0) + \frac{1}{2} \frac{d^3L}{\theta_0 \theta_0' \theta_0''} (\hat{\theta} - \theta_0)
\]

where \( |\hat{\theta} - \theta_0| > |\bar{\theta} - \theta_0| \), which requires that \( \bar{\theta} \) lie between \( \theta_0 \) 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

\[
\frac{dL}{\theta} = \sum_{t=1}^{T} \frac{dL_t}{\theta_0} + (\hat{\theta} - \theta_0) \sum_{t=1}^{T} \frac{d^2L_t}{\theta_0 \theta_0'} + (\hat{\theta} - \theta_0) \sum_{t=1}^{T} \frac{d^3L_t}{\theta_0 \theta_0' \theta_0''}
\]

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

\[
\frac{dL}{\theta} = 0
\]

These sufficient conditions consists of

\[
1 \sum_{t=1}^{T} \frac{dL_t}{\theta_0} \xrightarrow{p} 0
\]
(II) There exists a nonrandom matrix \( M(\theta_0) > 0 \) such that for all \( \epsilon > 0 \),

\[
\Pr \left[ -\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial \theta_0 \partial \theta_0'} \geq M(\theta_0) \right] > 1 - \epsilon
\]

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

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

\[
E \left[ \frac{\partial^3 l_t}{\partial \theta_1 \partial \theta_j \partial \theta_k} \right] < M \quad \text{for all } \theta \in \Theta.
\]

Then it follows that \( \hat{\theta} \overset{P}{\rightarrow} \theta_0 \).

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 \( \alpha_0 \).

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

\[
(A.50) \quad \frac{\partial L}{\partial \alpha_0} \bigg|_{\alpha_0 = \hat{\alpha}_0} = \sum_{t=1}^{T} - \frac{\alpha_0'W_{11,t} - \epsilon_{1t}'Y_{1t}}{2h_{1t}^2} - \frac{\epsilon_{1t}'\epsilon_{1t}}{h_{1t}^2} \frac{\partial h_{1t}^2}{\partial \alpha_0}
\]

\[
+ \frac{1}{2h_{1t}^2} \left[ \frac{\partial h_{1t}^2}{\partial \alpha_0} \left( \frac{\epsilon_{1t}'\epsilon_{1t}}{h_{1t}^2} - 1 \right) \right] = 0.
\]
When this derivative is evaluated at the true $a_o$, the first term is zero since $W_{1t}$ 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

\[
(A.61) \quad E \left[ \frac{\partial L}{\partial a_o} \right]_{I_{t-1}} = 0
\]

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

\[
(A.62) \quad \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \ell_t}{\partial a_o} \rightarrow 0.
\]

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

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

\[= -\lambda' A_o \lambda \]

where $A_o$ is defined to be

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

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

\[
(A.65) \quad \frac{\partial^2 L}{\partial a_o \partial a_o'} = \frac{T}{t-1} \left[ \frac{W_{11,t}}{a_o' W_{11,t} a_o} - \frac{a_o' W_{11,t} a_o}{(a_o' W_{11,t} a_o)^2} - \frac{Y_{1t}' Y_{1t}}{2h_{1t}^2} \right]
\]

\[+ \alpha_1 \left[ \frac{\epsilon_{lt}' Y_{1t}}{2h_{1t}^2} \right] T^{-1} \sum_{i=1}^{\delta_{I_t}'} Y_{1t-i} \epsilon_{lt-i}' \]

\[+ \frac{\alpha_1}{h_{1t}^2} \left[ \frac{\epsilon_{lt}' \epsilon_{lt}}{h_{1t}^2} - 1 \right] T^{-1} \sum_{i=1}^{\delta_{I_t}'} \epsilon_{1t-i}' Y_{1t-i} Y_{1t-i}' \]

\[+ \frac{T^{-1}}{\sum_{i=1}^{\delta_{I_t}'}} \delta_{I_t} Y_{1t-i}' Y_{1t-i}' \left[ \frac{\alpha_1}{h_{1t}^4} \left[ Y_{1t} \epsilon_{1t}' + 2 \left[ 1 - \frac{\epsilon_{1t}' \epsilon_{1t}}{h_{1t}^2} \right] \right] \right] \]

Taking negative expectations of (A.65) yields

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

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

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

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

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

(A.69) \[ M(a_o) = -\frac{1}{2} E \left[ \frac{\partial^2 L_t}{\partial a_o \partial a'} \right] \]

Then

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

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 \( \epsilon_{1t}^2 \). 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 \( \frac{\partial^3 h_{1t}^2}{\partial a_o^3} \) are bounded and that terms containing \( \frac{\partial^3 \epsilon_{1t}}{\partial a_o^3} \) have
bounded second moments. □

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

Proof of THEOREM III:

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

\[ (I) \quad T^{-1/2} \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta_o} \rightarrow A \times N(0, B_o) \quad \text{for nonrandom } B_o > 0. \]

\[ (II) \quad T^{-1} \sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial \theta_o \partial \theta'_o} \xrightarrow{p} -A_o \quad \text{for nonrandom } A_o > 0. \]

(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 \( B_o > 0 \). Then from Condition (I) of the proof of Theorem I, it follows that

\[ (A.71) \quad E \left[ \frac{\partial l_t}{\partial \theta_o} \right]_{I_t=1} = 0. \]

Weiss shows that if
\begin{align}
\text{(A.72)} \quad & E \left[ \frac{\partial \ell_t}{\partial \theta_o} \frac{\partial \ell_t}{\partial \theta_o'} \right] < \infty \\
\text{and} \\
\text{(A.73)} \quad & E \left[ T \frac{\partial L}{\partial \theta_o} \frac{\partial L}{\partial \theta_o'} \right] = B_o < \infty,
\end{align}

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

\begin{align}
\text{(A.74)} \quad & 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] \\
& \quad - \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] \\
& \quad - 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] \left[ \frac{\epsilon_{1t}^2}{h_{1t}^2} - 1 \right] \\
& \quad \left[ \frac{\epsilon_{1t}^2}{h_{1t}^4} - 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]
\end{align}

where the derivatives are evaluated at \( \theta_o \). Taking expectations yields
(A.75) \[ E \left[ \frac{\partial \ell_t}{\partial \theta} \frac{\partial \ell_t}{\partial \theta'} \right] = E \left[ \frac{1}{4} \frac{\partial \epsilon_{1t}^2}{\partial \theta} \frac{\partial \epsilon_{1t}^2}{\partial \theta'} \right], \]

which from Lemma III is clearly bounded.

Condition (II): Since by definition

(A.76) \[ A_o = -E \left[ \frac{\partial^2 \ell_t}{\partial \theta \partial \theta'} \bigg|_{I_{t-1}} \right] \]

then Lemma III verifies condition II. \( \Box \)

Proof of THEOREM IV:

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, \( \Gamma \).

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

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

Clearly, since \( E(\epsilon_{1t}^2 | I_{t-1}) = 0 \) and \( E(\epsilon_{1t}^2 | I_{t-1}) = h_{1t}^2 \),
(A.78) \[ E \left[ \frac{\partial L}{\partial \Gamma^\mu} \bigg| I_{t-1} \right] = 0 \]

Then, the ergodic theorem implies that

\[ \frac{1}{T} \sum_{t=1}^{T} \frac{\partial L}{\partial \Gamma^\mu} \overset{P}{\longrightarrow} 0. \]

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

\[ \frac{1}{T} \sum_{t=1}^{T} \lambda' \left[ \frac{\partial^2 L}{\partial \Gamma^\mu \partial \Gamma^{\mu'}} \right] \lambda \rightarrow E \lambda' \left[ \frac{\partial^2 L}{\partial \Gamma^\mu \partial \Gamma^{\mu'}} \right] \lambda = -\lambda' A_0 \lambda \quad \text{a.s.} \]

where \( A_0 \) is defined to be

\[ (A.81) \quad A_0 = -E \left[ \frac{\partial^2 L}{\partial \Gamma^\mu \partial \Gamma^{\mu'}} \right] \]

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

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

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

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

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

As in Theorem II, for a given \( \lambda \), let \( \delta \) be bounded by

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

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

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

for all \( T > T_1 \).

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

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

Then

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

for all \( T > T_1 \).

Condition (III): See Theorem II, condition III. \( \square \)
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, \( \Gamma \).

**Condition I:** Since it is assumed that \( B_0 > 0 \), then from Condition I from the proof of Theorem IV, it follows that

\[
(A.90) \quad E \left[ \frac{\partial l}{\partial \Gamma^\mu} \right] \bigg|_{t-1} = 0
\]

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

\[
(A.91) \quad E \left[ \frac{\partial l_t}{\partial \Gamma^\mu} \frac{\partial l_t}{\partial \Gamma^\mu'} \right] < \infty
\]

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

\[
(A.92) \quad E \left[ \frac{\partial l_t}{\partial \Gamma^\mu} \frac{\partial l_t}{\partial \Gamma^\mu'} \right] = E \left[ (\Gamma^{-1})' + \left[ \frac{\partial H_t}{\partial \Gamma^\mu} \left[ \frac{\epsilon_t' \epsilon_t}{H_t} - 1 \right] - \frac{\epsilon_t' Y_t}{H_t} \right] \right]
\]

\[
\left[ (\Gamma^{-1})' + \left[ \frac{\partial H_t}{\partial \Gamma^\mu} \left[ \frac{\epsilon_t' \epsilon_t}{H_t} - 1 \right] - \frac{\epsilon_t' Y_t}{H_t} \right] \right]
\]
where the derivatives are evaluated at the true $\Gamma^\mu$. Taking expectations yields

\[
E \left[ \left( \Gamma^{-1} \right) \left( \Gamma^{-1} \right) ' + \frac{Y_t^t \epsilon_t^t \epsilon_t^t Y_t}{H_t^t H_t^t} - \frac{1}{4} \left[ \Gamma^{-1} \frac{\partial H_t}{\partial \Gamma^\mu} \right] \left[ \epsilon_t^t \epsilon_t^t - 1 \right] \right] \\
+ \left[ \epsilon_t^t \epsilon_t^t - 1 \right] \left[ \Gamma^{-1} \frac{\partial H_t}{\partial \Gamma^\mu} \right] + \left( \Gamma^{-1} \right) \left[ \Gamma^{-1} \frac{\partial H_t}{\partial \Gamma^\mu} \right] \left[ \epsilon_t^t \epsilon_t^t - 1 \right] \\
- 2 \left( \Gamma^{-1} \right) \left[ \epsilon_t^t Y_t \right] \frac{H_t^t}{H_t^t} 
\]

Condition II: By definition

\[(A.93) \quad E \left[ \frac{\partial \phi_t}{\partial \Gamma^\mu} \frac{\partial \phi_t}{\partial \Gamma^\mu'} \right] = E \left[ \left( \Gamma^{-1} \right) \left( \Gamma^{-1} \right) ' + \frac{Y_t^t \epsilon_t^t \epsilon_t^t Y_t}{H_t^t H_t^t} \right] < \infty \]

\[(A.94) \quad A_o = - E \left[ \frac{\partial^2 \phi_t}{\partial \Gamma^\mu \partial \Gamma^\mu'} \bigg|_{I_t-1} \right] \]

An argument identical to Lemma III verifies Condition II.

Condition III: Similar to Theorem II, Condition III. \( \square \)
# International Finance Discussion Papers

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

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.