QUANTITATIVE ANALYSIS UNIT

Q A U

GARCH-BASED IDENTIFICATION AND ESTIMATION OF TRIANGULAR SYSTEMS

FEDERAL RESERVE BANK OF BOSTON™

Working Paper No. QAU08-4

Todd Prono
Federal Reserve Bank of Boston

GARCH-Based Identification and Estimation of Triangular Systems

Todd Prono[1]
Federal Reserve Bank of Boston
September 2008

Abstract

Diagonal GARCH is shown to support identification of the triangular system and is argued as a higher moment analog to traditional exclusion restrictions used for determining suitable instruments. The estimator for this result is ML in the case where a distribution for the GARCH process is known and GMM otherwise. For the GMM estimator, an alternative weighting matrix is proposed.

JEL Codes: C13, C32. Keywords: Triangular Systems, Endogeneity, Identification, Heteroskedasticity, Generalized Method of Moments, GARCH.

Let $Y_{1,t}$ and $Y_{2,t}$ be observed endogenous variables, X_t' a vector of predetermined variables that can include lags of the endogenous variables, and $\epsilon_t = \begin{bmatrix}\epsilon_{1,t} & \epsilon_{2,t}\end{bmatrix}'$ unobserved errors. Specifying β_{1o} as the true value of β_1 and similarly for other parameters, consider the following model

$$Y_{1,t} = X_t'\beta_{1o} + Y_{2,t}\gamma_o + \epsilon_{1,t} \quad (1)$$

$$Y_{2,t} = X_t'\beta_{2o} + \epsilon_{2,t} \quad (2)$$

where the errors may be correlated and no exclusion restrictions are available for β_{1o}. Identification is shown if the errors follow a diagonal GARCH process. Works closely related to this one include Klein and Vella (2006) and Lewbel (2004). A distinguishing feature relative to the latter is that the covariance between errors is not assumed to be constant, while relative to the former, conditional variances of the errors are not sufficient for describing time-variation in the covariance. Sentana and Fiorentini (2001) explore GARCH-based identification of latent factor models. Their results also depend chiefly upon a constant covariance and necessarily involve a distributional assumption. Rigobon (2003) and Rigobon and Sack (2003) are examples of heteroskedasticity-based identification dependent upon discrete sets of independent variance regimes.

[1]Supervision, Regulation and Credit, 600 Atlantic Avenue, Boston, MA 02210 USA. Tel: (617)-973-3869 fax: (617)-619-8869 email: todd.prono@bos.frb.org

Assumption A1: *$E\left[X_t X_t'\right]$ and $E\left[X_t Y_t'\right]$ are finite and identified from the data. $E\left[X_t X_t'\right]$ is nonsingular.*

Define $S_{t-1} = \{X_t, X_{t-1}, \ldots, \epsilon_{t-1}, \epsilon_{t-2}, \ldots\}$. Consider the following definitions from Drost and Nijman (1993).

Definition D1 (Strong GARCH): $\epsilon_t = H_t^{1/2}\xi_t, \quad \xi_t \sim i.i.d.\ D\left(0,\ 1\right)$, where D specifies a known distribution.

Definition D2 (Semi-strong GARCH): $E\left[\epsilon_t \mid S_{t-1}\right] = 0, \quad E\left[\epsilon_t\epsilon_t^{'} \mid S_{t-1}\right] = H_t$.

Under either definition, parameterize H_t as

Assumption A2:

$$H_t = C_{0o}'C_{0o} + \sum_{k=1}^{2} A_{ko}'\epsilon_{t-1}\epsilon_{t-1}'A_{ko} + \sum_{k=1}^{2} B_{ko}'H_{t-1}B_{ko} \tag{3}$$

where $C_{0o} = \begin{bmatrix} c_{11,0o} & 0 \\ c_{21,0o} & c_{22,0o} \end{bmatrix}$, $A_{1o} = \begin{bmatrix} a_{11,1o} & 0 \\ 0 & a_{22,1o} \end{bmatrix}$, $A_{2o} = \begin{bmatrix} a_{11,2o} & 0 \\ 0 & 0 \end{bmatrix}$, $B_{1o} = \begin{bmatrix} b_{11,1o} & 0 \\ 0 & b_{22,1o} \end{bmatrix}$, *and* $B_{2o} = \begin{bmatrix} b_{11,2o} & 0 \\ 0 & 0 \end{bmatrix}$. *The parameters c_{1o}, c_{2o}, $a_{22,1o}$, $a_{11,2o}$, $b_{22,1o}$, and $b_{11,2o}$ are strictly positive.*

A2 defines a first-order diagonal bivariate BEKK model as detailed in Proposition 2.3 of Engle and Kroner (1995). This particular GARCH form is chosen because it establishes H_t as positive definite under very mild conditions. Applying the $vech\left(\cdot\right)$ operator to (3) and simplifying the result produces

$$h_t = C + Ae_{t-1} + Bh_{t-1} \tag{4}$$

where A and B are diagonal matrices whose nonzero elements are composite functions of the parameters in A_{ko} and B_{ko}, respectively, and $e_{t-1} = vech\left(\epsilon_{t-1}\epsilon_{t-1}^{'}\right)$.[2]

Assumption A3: *(i) The eigenvalues of $A + B$ are less than one in modulus. (ii) Let a_{ij} be the element in the ith row and jth column of the matrix A, and similarly define b_{ij}. $a_{33}+b_{33} \neq a_{22}+b_{22}$.*

(4) implies that $e_t = h_t + \omega_t$, where $E\left[\omega_t \mid S_{t-1}\right] = 0$ and $E\left[\omega_t\omega_s^{'} \mid S_{t-1}\right] = 0\ \forall\ s \neq t$. Let $\overline{e}_t = \left[\begin{array}{cc} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{array}\right]'$. Similarly define $\overline{h}_t$ and $\overline{\omega}_t$ as vectors of the second and third elements of h_t and ω_t, respectively.

Assumption A4: *(i) $E\left[\overline{\omega}_t\overline{\omega}_t^{'}\right] = \Sigma_{\overline{\omega}} < \infty$. (ii) $Cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$ is nonsingular if either $a_{11,1o}$ or $b_{11,1o}$ is nonzero.*

A3(i) defines h_t (or, equivalently, H_t) as mean stationary according to Proposition 2.7 of Engle and Kroner (1995). Given A4(i), $\overline{\omega}_t$ is covariance stationary. A3(i) and A4(i) together determine $\overline{e}_t$

[2]The $vech\left(\cdot\right)$ operator vertically stacks the elements on or below the principal diagonal of a square matrix into a column vector.

to be covariance stationary (see the Lemma and its proof in the Appendix), a condition that requires $\epsilon_{2,t}$ to be fourth moment stationary. Under D2, A4(i) is primitive, while under D1, it is substantive. For the latter, Theorem 3 of Hafner (2003) is necessary and sufficient for A4(i) if D belongs to the class of spherical distributions. Finally, note that if $a_{11,1o} = b_{11,1o} = 0$, then $Cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$ is singular.

Proposition 1 *Given D1 and A1–A3 for the model of (1) and (2), the structural parameters β_{1o}, β_{2o}, and γ_o are identified.*

Proof. $\beta_{2o} = E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{2,t}\right]$. The reduced form residuals of (1) and (2) are

$$R_{i,t} = Y_{i,t} - X_t' E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{i,t}\right], \quad i = 1, 2. \tag{5}$$

Let $R_t = \left[\begin{array}{cc} R_{1,t} & R_{2,t} \end{array}\right]'$, and note that (5) relates structural errors to their reduced form counterparts as

$$\epsilon_t = \Gamma_o^{-1} R_t \tag{6}$$

where $\Gamma_o = \left[\begin{array}{cc} 1 & \gamma_o \\ 0 & 1 \end{array}\right]$. Given (6) and the definition of e_{t-1}, the reduced form of (4) is

$$h_{r,t} = C_r + A_r r_{t-1} + B_r h_{r,t-1}.$$

The matrices A_r and B_r are upper triangular. Components along their principal diagonal equal the corresponding components of A and B, respectively (i.e., $dg\left(A_r\right) = A$ and $dg\left(B_r\right) = B$, where $dg\left(Z\right)$ forms a diagonal matrix from the elements along the principal diagonal of Z). Off diagonal elements of A_r and B_r are composite functions of the diagonal elements to A and B, respectively, as well as γ_o. From the elements of A_r and B_r,

$$\gamma_o = \frac{a_{23,r} + b_{23,r}}{\left(a_{33,r} - b_{33,r}\right) - \left(a_{22,r} - b_{22,r}\right)}. \tag{7}$$

Given (7), $\beta_{1o} = E\left[X_t X_t'\right]^{-1} E\left[X_t\left(Y_{1,t} - Y_{2,t}\gamma_o\right)\right]$. ■

Proposition 2 *Given D2 and A1–A4 for the model of (1) and (2), the structural parameters β_{1o}, β_{2o}, and γ_o are identified.*

Proof. If either $a_{11,1o}$ or $b_{11,1o}$ is nonzero as in A4(ii), then $E\left[\epsilon_{1,t}\epsilon_{2,t} \mid S_{t-1}\right]$ is time-varying. In this case, consider $Cov\left(\overline{e}_t,\ \overline{e}_{t-i}\right) = Cov\left(\overline{h}_t,\ \overline{e}_{t-i}\right)$, noting that $\overline{e}_t$ is covariance stationary. For $i \geq 1$, recursive substitution into $\overline{h}_t$ reveals that

$$Cov\left(\overline{e}_t,\ \overline{e}_{t-i}\right) = \left(\overline{A} + \overline{B}\right)^{i-1} Cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right) \tag{8}$$

where $\overline{A}$ is a 2×2 diagonal matrix formed from the elements a_{22} and a_{33} in A and similarly for $\overline{B}$. Given (6), the definition of e_{t-1}, and the relation between e_{t-1} and $\overline{e}_{t-1}$, the reduced form of (8) for $i = 2$ is

$$Cov\left(\overline{r}_t,\ \overline{r}_{t-2}\right) = \left(\overline{A}_r + \overline{B}_r\right) Cov\left(\overline{r}_t,\ \overline{r}_{t-1}\right)$$

where the relationship between $\overline{A}_r$ and A_r is equivalent to that between $\overline{A}$ and A. An analogous relationship exists between $\overline{B}_r$ and B_r. Identification of $\overline{A}_r + \overline{B}_r$ follows from the nonsingularity of $Cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$, with γ_o determined by (7).

Next, consider the case where $a_{11,1o} = b_{11,1o} = 0$. Define $Z_{t-1} = \left[\begin{array}{ccc} \epsilon^2_{2,t-1} & \cdots & \epsilon^2_{2,t-l} \end{array}\right]'$ for finite $l \geq 1$. Since $E\left[\epsilon_{1,t}\epsilon_{2,t} \mid S_{t-1}\right] = c_{21,0o}c_{22,0o}$, it follows that

$$Cov(\epsilon_{1,t}\epsilon_{2,t},\ Z_{t-1}) = 0. \tag{9}$$

From (6), $\epsilon_{1,t} = R_{1,t} - R_{2,t}\gamma_o$ and $R_{2,t} = \epsilon_{2,t}$. Substitution of these results into (9) produces

$$Cov(R_{1,t}\epsilon_{2,t},\ Z_{t-1}) = Cov(\epsilon^2_{2,t},\ Z_{t-1})\gamma_o.$$

Let $\Omega = Cov(\epsilon^2_{2,t},\ Z_{t-1})$, and note that $\Omega \neq 0$ given A2. Then γ_o is identified as $\gamma_o = \left(\Omega'\Omega\right)^{-1}\Omega' Cov(R_{1,t}\epsilon_{2,t},\ Z_{t-1})$.

Finally, in either case, β_{1o} and β_{2o} are identified according to the Proof of Proposition 1. ■

Propositions 1 and 2 are second-moment analogs to exclusion restrictions for β_{1o}. The diagonal GARCH model in (3) restricts each $h_{ij,t}$ in H_t to be a function of only past values of itself and of $\epsilon_{i,t}\epsilon_{j,t}$. Identification depends on these types of restrictions. If, instead, each $h_{ij,t}$ depends on past values of $h_{ij,t}$ and $\epsilon_{i,t}\epsilon_{j,t}$ $\forall\ i, j = 1, 2$ such that each $h_{ij,t}$ depends on six covariates as opposed to two in the diagonal case, then the number of reduced form parameters in A_r and B_r is less than the total number of structural parameters, and γ_o remains unidentified.

Let $H_{r,t}$ be the reduced form of H_t. In general, if H_t is diagonal (as it is in A2), $H_{r,t}$ will not be. Proposition 2.1 of Iglesias and Phillips (2004) demonstrates this result. Departures from diagonality in $H_{r,t}$ are precisely what identify γ_o. The numerator of (7), for example, is determined by off-diagonal elements. Suppose that $a_{33} = a_{22}$ and $b_{33} = b_{22}$ such that A3(ii) is violated. Then $\overline{A}_r = \overline{A}$, $\overline{B}_r = \overline{B}$, and γ_o is not identified from $\overline{A}_r$ or $\overline{B}_r$because it does not appear in either. This example represents a special case where diagonality in $E\left[\epsilon_{1,t}\epsilon_{2,t} \mid S_{t-1}\right]$ passes through to $E\left[R_{1,t}\epsilon_{2,t} \mid S_{t-1}\right]$. A3(ii), therefore, is necessary to preserve the off-diagonal elements in $E\left[R_{1,t}\epsilon_{2,t} \mid S_{t-1}\right]$ responsible for identification.

Owing to D2, Proposition 2 is a more general result than Proposition 1. The cost of this generality is paid in terms of stationary conditions for higher moments. Cragg (1997) and Lewbel (1997) require similar conditions for identification of the errors-in-variables model without distributional assumptions. If $a_{11,1o} = b_{11,1o} = 0$, Proposition 2 is a special case of Theorem 1 in Lewbel (2004).

Define $\theta = \{\beta_1,\ \beta_2,\ \gamma,\ C_0,\ A_k,\ B_k\}$ and Θ to be the set of all possible values for θ. In D1, specify $D = N$, the normal distribution. Let L_t be the log likelihood for observation t and L the joint log likelihood. Then $L = \sum\limits_{t=1}^{T} L_t$ and

$$L_t = -\frac{T}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|H_{r,t}\right| - \frac{1}{2}R_t' H_{r,t}^{-1} R_t.$$

Proposition 1 together with Theorems 1–3 of Lumsdaine (1996) can be used to establish

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} L$$

as a consistent and asymptotically normal estimator. Standard regularity conditions impose compactness on Θ. This condition needs to be reconciled with A3(ii). Suppose $|a_{22}| \leq a_{33}$ and $|b_{22}| \leq b_{33}$. A possible reconciliation is to define Θ such that every $\left(\frac{a_{22}}{a_{33}},\ \frac{b_{22}}{b_{33}}\right) \in \theta \in \Theta$ is exclusive of an open neighborhood of one.

Let $\Phi = \overline{A} + \overline{B}$ and $\sigma_{\overline{e}} = \left[\begin{array}{cc} \sigma_{12} & \sigma_{22} \end{array}\right]'$, where $\sigma_{12} = \frac{c_{21,0}c_{22,0}}{1-\phi_{11}}$ and $\sigma_{22} = \frac{c_{22,0}^2}{1-\phi_{22}}$. Define $Z = \left[\begin{array}{cc} \zeta_{11o} & 0 \\ 0 & \zeta_{22o} \end{array}\right]$, where

$$\zeta_{iio}^2 = E\left[\left(\epsilon_{i,t}\epsilon_{2,t} - \sigma_{i2}\right)^2\right], \quad i = 1, 2.$$

In addition, let $\psi = \{\beta_1,\ \beta_2,\ \gamma,\ \overline{C},\ \Phi\}$ and Ψ be the set of all possible values for ψ. Construct the following vector functions

$$U_1\left(\psi,\ Y_t,\ S_{t-1}\right) = X_t \otimes \epsilon_t$$

$$U_2\left(\psi,\ Y_t,\ S_{t-1}\right) = \overline{e}_t - \sigma_{\overline{e}}$$

$$U_3\left(\psi,\ Y_t,\ S_{t-1}\right) = vec\left[\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-2} - \sigma_{\overline{e}}\right)' - \Phi\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-1} - \sigma_{\overline{e}}\right)'\right]$$

and stack them into a single vector U_t.[3] $E\left[U_{2,t}\right] = 0$ grants that $\sigma_{\overline{e}} = E\left[\overline{e}_t\right]$. Therefore, $E\left[U_t\right] = 0$ is equivalent to $E\left[X_t\epsilon_{1,t}\right] = 0$ and $E\left[X_t\epsilon_{2,t}\right] = 0$ from (1) and (2) as well as $Cov\left(\overline{e}_t,\ \overline{e}_{t-2}\right) = \left(\overline{A} + \overline{B}\right) Cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$ (see (8) with $i = 2$). From Proposition 2, the only $\psi \in \Psi$ that satisfies $E\left[U_t\right] = 0$ is $\psi = \psi_o$.

Let $g = \frac{1}{T}\sum_{t=1}^{T} U_t$ and W be positive definite. Then Proposition 2 and Theorem 2.6 of Newey and McFadden (1994) can be used to establish

$$\widehat{\psi} = \underset{\psi \in \Psi}{\arg\min}\ g'Wg$$

as a consistent estimator.[4] Once again, Ψ needs to be compact. Reconciling this condition with A3(ii) follows a parallel argument to the one given above for Θ. Namely, assume $|\phi_{11}| \leq \phi_{22}$, and define Ψ such that every $\frac{\phi_{11}}{\phi_{22}} \in \psi \in \Psi$ is exclusive of an open neighborhood of one.

Suppose W is a consistent estimator of $W_o = E\left[U\left(\psi_o,\ Y_t,\ S_{t-1}\right) U\left(\psi_o,\ Y_t,\ S_{t-1}\right)'\right]$. Then asymptotic normality of $\widehat{\psi}$ follows from Theorem 3.4 of Newey and McFadden (1994). A necessary condition for this theorem to hold, however, is that $\epsilon_{2,t}$ be eighth moment stationary. If stationary higher moment conditions beyond those required under Proposition 2 prove overly restrictive, consistency of $\widehat{\psi}$ follows if $W = I$, where I is the identity matrix, in which case $\widehat{\psi}$ is the product of single-step GMM. Standard errors can be obtained by employing the nonoverlapping block bootstrap of Carlstein (1986), making sure to recenter the bootstrapped version of the

[3] The matrix operator $\otimes$ is the kronecker product. The $vec(\cdot)$ operator stacks the columns of an $(m \times n)$ matrix into an $(mn \times 1)$ vector.

[4] $\widehat{\psi}$ is the GMM estimator of Hansen (1982).

moment conditions relative to the population version as in Hall and Horowitz (1996).

Suppose X_t is a $k \times 1$ vector, and define I_j as the $j \times j$ identity matrix. Let $\widetilde{\zeta}_{ii}$ be a preliminary estimate of ζ_{iio}. In the case of single-step GMM estimation of $\widehat{\psi}$, an alternative choice for W is

$$W\left(\widetilde{Z}\right) = \begin{bmatrix} I_{2k\times 2k} & \cdots & 0 \\ \vdots & I_{2\times 2} & \vdots \\ 0 & \cdots & \left(\widetilde{Z}\otimes\widetilde{Z}\right)^{-1} \end{bmatrix}.$$

The weights $\left(\widetilde{Z}\otimes\widetilde{Z}\right)^{-1}$ impact the moments that define the autocovariances of $\overline{e}_t$, transforming these autocovariances into autocorrelations. The choice of $W\left(\widetilde{Z}\right)$ over I is motivated by the results of a Monte Carlo study (see the first row of Table 1) showing improved finite sample properties of $\widehat{\gamma}$ based on the former.

West (2002) demonstrates efficiency gains from using higher order lag terms to estimate AR processes that also display GARCH. Given (8), higher order lagged moments are available for inclusion in g by appending

$$U_{3+q}\left(\psi,\ Y_t,\ S_{t-1}\right) = vec\left[\left(\overline{e}_t-\sigma_{\overline{e}}\right)\left(\overline{e}_{t-q}-\sigma_{\overline{e}}\right)' - \Phi^{q-1}\left(\overline{e}_t-\sigma_{\overline{e}}\right)\left(\overline{e}_{t-1}-\sigma_{\overline{e}}\right)'\right], \quad q\geq 3$$

to U_t. $W\left(\widetilde{Z}\right)$ then needs to be redefined to include $(q-2)\times$ as many weighting matrices $\left(\widetilde{Z}\otimes\widetilde{Z}\right)^{-1}$ along the diagonal. The same Monte Carlo study mentioned above verifies reduced variability in $\widehat{\gamma}$ as q grows (see rows two and three of Table 1). However, this study also shows that the bias in $\widehat{\gamma}$ is increasing in q. Newey and Smith (2001) demonstrate that the GMM estimator can have large biases in the case of IV models with many instruments. Their theoretical result together with the Monte Carlo evidence presented here further supports the analogy between identification through GARCH and traditional exclusion restrictions. Existence of this bias advocates a modest value for q.

Appendix

Lemma 3 *Given A3(i) and A4(i), $\overline{e}_t$ is covariance stationary.*

Proof. From (4),

$$\overline{h}_t = \overline{C} + \overline{A}\overline{e}_{t-1} + \overline{B}\overline{h}_{t-1} \tag{10}$$

where $\overline{A}$ is a 2×2 diagonal matrix formed from the elements a_{22} and a_{33} in A and similarly for $\overline{B}$. Recursive substitution into (10) produces

$$\overline{h}_t = \sum_{i=1}^{\infty} \overline{B}^{i-1} \left(\overline{C} + \overline{A}\overline{e}_{t-i}\right). \tag{11}$$

Following the steps outlined in the proof to Proposition 2.7 of Engle and Kroner (1995), (11) can be used to show that

$$E_{t-\tau}\overline{e}_t = \left[I + \left(\overline{A}+\overline{B}\right) + \cdots + \left(\overline{A}+\overline{B}\right)^{\tau-2}\right]\overline{C} + \left(\overline{A}+\overline{B}\right)^{\tau-1} \sum_{i=1}^{\infty} \overline{B}^{i-1} \left(\overline{C} + \overline{A}\overline{e}_{t-i-\tau+1}\right)$$

where $E_{t-\tau}$ is the expectations operator conditional on the information set $S_{t-\tau}$. For a square matrix Z, it is well known that $Z^{\tau} \to 0$ as $\tau \to \infty$ if and only if the eigenvalues of Z are less than one in modulus. This same condition grants $(I + Z + \cdots + Z^{\tau-1}) \to (I - Z)^{-1}$ as $\tau \to \infty$ for the appropriately sized identity matrix I. Given A3(i), therefore, $E_{t-\tau}\overline{e}_t \xrightarrow{p} \left[I - \left(\overline{A}+\overline{B}\right)\right]^{-1}\overline{C}$ (as $\tau \to \infty$).

Since (4) implies that $e_t = h_t + \omega_t$, where $E\left[\omega_t \mid S_{t-1}\right] = 0$, and given A4(i),

$$E\left[\overline{e}_t\overline{e}_t^{'}\right] = E\left[\overline{h}_t\overline{h}_t^{'}\right] + \Sigma_{\overline{\omega}}.$$

Let $\sigma_{\overline{e}} = \left[I - \left(\overline{A}+\overline{B}\right)\right]^{-1}\overline{C}$.

$$\begin{aligned} E\left[\overline{h}_t\overline{h}_t^{'}\right] &= \eta + \overline{A}E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\overline{A}^{'} + \overline{A}\Sigma_{\overline{\omega}}\overline{A}^{'} + \overline{A}E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\overline{B}^{'} \\ &\quad + \overline{B}E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\overline{A}^{'} + \overline{B}E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\overline{B}^{'} \end{aligned} \tag{12}$$

where $\eta = \overline{C}\overline{C}^{'} + \left(\overline{A}+\overline{B}\right)\sigma_{\overline{e}}\overline{C}^{'} + \overline{C}\sigma_{\overline{e}}^{'}\left(\overline{A}+\overline{B}\right)^{'}$. Applying the $vec(\cdot)$ operator to (12) and simplifying yields

$$\begin{aligned} vec\left(E\left[\overline{h}_t\overline{h}_t^{'}\right]\right) &= \eta + (D)\,vec\left(E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\right) + \left(\overline{A}\otimes\overline{A}\right)vec\left(\Sigma_{\overline{\omega}}\right) \\ &= [I + D]\left(\eta + \left(\overline{A}\otimes\overline{A}\right)vec\left(\Sigma_{\overline{\omega}}\right)\right) + \left(D^2\right)vec\left(E\left[\overline{h}_{t-2}\overline{h}_{t-2}^{'}\right]\right) \\ &= \left[I + D + D^2\right]\left(\eta + \left(\overline{A}\otimes\overline{A}\right)vec\left(\Sigma_{\overline{\omega}}\right)\right) + \left(D^3\right)vec\left(E\left[\overline{h}_{t-3}\overline{h}_{t-3}^{'}\right]\right) \\ &= \ldots \\ &= \left[I + D + \cdots + D^{\tau-1}\right]\left(\eta + \left(\overline{A}\otimes\overline{A}\right)vec\left(\Sigma_{\overline{\omega}}\right)\right) + \left(D^{\tau}\right)vec\left(E\left[\overline{h}_{t-\tau}\overline{h}_{t-\tau}^{'}\right]\right) \end{aligned}$$

where $D = (\overline{A} + \overline{B}) \otimes (\overline{A} + \overline{B})$. Given A3(i), the eigenvalues of D are less than one in modulus, granting that $vec\left(E\left[\overline{h}_t \overline{h}_t^{'}\right]\right)$ converges to $[I - D]^{-1}\left(\eta + (\overline{A} \otimes \overline{A})\, vec\,(\Sigma_{\overline{\omega}})\right)$ as $\tau \to \infty$.

Note that

$$Cov\left(\overline{e}_t,\ \overline{e}_{t-\tau}\right) = E\left[\overline{e}_t \overline{e}_{t-\tau}^{'}\right] - \sigma_{\overline{e}} \sigma_{\overline{e}}^{'}$$

Consider the case where $\tau = 1$.

$$E\left[\overline{e}_t \overline{e}_{t-1}^{'} \mid S_{t-1}\right] = \overline{C}\overline{e}_{t-1}^{'} + \overline{A}\overline{e}_{t-1}\overline{e}_{t-1}^{'} + \overline{B}\overline{h}_{t-1}\overline{e}_{t-1}^{'}.$$

By iterated expectations,

$$E\left[\overline{e}_t \overline{e}_{t-1}^{'}\right] = \overline{C}\sigma_{\overline{e}}^{'} + (\overline{A} + \overline{B})\, \Sigma_{\overline{h}} + \overline{A}\Sigma_{\overline{\omega}}$$

and, as a result,

$$Cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right) = \left(\overline{C} - \sigma_{\overline{e}}\right)\sigma_{\overline{e}}^{'} + (\overline{A} + \overline{B})\, \Sigma_{\overline{h}} + \overline{A}\Sigma_{\overline{\omega}}$$

where $\Sigma_{\overline{h}} = E\left[\overline{h}_t \overline{h}_t^{'}\right]$. Next, consider the case where $\tau \geq 2$.

$$\begin{aligned}
E\left[\overline{h}_t \mid S_{t-\tau}\right] &= E\left[\overline{C} + \overline{A}\overline{e}_{t-1} + \overline{B}\overline{h}_{t-1} \mid S_{t-\tau}\right] \\
&= \overline{C} + (\overline{A} + \overline{B})\, E\left[\overline{h}_{t-1} \mid S_{t-\tau}\right] \\
&= \left[I + (\overline{A} + \overline{B})\right]\overline{C} + (\overline{A} + \overline{B})^2\, E\left[\overline{h}_{t-2} \mid S_{t-\tau}\right] \\
&= \ldots \\
&= \left[I + (\overline{A} + \overline{B}) + \ldots + (\overline{A} + \overline{B})^{\tau-1}\right]\overline{C} + (\overline{A} + \overline{B})^{\tau-1}\left[\overline{A}\overline{e}_{t-\tau} + \overline{B}\overline{h}_{t-\tau}\right] \\
&= \left[I - (\overline{A} + \overline{B})^{\tau}\right]\sigma_{\overline{e}} + (\overline{A} + \overline{B})^{\tau-1}\left[\overline{A}\overline{e}_{t-\tau} + \overline{B}\overline{h}_{t-\tau}\right].
\end{aligned}$$

By iterated expectations,

$$\begin{aligned}
E\left[\overline{e}_t \overline{e}_{t-\tau}^{'}\right] &= E\left[E\left[\overline{e}_t \overline{e}_{t-\tau}^{'} \mid S_{t-\tau}\right]\right] \\
&= E\left[E\left[\overline{h}_t \mid S_{t-\tau}\right]\overline{e}_{t-\tau}^{'}\right] \\
&= \left[I - (\overline{A} + \overline{B})^{\tau}\right]\sigma_{\overline{e}}\sigma_{\overline{e}}^{'} + (\overline{A} + \overline{B})^{\tau-1}\left[(\overline{A} + \overline{B})\, E\left[\overline{h}_{t-\tau}\overline{h}_{t-\tau}^{'}\right] + \overline{A}E\left[\overline{\omega}_{t-\tau}\overline{\omega}_{t-\tau}^{'}\right]\right].
\end{aligned}$$

As a result,

$$Cov\left(\overline{e}_t,\ \overline{e}_{t-\tau}\right) = (\overline{A} + \overline{B})^{\tau-1}\left[(\overline{A} + \overline{B})\left(\Sigma_{\overline{h}} - \sigma_{\overline{e}}\sigma_{\overline{e}}^{'}\right) + \overline{A}\Sigma_{\overline{\omega}}\right] \tag{13}$$

which converges to zero as $\tau \to \infty$, since $(\overline{A} + \overline{B})^{\tau-1} \to 0$ (as $\tau \to \infty$). Given (13), note that for $0 \leq i \leq \tau - 1$,

$$Cov\left(\overline{e}_t,\ \overline{e}_{t-\tau}\right) = (\overline{A} + \overline{B})^{i}\, Cov\left(\overline{e}_t,\ \overline{e}_{t-\tau+i}\right).$$

■

References

[1] Carlstein, E., 1986, The use of subseries methods for estimating the variance of a general statistic from a stationary time series, Annals of Statistics, 14, 1171-1179.

[2] Cragg, J.G., 1997, Using higher moments to estimate the simple errors-in-variables model, RAND Journal of Economics, 28, 571-591.

[3] Drost, F.C. and T.E. Nijman, 1993, Temporal aggregation of GARCH processes, Econometrica, 61, 909-927.

[4] Engle, R.F and K.F. Kroner, 1995, Multivariate simultaneous generalized GARCH, Econometric Theory, 11, 121-150.

[5] Hafner, C.M., 2003, Fourth moment structure of multivariate GARCH models, Journal of Financial Econometrics, 1, 26-54.

[6] Hall, P. and J.L. Horowitz, 1996, Bootstrap critical values for tests based on generalized-method-of-moments estimators, Econometrica, 64, 891-916.

[7] Hansen, L., 1982, Large sample properties of generalized method of moments estimators, Econometrica, 50, 1029-1054.

[8] Iglesias, E.M. and G.D.A Phillips, 2004, Simultaneous equations and weak instruments under conditionally heteroskedastic disturbances, unpublished manuscript.

[9] Lewbel, A., 2004, Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models, unpublished manuscript.

[10] Lewbel, A., 1997, Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D, Econometrica, 65, 1201-1213.

[11] Lumsdaine, R.L., 1996, Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, Econometrica, 64, 575-596.

[12] Newey, W.K. and D.L. McFadden, 1994, Large sample estimation and hypothesis testing, in: R.F. Engle and D.L. McFadden, eds., Handbook of Econometrics, Vol. IV, (Elsevier Science,), 2112-2244.

[13] Newey, W.K. and R.J. Smith, 2000, Asymptotic bias and equivalence of GMM and GEL estimators, unpublished manuscript.

[14] Rigobon, R., 2003, Identification through heteroskedasticity, Review of Economics and Statistics, 85, 777-792.

[15] Rigobon, R. and B. Sack, 2003, Measuring the response of monetary policy to the stock market, Quarterly Journal of Economics, 118, 639-669.

[16] Sentana, E. and G. Fiorentini, 2001, Identification, estimation and testing of conditionally heteroskedastic factor models, Journal of Econometrics, 102, 143-164.

[17] West, K.D., 2002, Efficient GMM estimation of weak AR processes, Economic Letters, 75, 415-418.

TABLE 1

Lags	Wght. Matrix	Med. Bias	MDAE	Dec. Range	SD
2	I	0.000	0.117	4.344	1.126
	$W\left(\widetilde{Z}\right)$	0.000	0.093	1.167	1.223
8	$W\left(\widetilde{Z}\right)$	0.118	0.136	0.414	0.167
16	$W\left(\widetilde{Z}\right)$	0.143	0.147	0.330	0.130

Notes: Consider (1) and (2) where $\beta_{1o} = \beta_{2o} = 0$ and $\gamma_o = 1$. Parameterize the errors according to D1 where $D = N$, the normal distribution. Let $a_{11,1o} = 0.13$, $a_{22,1o} = 0.32$, $a_{11,2o} = 0.18$, $b_{11,1o} = 0.89$, $b_{22,1o} = 0.89$, and $b_{11,2o} = 0.32$. Select values for C_{0o} such that $Var\left(\epsilon_{1,t}\right) = Var\left(\epsilon_{1,t}\right) = 1$ and $Cov\left(\epsilon_{1,t},\, \epsilon_{2,t}\right) = 0.20$. Monte Carlo studies were conducted across 5000 trials for $T = 1260$ observations. Results for $\widehat{\gamma}$ are shown. Med. Bias is the median bias of $\widehat{\gamma}$ relative to the true value. MDAE is the median absolute error of $\widehat{\gamma}$ relative to the true value. Dec. Range is the decile range, defined as the difference between the 0.10 and 0.90 quantiles of $\widehat{\gamma}$. SD is the standard deviation of $\widehat{\gamma}$. Robust measures of central tendency and dispersion are reported because of concerns over the existence of moments. The standard deviation, while not a robust measure, is also reported to give an indication of outliers.