GARCH-Based Identification and Estimation of Triangular Systems

Todd, Prono

Commodity Futures Trading Commission

September 2009

GARCH-Based Identification and Estimation of Triangular Systems[1]

Todd Prono[2]
Commodity Futures Trading Commission

Revised September 2009

Abstract

The diagonal GARCH(1,1) model is shown to support identification of the triangular system and is argued as a higher moment analog to traditional exclusion restrictions. Estimators for this result include QML and GMM. For the GMM estimator, only partial parameterization of the conditional covariance matrix is required. An alternative weighting matrix for the GMM estimator is also proposed.

JEL Codes: C13, C32. Keywords: Triangular Systems, Endogeneity, Identification, Heteroskedasticity, Quasi Maximum Likelihood, Generalized Method of Moments, GARCH, QML, GMM.

1. Introduction

Let $Y_{1,t}$ and $Y_{2,t}$ be observed endogenous variables, X_t a vector of predetermined variables that includes lags of the endogenous variables, and $\epsilon_t = \begin{bmatrix}\epsilon_{1,t} & \epsilon_{2,t}\end{bmatrix}'$ a vector of unobserved errors. Specifying $\beta_{1,0}$ as the true value of β_1 and similarly for other parameters, consider the model

$$Y_{1,t} = X_t'\beta_{1,0} + Y_{2,t}\gamma_0 + \epsilon_{1,t} \tag{1}$$

$$Y_{2,t} = X_t'\beta_{2,0} + \epsilon_{2,t} \tag{2}$$

where the errors may be correlated and no exclusion restrictions are available for $\beta_{1,0}$. Identification is shown if the errors follow a diagonal GARCH(1,1) process. This specification is the

[1]I owe graditude to Robin Lumsdaine, Arthur Lewbel, two anonymous referees, and seminar participants at the 2007 Summer Meeting of the Econometric Society for helpful comments and discussions. A substantial amount of work on this paper was completed during my tenure at the Federal Reserve Bank of Boston.

[2]Office of the Chief Economist, 1155 21st Street, N.W., Washington, DC 20581 USA. Tel: (202) 418-5460 email: tprono@cftc.gov

standard workhorse for modeling second moment dynamics of financial and macroeconomic time-series because of its parsimony and forecasting power.[3] Sentana and Fiorentini (2001) study how this specification identifies APT-style factor models. Their approach depends chiefly on constant conditional covariances. In contrast, the approach considered here allows the conditional covariance to be time-varying given a fairly general parametric form.

Other works closely related to this one are Klein and Vella (2006) and Lewbel (2008). Klein and Vella (2006) prove identification of the triangular system given a particular semi-parametric functional form for the heteroskedasticity in ϵ_t. Their estimator, however, is complicated to implement and relates the dynamics of the conditional covariance directly to the dynamics of each conditional error variance as in Bollerslev (1990). The estimators proposed here are straightforward applications of either QML or single-step GMM. In addition, the dynamics of the conditional covariance are not constrained by the processes that describe the two conditional variances. Lewbel (2008) also discusses heteroskedasticity-based identification of the triangular system, and his results can be viewed as a generalization of the results presented here, but only when the conditional covariance is constant.

2. Identification

Consider the following assumptions for the model of (1) and (2).

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 non-singular.

Define S_{t-1} as the σ-field generated by $\{X_t, X_{t-1}, \ldots, \epsilon_{t-1}, \epsilon_{t-2}, \ldots\}$. 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 $D\left(0,\ 1\right)$ specifies a distribution with zero mean and unit variance.

[3] Applications of multivariate GARCH(1,1) models to financial data include Bollerslev, Engle, and Wooldridge (1988) and Bollerslev (1990).

Definition D2 (Semi-strong GARCH):

$$E\left[\epsilon_t \mid S_{t-1}\right] = 0, \quad E\left[\epsilon_t \epsilon_t^{'} \mid S_{t-1}\right] = H_t.$$

Let

$$vech\left(H_t\right) = h_t, \qquad vech\left(\epsilon_t \epsilon_t^{'}\right) = e_t.$$

Throughout this paper, $vech\left(\cdot\right)$ denotes the matrix operator that stacks the lower triangle, including the diagonal, of a symmetric matrix into a column vector, while $vec\left(\cdot\right)$ is the matrix operator that stacks the columns of a matrix into a column vector. In addition, $A = \left[a_{jk}\right]$ denotes any matrix A.

Assumption A2:

$$h_t = C_0 + A_0 e_{t-1} + B_0 h_{t-1}, \tag{3}$$

where C_0 is a 3×1 vector of constants, and A_0 and B_0 are both 3×3 diagonal matrices.

Assumption A3: H_t is positive definite almost surely.

A2 defines a bivariate diagonal GARCH(1,1) model. A3 places restrictions on the parameters $c_{j1,0}$, $a_{jk,0}$, and $b_{jk,0}$ from that model. One way to satisfy A3 is to specify (3) according to a bivariate diagonal BEKK(1,1) model of Engle and Kroner (1995).[4]

Assumption A4: (i) The eigenvalues of $A_0 + B_0$ are less than one in modulus. (ii) $a_{33,0} + b_{33,0} \neq a_{22,0} + b_{22,0}$.

Given A4(i), the errors from the triangular system are covariance stationary.[5] A4(ii) imposes a restriction on the parameters governing the covariance between these errors and the variance of the errors in (2).

Proposition 1 *Given D1 and A1–A4 for the model of (1) and (2), the structural parameters $\beta_{1,0}$, $\beta_{2,0}$, and γ_0 are identified.*

[4] See Proposition 2.6 of the aforementioned work. In general, the BEKK models parameterize multivariate GARCH processes to ensure positivity.

[5] See Proposition 2.7 of Engle and Kroner (1995).

Proof. From A1,

$$\beta_{2,0} = E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{2,t}\right]. \tag{4}$$

In addition, 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 $R_t = \begin{bmatrix} R_{1,t} & R_{2,t} \end{bmatrix}'$. (5) relates structural errors to their reduced form counterparts as

$$\epsilon_t = \Gamma_0^{-1} R_t, \tag{6}$$

where $\Gamma_0 = \begin{bmatrix} 1 & \gamma_0 \\ 0 & 1 \end{bmatrix}$. Substitution of (6) into (3) produces the reduced form equation

$$h_{r,t} = C_{0,r} + A_{0,r} r_{t-1} + B_{0,r} h_{r,t-1}, \tag{7}$$

where $h_{r,t} = vech\left(E\left[R_t R_t' \mid S_{t-1}\right]\right)$ and $r_{t-1} = vech\left(R_{t-1} R_{t-1}'\right)$. Both matrices $A_{0,r}$ and $B_{0,r}$ are 3×3 upper triangular. Let $dg\left(\cdot\right)$ be the matrix operator that forms a diagonal matrix from the elements along the principal diagonal of any matrix. Then $dg\left(A_{0,r}\right) = A_0$, $dg\left(B_{0,r}\right) = B_0$, and the off diagonal elements of $A_{0,r}$ and $B_{0,r}$ are composite functions of the respective diagonal elements as well as γ_0. Therefore, the parameter matrix $A_{0,r}$ defines a system of 4 linearly independent reduced form equations in 4 structural unknowns.[6] If $a_{33,0} \neq a_{22,0}$, then γ_0 is identified from that system as

$$\gamma_0 = \frac{a_{23,0,r}}{a_{33,0,r} - a_{22,0,r}}. \tag{8}$$

If $a_{33,0} = a_{22,0}$, the parameter matrix $B_{0,r}$ defines a system of 4 linearly independent reduced form equations in 4 structural unknowns with a solution for γ_0 in the same form as (8). A4(ii) establishes existence of this solution. $\beta_{1,0}$ is then identified as

$$\beta_{1,0} = E\left[X_t X_t'\right]^{-1} E\left[X_t\left(Y_{1,t} - Y_{2,t}\gamma_0\right)\right]. \tag{9}$$

[6]Those four unknowns are $a_{11,0}$, $a_{22,0}$, $a_{33,0}$, and γ_0.

■

Iglesias and Phillips (2004) demonstrate that if the structural errors from a triangular system follow a diagonal GARCH process, the reduced form errors, while still GARCH, are no longer diagonal GARCH. The reduced form parameter matrices $A_{0,r}$ and $B_{0,r}$ both illustrate this point and evidence how departures from diagonality permit identification. In particular, since $dg\left(A_{0,r}\right) = A_0$, elements in the upper triangle of $A_{0,r}$ are restricted by the diagonal terms. It is from these restrictions that identification follows. In discussing how the relationship between structural and reduced form GARCH models can identify simultaneous systems, Rigobon (2002) states "the model of heteroskedasticity of the structural residuals impose[s] important constraints on how the reduced form heteroskedasticity can evolve" (p.433). The constraint in Proposition 1 is the exclusion of all off-diagonal terms in the formulation of h_t.

The traditional method for identifying the triangular system is to impose exclusion restrictions (i.e., zero restrictions) on some of the parameters in $\beta_{1,0}$. Diagonality of the parameter matrices A_0 and B_0 in (3) is the extension of exclusion restrictions onto the second moments. Without these restrictions, the triangular system would remain unidentified, which is to say that the existence of conditional heteroskedasticity alone is not sufficient for identification. For instance, suppose h_t follows a fully general GARCH model, which requires A_0 and B_0 to be composed entirely of nonzero terms. Then the structural form GARCH model imposes no constraints on how the reduced form can evolve. In the context of the proof to Proposition 1, the lack of these constraints translates $A_{0,r}$ into a system of 9 equations in 10 unknowns (the 9 structural parameters in A_0 and γ_0).[7]

Apparent from the proof to Proposition 1, while identification depends on h_t following a diagonal GARCH model, the exact order of that model is unimportant. A general specification for h_t as a diagonal GARCH(p, q) process with $p, q \geq 1$ still supports identification. The special case of $p = q = 1$ is chosen here for simplicity.

Proposition 1 identifies the triangular system from the conditional moment restrictions of a diagonal GARCH model. The following assumptions are necessary to base identification on the unconditional autocovariances implied by a diagonal GARCH(1,1) model.

[7] Rigobon (2002) formalizes this result in an appendix.

To begin, (3) implies that

$$e_t = h_t + \omega_t, \tag{10}$$

where $E\left[\omega_t \mid S_{t-1}\right] = 0$ and $E\left[\omega_t \omega_s' \mid S_{t-1}\right] = 0 \;\forall\; s \neq t$. Consider only the conditional covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ as well as the conditional variance of $\epsilon_{2,t}$. In doing so, let $\overline{e}_t = \left[\begin{array}{cc} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{array}\right]'$ and similarly define $\overline{h}_t$ and $\overline{\omega}_t$ as vectors of the second and third elements of h_t and ω_t, respectively. In addition, let $Z_{t-2} = \left[\overline{e}_{t-2}' \cdots \overline{e}_{t-L}'\right]'$ for a finite $L \geq 2$, and define $Cov\left[\overline{e}_t,\ Z_{t-i}\right] \equiv E\left[\left(\overline{e}_t - E\left[\overline{e}_t\right]\right)\left(Z_{t-i} - E\left[Z_t\right]\right)'\right]$ for $i \geq 1$.

Assumption A5: (i) $E\left[\overline{\omega}_t \overline{\omega}_t'\right] = \Sigma_{\overline{\omega}} < \infty$. (ii) $Cov\left[\overline{e}_t,\ Z_{t-1}\right]$ has full row rank if either $a_{22,0}$ or $b_{22,0}$ is nonzero.[8]

Given A5(i), $\overline{\omega}_t$ is covariance stationary. A4(i) and A5(i) together determine $\overline{e}_t$ 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.[9]

Proposition 2 *Given D2 and A1–A5 for the model of (1) and (2), the structural parameters $\beta_{1,0}$, $\beta_{2,0}$, and γ_0 are identified.*

Proof. From (19) follows

$$Cov\left[\overline{e}_t,\ \overline{e}_{t-\tau}\right] = \left(\overline{A}_0 + \overline{B}_0\right) Cov\left[\overline{e}_t,\ \overline{e}_{t-(\tau-1)}\right], \tag{11}$$

where $\overline{A}_0$ is a 2×2 diagonal matrix formed from the elements $a_{22,0}$ and $a_{33,0}$ in A_0 and similarly for $\overline{B}_0$. (11) grants that

$$Cov\left[\overline{e}_t,\ Z_{t-2}\right] = \left(\overline{A}_0 + \overline{B}_0\right) Cov\left[\overline{e}_t,\ Z_{t-1}\right]. \tag{12}$$

Substitution of the results from (6) into (12) produces

$$Cov\left[\overline{r}_t, Z_{r,t-2}\right] = \left(\overline{A}_{0,r} + \overline{B}_{0,r}\right) Cov\left[\overline{r}_t,\ Z_{r,t-1}\right],$$

[8] If $a_{22,0} = b_{22,0} = 0$, then $Cov\left[\overline{e}_t,\ Z_{t-1}\right]$ has a row rank of one.

[9] The Lemma is an extension of Theorem 3 in Hafner (2003) to the semi-strong diagonal GARCH(1,1) model.

where $\overline{r}_t$ and $Z_{r,t-i}$ (for $i = 1, 2$) are the reduced forms of $\overline{e}_t$ and Z_{t-i}, respectively. $\overline{A}_{0,r}$ is defined from

$$A_{0,r} = \begin{bmatrix} a_{11,0,r} & a_{12,0,r} & a_{13,0,r} \\ 0 & a_{22,0,r} & a_{23,0,r} \\ 0 & 0 & a_{33,0,r} \end{bmatrix}$$

in (7) as

$$\overline{A}_{0,r} = \begin{bmatrix} a_{22,0,r} & a_{23,0,r} \\ 0 & a_{33,0,r} \end{bmatrix} = \begin{bmatrix} a_{22,0} & \gamma_0 \left(a_{33,0} - a_{22,0}\right) \\ 0 & a_{33,0} \end{bmatrix}.$$

Since $\overline{B}_{0,r}$ is afforded a parallel definition in terms of the elements of $B_{0,r}$, identification of γ_0 as

$$\gamma_0 = \frac{a_{23,0,r} + b_{23,0,r}}{\left(a_{33,0,r} + b_{33,0,r}\right) - \left(a_{22,0,r} + b_{22,0,r}\right)} \tag{13}$$

follows given A4(ii) if and only if $\overline{A}_{0,r} + \overline{B}_{0,r}$ is identified. Let $\Omega\left(i\right)_r = Cov\left[\overline{r}_t,\ Z_{r,t-i}\right]$. Then, given A5(ii), $\overline{A}_{0,r} + \overline{B}_{0,r}$ is identified as

$$\left(\overline{A}_{0,r} + \overline{B}_{0,r}\right) = \Omega\left(2\right)_r \Omega\left(1\right)_r' \left[\Omega\left(1\right)_r \Omega\left(1\right)_r'\right]^{-1}.$$

Next, consider the case where $a_{22,0} = b_{22,0} = 0$. Define $Z_{2,t-1} = \begin{bmatrix} \epsilon^2_{2,t-1} & \cdots & \epsilon^2_{2,t-L} \end{bmatrix}'$. Since $h_{12,t} = E\left[\epsilon_{1,t}\epsilon_{2,t} \mid S_{t-1}\right]$ is constant,

$$Cov\left[\epsilon_{1,t}\epsilon_{2,t},\ Z_{2,t-1}\right] = 0. \tag{14}$$

From (6), $\epsilon_{1,t} = R_{1,t} - R_{2,t}\gamma_0$ and $R_{2,t} = \epsilon_{2,t}$. Substitution of these results into (14) produces

$$Cov\left[R_{1,t}\epsilon_{2,t},\ Z_{2,t-1}\right] = Cov\left[\epsilon^2_{2,t},\ Z_{2,t-1}\right]\gamma_0.$$

Let $\Omega\left(i\right)_2 = Cov\left[\epsilon^2_{2,t},\ Z_{2,t-i}\right]$, and note that $\Omega\left(i\right)_2 \neq 0$ given A2. Then γ_0 is identified as

$$\gamma_0 = \left[\Omega\left(1\right)_2' \Omega\left(1\right)_2\right]^{-1} \Omega\left(1\right)_2' Cov(R_{1,t}\epsilon_{2,t},\ Z_{t-1}). \tag{15}$$

Regardless of whether γ_0 is identified by (13) or (15), $\beta_{2,0}$ is identified by (4), and, given identification of γ_0, $\beta_{1,0}$ is identified by (9). ■

From the proof to Proposition 2, two observations are important. First, identification of γ_0 depends on identification of the sum of the reduced form ARCH and GARCH terms from the error covariance as well as the variance of $\epsilon_{2,t}$. Separate identification of these ARCH and GARCH terms is not necessary. Second, the conditional variance of $\epsilon_{1,t}$ plays no role in identification. Therefore, its parameterization in (3) need not imply a finite fourth moment for $\epsilon_{1,t}$ as is the requirement for $h_{22,t}$ in regards to $\epsilon_{2,t}$. Moreover, the variance of $\epsilon_{1,t}$ could be homoskedastic, or it could follow some alternative heteroskedastic process.[10] In either case, Proposition 2 continues to hold. This second observation stands in contrast to Proposition 1, which bases identification on a parameterization of the complete error variance-covariance matrix.

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. Finally, if $a_{22,0} = b_{22,0} = 0$, Proposition 2 is a special case of Theorem 1 in Lewbel (2004).

2. Estimation

For the observed data $\{(Y_t,\ X_t)\,,\ t = 1, \ldots, T\}$, let $I_{t-1} = \{X_t, \ldots, X_1\}$. Consider estimation of the triangular system given Propositions 1 and 2. Beginning with Proposition 1, define $\theta = \{\beta_1,\ \beta_2,\ \gamma,\ C,\ A,\ B\}$ and Θ to be the set of all possible values for θ. Let $L = \sum_{t=1}^{T} l_t\left(Y_t,\ I_{t-1};\ \theta\right)$, where

$$l_t\left(Y_t,\ I_{t-1};\ \theta\right) = -\ln\left(2\pi\right) - \frac{1}{2}\ln\left|H_{r,t}\right| - \frac{1}{2}R_t' H_{r,t}^{-1} R_t,$$

and $H_{r,t}$ is the reduced form of H_t. Assuming certain regularity conditions discussed in Bollerslev and Wooldridge (1992) and given Proposition 1, the estimator

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

[10]For example, $h_{11,t}$ could follow a discrete regime switching model as in Rigobon (2003) and Rigobon and Sack (2003).

is weakly consistent. From Theorem 2.1 of Bollerslev and Wooldridge (1992), $\widehat{\theta}$ is also asymptotically normal. Neither weak consistency nor asymptotic normality depend on D being the normal distribution in D1.

Under Proposition 1, the diagonal GARCH model is the key identifying assumption. Defending this specification is its widespread use in financial and macroeconomic volatility forecasting. The importance of this assumption, however, warrants formal diagnostics for assessing its appropriateness to the given data under study. Towards that end, the robust Lagrange multiplier tests proposed by Bollerslev and Wooldridge (1992) are applicable.

Next, consider estimation under Proposition 2. Let $\epsilon_t = \begin{bmatrix} \epsilon_{1,t} & \epsilon_{2,t} \end{bmatrix}'$, and $\overline{e}_t = \begin{bmatrix} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{bmatrix}'$, where

$$
\begin{aligned}
\epsilon_{1,t} &= Y_{1,t} - X_t'\beta_1 - Y_{2,t}\gamma \\
\epsilon_{2,t} &= Y_{2,t} - X_t'\beta_2
\end{aligned}
$$

Define $\psi = \{\beta_1,\ \beta_2,\ \gamma,\ \overline{C},\ \overline{A} + \overline{B}\}$, where $\overline{C} = \begin{bmatrix} c_{21} & c_{31} \end{bmatrix}'$, and Ψ as the set of all possible values for ψ. In addition, $\overline{\sigma} = \left[I - \left(\overline{A} + \overline{B}\right)\right]^{-1} \overline{C}$, where I is the identity matrix, and $z_{t-2} = \left[\left(\overline{e}_{t-2} - \overline{\sigma}\right)' \cdots \left(\overline{e}_{t-L} - \overline{\sigma}\right)'\right]'$. Consider the following vector functions

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

$$U_2\left(Y_t,\ I_{t-1};\ \psi\right) = \overline{e}_t - \overline{\sigma}$$

$$U_3\left(Y_t,\ I_{t-1};\ \psi\right) = vec\left(\left(\overline{e}_t - \overline{\sigma}\right) z_{t-2}' - \left(\overline{A} + \overline{B}\right)\left(\overline{e}_t - \overline{\sigma}\right) z_{t-1}'\right)$$

stacked into a single vector $U\left(Y_t,\ I_{t-1};\ \psi\right)$.[11] Proposition 2 establishes $\psi = \psi_0$ as the only $\psi \in \Psi$ satisfying $E\left[U\left(Y_t,\ I_{t-1};\ \psi\right)\right] = 0$.

Construct the standard Hansen (1982) GMM estimator

$$\widehat{\psi} = \underset{\psi \in \Psi}{\arg\min} \left[T^{-1} \sum_{t=1}^{T} U\left(Y_t,\ I_{t-1};\ \psi\right)\right]' W_T \left[T^{-1} \sum_{t=1}^{T} U\left(Y_t,\ I_{t-1};\ \psi\right)\right],$$

[11] The matrix operator $\otimes$ denotes the Kronecker product.

for some sequence of positive definite W_T. If (i) $\psi_0 \in \text{int}\ \Psi$, (ii) $W_T \xrightarrow{p} W_0$, and (iii) $U(\cdot;\ \psi)$ satisfies the weak uniform law of large numbers from Wooldridge (1990, Definition A.1), then $\widehat{\psi}$ is weakly consistent. Asymptotic normality can also follow, but only if $\epsilon_{2,t}$ is eighth moment stationary. If higher moment stationary conditions beyond those required under Proposition 2 prove overly restrictive, standard errors for $\widehat{\psi}$ can be obtained by employing the nonoverlapping block bootstrap of Carlstein (1986), making sure to recenter the bootstrap version of the moment conditions relative to the population version as in Hall and Horowitz (1996).

Weak consistency of $\widehat{\theta}$ and $\widehat{\psi}$ requires Θ and Ψ, respectively, to be compact. This condition needs to be reconciled with A4(ii). Suppose $a_{22} + b_{22} \geq 0$. Then a possible reconciliation might be to redefine Θ and Ψ such that $\frac{a_{22}+b_{22}}{a_{33}+b_{33}}$ is finite, nonnegative, and exclusive of an open neighborhood of one.

If $W_T = I$, then $\widehat{\psi}$ is the product of single-step GMM. Let $L = 2$, and suppose X_t is a $k \times 1$ vector. Define I_j as the $j \times j$ identity matrix. Let $\widetilde{\lambda}_{ij}$ be a preliminary estimate of $\lambda_{ij,0}$, where $\lambda^2_{ij,0} = E\left[\left(\epsilon_{i,t}\epsilon_{j,t} - \sigma_{ij}\right)^2\right]$ for $i, j = 1, 2$, and construct $\widetilde{\Lambda} = \begin{bmatrix} \widetilde{\lambda}_{12} & 0 \\ 0 & \widetilde{\lambda}_{22} \end{bmatrix}$. Allowing

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

results in improved finite sample properties for $\widehat{\psi}$ over $W_T = I$. The first two rows of Table 1 illustrate this result for $\widehat{\gamma}$, as noted by the marked reduction in median absolute error (MDAE) and decile range. The weights $\left(\widetilde{\Lambda}\otimes\widetilde{\Lambda}\right)^{-1}$ impact the sample moments defined by $U_3(\cdot;\ \psi)$, transforming these sample autocovariances into sample autocorrelations. For $L > 2$, $W\left(\widetilde{\Lambda}\right)$ needs to be redefined to include $L - 1$ additional weighting matrices $\left(\widetilde{\Lambda}\otimes\widetilde{\Lambda}\right)^{-1}$ along the diagonal.

Under Proposition 2, the diagonal GARCH(1,1) process is a key identifying assumption. For $L > 2$, this assumption is testable by a bootstrap version of the χ^2 difference test proposed by Newey and West (1987).

West (2002) demonstrates efficiency gains from using higher order lag terms to estimate finite order AR models by GMM in the presence of GARCH errors. Table 1 provides a similar commentary for the triangular system, as noted by a reduction in the variability of $\widehat{\gamma}$ as L grows. However, also noted in Table 1 is an increase in the bias of $\widehat{\gamma}$ as L increases. 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 L.

Appendix

Lemma *Given A4(i) and A5(i), $\overline{e}_t$ is covariance stationary.*

Proof. Given the definitions of $\overline{e}_t$ and $\overline{h}_t$, it follows from (3) that

$$\overline{h}_t = \overline{C}_0 + \overline{A}_0 \overline{e}_{t-1} + \overline{B}_0 \overline{h}_{t-1}, \tag{16}$$

where $\overline{A}_0$ is a 2×2 diagonal matrix formed from the elements $a_{22,0}$ and $a_{33,0}$ in A_0; $\overline{B}_0$ is similarly defined in terms of the elements in $\overline{B}_0$, and $\overline{C}_0$ is a 2×1 vector of constant terms for the conditional error covariance and the conditional variance of $\epsilon_{2,t}$. Recursive substitution into (16) produces

$$\overline{h}_t = \sum_{i=1}^{\infty} \overline{B}_0^{i-1} \left(\overline{C}_0 + \overline{A}_0 \overline{e}_{t-i} \right). \tag{17}$$

Following the steps outlined in the proof to Proposition 2.7 of Engle and Kroner (1995), (17) can be used to show that

$$E_{t-\tau}\left[\overline{e}_t\right] = \left[I + \left(\overline{A}_0 + \overline{B}_0\right) + \cdots + \left(\overline{A}_0 + \overline{B}_0\right)^{\tau-2} \right] \overline{C}_0 + \left(\overline{A}_0 + \overline{B}_0\right)^{\tau-1} \sum_{i=1}^{\infty} \overline{B}_0^{i-1} \left(\overline{C}_0 + \overline{A}_0 \overline{e}_{t-i-\tau+1} \right),$$

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 $\left(I + Z + \cdots + Z^{\tau-1}\right) \to (I - Z)^{-1}$ as $\tau \to \infty$. Given A3(i), therefore, $E_{t-\tau}\left[\overline{e}_t\right] \xrightarrow{p} \left[I - \left(\overline{A}_0 + \overline{B}_0\right)\right]^{-1} \overline{C}_0$ (as $\tau \to \infty$).

From (10),

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

given A5(i). Let $\overline{\sigma}_0 = \left[I - \left(\overline{A}_0 + \overline{B}_0\right)\right]^{-1} \overline{C}_0$.

$$\begin{aligned} E\left[\overline{h}_t \overline{h}_t^{'}\right] &= \eta_0 + \overline{A}_0 E\left[\overline{h}_{t-1} \overline{h}_{t-1}^{'}\right] \overline{A}_0 + \overline{A}_0 \Sigma_{\overline{\varpi}} \overline{A}_0 + \overline{A}_0 E\left[\overline{h}_{t-1} \overline{h}_{t-1}^{'}\right] \overline{B}_0 \\ &\quad + \overline{B}_0 E\left[\overline{h}_{t-1} \overline{h}_{t-1}^{'}\right] \overline{B}_0 + \overline{B}_0 E\left[\overline{h}_{t-1} \overline{h}_{t-1}^{'}\right] \overline{B}_0 \end{aligned} \tag{18}$$

where $\eta_0 = \overline{C}_0\overline{C}_0^{'} + \left(\overline{A}_0 + \overline{B}_0\right)\overline{\sigma}_0\overline{C}_0^{'} + \overline{C}_0\overline{\sigma}_0^{'}\left(\overline{A}_0 + \overline{B}_0\right)$. Applying the $vec(\cdot)$ operator to (18) and simplifying yields

$$\begin{aligned}
vec\left(E\left[\overline{h}_t\overline{h}_t^{'}\right]\right) &= \eta_0 + D_0 vec\left(E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\right) + \left(\overline{A}_0 \otimes \overline{A}_0\right) vec\left(\Sigma_{\overline{\varpi}}\right) \\
&= \left[I + D_0\right]\left(\eta_0 + \left(\overline{A}_0 \otimes \overline{A}_0\right) vec\left(\Sigma_{\overline{\varpi}}\right)\right) + \left(D_0^2\right) vec\left(E\left[\overline{h}_{t-2}\overline{h}_{t-2}^{'}\right]\right) \\
&= \left[I + D_0 + D_0^2\right]\left(\eta_0 + \left(\overline{A}_0 \otimes \overline{A}_0\right) vec\left(\Sigma_{\overline{\varpi}}\right)\right) + \left(D_0^3\right) vec\left(E\left[\overline{h}_{t-3}\overline{h}_{t-3}^{'}\right]\right) \\
&= \ldots \\
&= \left[I + D_0 + \cdots + D_0^{\tau-1}\right]\left(\eta_0 + \left(\overline{A}_0 \otimes \overline{A}_0\right) vec\left(\Sigma_{\overline{\varpi}}\right)\right) + \left(D_0^{\tau}\right) vec\left(E\left[\overline{h}_{t-\tau}\overline{h}_{t-\tau}^{'}\right]\right)
\end{aligned}$$

where $D_0 = \left(\overline{A}_0 + \overline{B}_0\right) \otimes \left(\overline{A}_0 + \overline{B}_0\right)$. Given A3(i), the eigenvalues of D_0 are less than one in modulus, granting that $vec\left(E\left[\overline{h}_t\overline{h}_t^{'}\right]\right)$ converges to $\left[I - D_0\right]^{-1}\left(\eta_0 + \left(\overline{A}_0 \otimes \overline{A}_0\right) vec\left(\Sigma_{\overline{\varpi}}\right)\right)$ as $\tau \to \infty$.

Note that

$$Cov\left[\overline{e}_t,\ \overline{e}_{t-\tau}\right] = E\left[\overline{e}_t\overline{e}_{t-\tau}^{'}\right] - \overline{\sigma}_0\overline{\sigma}_0^{'}$$

Consider the case where $\tau = 1$.

$$E\left[\overline{e}_t\overline{e}_{t-1}^{'} \mid S_{t-1}\right] = \overline{C}_0\overline{e}_{t-1}^{'} + \overline{A}_0\overline{e}_{t-1}\overline{e}_{t-1}^{'} + \overline{B}_0\overline{h}_{t-1}\overline{e}_{t-1}^{'}.$$

By iterated expectations,

$$E\left[\overline{e}_t\overline{e}_{t-1}^{'}\right] = \overline{C}_0\overline{\sigma}_0^{'} + \left(\overline{A}_0 + \overline{B}_0\right)\Sigma_{\overline{h}} + \overline{A}_0\Sigma_{\overline{\varpi}}$$

and, as a result,

$$Cov\left[\overline{e}_t,\ \overline{e}_{t-1}\right] = \left(\overline{C}_0 - \overline{\sigma}_0\right)\overline{\sigma}_0^{'} + \left(\overline{A}_0 + \overline{B}_0\right)\Sigma_{\overline{h}} + \overline{A}_0\Sigma_{\overline{\varpi}}$$

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

$$\begin{aligned}
E\left[\overline{h}_t \mid S_{t-\tau}\right] &= E\left[\overline{C}_0 + \overline{A}_0\overline{e}_{t-1} + \overline{B}_0\overline{h}_{t-1} \mid S_{t-\tau}\right] \\
&= \overline{C}_0 + \left(\overline{A}_0 + \overline{B}_0\right) E\left[\overline{h}_{t-1} \mid S_{t-\tau}\right] \\
&= \left[I + \left(\overline{A}_0 + \overline{B}_0\right)\right]\overline{C}_0 + \left(\overline{A}_0 + \overline{B}_0\right)^2 E\left[\overline{h}_{t-2} \mid S_{t-\tau}\right] \\
&= \ldots \\
&= \left[I + \left(\overline{A}_0 + \overline{B}_0\right) + \ldots + \left(\overline{A}_0 + \overline{B}_0\right)^{\tau-1}\right]\overline{C}_0 + \left(\overline{A}_0 + \overline{B}_0\right)^{\tau-1}\left[\overline{A}_0\overline{e}_{t-\tau} + \overline{B}_0\overline{h}_{t-\tau}\right] \\
&= \left[I - \left(\overline{A}_0 + \overline{B}_0\right)^{\tau}\right]\overline{\sigma}_0 + \left(\overline{A}_0 + \overline{B}_0\right)^{\tau-1}\left[\overline{A}_0\overline{e}_{t-\tau} + \overline{B}_0\overline{h}_{t-\tau}\right].
\end{aligned}$$

By iterated expectations,

$$\begin{aligned}
E\left[\overline{e}_t\overline{e}_{t-\tau}^{'}\right] &= E\left[E\left[\overline{e}_t\overline{e}_{t-\tau}^{'} \mid S_{t-\tau}\right]\right] \\
&= E\left[E\left[\overline{h}_t \mid S_{t-\tau}\right]\overline{e}_{t-\tau}^{'}\right] \\
&= \left[I - \left(\overline{A}_0 + \overline{B}_0\right)^{\tau}\right]\overline{\sigma}_0\overline{\sigma}_0^{'} + \left(\overline{A}_0 + \overline{B}_0\right)^{\tau-1}\left[\left(\overline{A}_0 + \overline{B}_0\right)E\left[\overline{h}_{t-\tau}\overline{h}_{t-\tau}^{'}\right] + \overline{A}_0 E\left[\overline{\omega}_{t-\tau}\overline{\omega}_{t-\tau}^{'}\right]\right].
\end{aligned}$$

As a result,

$$Cov\left[\overline{e}_t,\ \overline{e}_{t-\tau}\right] = \left(\overline{A}_0 + \overline{B}_0\right)^{\tau-1}\left[\left(\overline{A}_0 + \overline{B}_0\right)\left(\Sigma_{\overline{h}} - \overline{\sigma}_0\overline{\sigma}_0^{'}\right) + \overline{A}_0\Sigma_{\overline{\omega}}\right] \tag{19}$$

which converges to zero as $\tau \to \infty$, since $\left(\overline{A}_0 + \overline{B}_0\right)^{\tau-1} \to 0$ (as $\tau \to \infty$). ■

References

[1] Bollerslev, T., 1990, Modelling the coherence in short run nominal exchange rates: a multivariate generalized ARCH model, Review of Economics and Statistics, 72, 498-505.

[2] Bollerslev, T., R.F Engle and J.M. Wooldridge, 1988, A capital asset pricing model with time-varying covariances, Journal of the Political Economy, 96, 116-131.

[3] Bollerslev, T. and J.M. Wooldridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, Econometric Reviews, 11, 143-172.

[4] 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.

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

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

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

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

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

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

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

[12] Klein, R. and F. Vella, 2006, Estimating a class of triangular simultaneous equations models without exclusion restrictions, unpublished manuscript

[13] Lewbel, A., 2008, Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models, unpublished manuscript.

[14] 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.

[15] 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.

[16] Newey, W.K. and K.D. West, 1987, Hypothesis testing with efficient method of moments, International Economic Review, 28, 777-787.

[17] Newey, W.K. and R.J. Smith, 2001, Asymptotic bias and equivalence of GMM and GEL estimators, unpublished manuscript.

[18] Prono, T., 2009, Market proxies, correlation, and relative mean-variance efficiency: still living with the Roll critique, Federal Reserve Bank of Boston Working Paper No. QAU09-3.

[19] Rigobon, R., 2002, The curse of non-investment grade countries, Journal of Development Economics, 69, 423-449.

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

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

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

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

[24] Wooldridge, J.M., 1990, A unified approach to robust, regression-based specification tests, Econometric Theory, 6, 17-43.

TABLE 1. Simulation Results

GMM Estimator

		$\widehat{\gamma}$			
L	W_T	Med. Bias	MDAE	Dec. Range	SD
2	I	0.000	0.117	4.344	1.126
	$W\left(\widetilde{\Lambda}\right)$	0.000	0.093	1.167	1.223
4	$W\left(\widetilde{\Lambda}\right)$	0.069	0.131	0.567	0.252
8	$W\left(\widetilde{\Lambda}\right)$	0.118	0.136	0.414	0.167
16	$W\left(\widetilde{\Lambda}\right)$	0.143	0.147	0.330	0.130

Notes: Consider (1) and (2) where $\beta_{1,0} = \beta_{2,0} = 0$ and $\gamma_0 = 1$. Parameterize the errors according to D1 where $D = N$, the normal distribution. Let $a_{11,0} = 0.05$, $a_{22,0} = 0.04$, $a_{33,0} = 0.10$, $b_{11,0} = 0.90$, and $b_{22,0} = b_{33,0} = 0.80$. These values reflect the low ARCH and high GARCH terms typically encountered in empirical work. Values for C_0 are selected such that $Var\left[\epsilon_{1,t}\right] = Var\left[\epsilon_{2,t}\right] = 1$ and $Cov\left[\epsilon_{1,t},\ \epsilon_{2,t}\right] = 0.20$, while $\lambda_{12,0} = 0.91$ and $\lambda_{22,0} = 2.19$. All of the aforementioned parameter values are used as the starting values for the optimizing iterations in each simulation. Given these values, (3) has a representation as a diagonal BEKK(1,1) model according to Proposition 2.6 of Engle and Kroner (1995). This BEKK representation is used in the simulations to satisfy A3. Monte Carlo studies are conducted across 5000 trials for $T = 1260$ observations. This sample size corresponds to five years worth of trading days and is motivated by an application of $\widehat{\psi}$ in testing the CAPM as discussed by Prono (2009). Results for $\widehat{\gamma}$ are shown. Robust measures of central tendency and dispersion are reported because of concerns over the existence of moments. 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 $\widehat{\gamma}$. The standard deviation, while not a robust measure, is reported to give an indication of outliers.