MPRA
Munich Personal RePEc Archive

GARCH-Based Identification and Estimation of Triangular Systems

Todd, Prono

Commodity Futures Trading Commission

September 2009

Online at https://mpra.ub.uni-muenchen.de/27482/
MPRA Paper No. 27482, posted 17 Dec 2010 00:48 UTC

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

Todd Prono[2]
Commodity Futures Trading Commission

Revised December 2010

Abstract

The diagonal GARCH(1,1) model is shown to support identification of the triangular system and is argued as a second moment analog to traditional exclusion restrictions. Estimators for this result include QML and GMM. The GMM estimator contains many (potential weak) moment conditions that can be the source of bias. As a result, a jackknife GMM estimator is proposed that remains consistent in the presence of many such moments. A small Monte Carlo study of the GMM and jackknife GMM estimators is also included.

JEL Codes: C3, C13, C32. Keywords: Triangular models, heteroskedasticity, identification.

1. Introduction

Let $Y_t = \begin{bmatrix} Y_{1,t} & Y_{2,t} \end{bmatrix}'$, a vector of endogenous variables. Let X_t be a vector of predetermined variables that can include lags of Y_t, and let $\epsilon_t = \begin{bmatrix} \epsilon_{1,t}, & \epsilon_{2,t} \end{bmatrix}'$, a vector of unobserved errors. In addition, let $\theta_0 = \left\{\beta_0,\ \delta_0,\ \omega_{ij,0},\ a_{ij,0},\ b_{ij,0}\right\}\ \forall\, i,j = 1,2$ be a set of true parameter values. For the sequence $\{Y_t,\ X_t\}_{t\in\mathbb{Z}}$ with associated σ-algebra $\mathcal{F}_t$, consider the triangular system:

$$Y_{1,t} = X_t'\beta_{1,0} + Y_{2,t}\beta_{2,0} + \epsilon_{1,t} \equiv Z_t\beta_0 + \epsilon_{1,t} \tag{1}$$

$$Y_{2,t} = X_t'\delta_0 + \epsilon_{2,t} \tag{2}$$

[1] I owe gratitude to Robin Lumsdaine, Arthur Lewbel, David Reiffen, two anonymous referees, and seminar participants at the 2007 Summer Meeting of the Econometric Society for helpful comments and discussions.
[2] Office of the Chief Economist, 1155 21st Street, N.W., Washington, DC 20581 USA. Tel: (202) 418-5460 email: tprono@cftc.gov

The views expressed herein are solely those of the author and do not reflect official positions of the Commodity Futures Trading Commission. In addition, the usual disclaimer applies.

where the errors are correlated. I show that when there are no exclusionary restrictions available for $\beta_{1,0}$, it remains possible to identify β_0 if

$$E\left[\epsilon_t \mid \mathcal{F}_{t-1}\right] = 0, \quad E\left[\epsilon_t \epsilon_t^{'} \mid \mathcal{F}_{t-1}\right] = H_t = \left[h_{ij,t}\right], \tag{3}$$

where H_t is time varying, and

$$h_{ij,t} = h_{ij,t}\left(\theta_0\right) = \omega_{ij,0} + a_{ij,0}\epsilon_{i,t-1}\epsilon_{j,t-1} + b_{ij,0}h_{ij,t-1}\left(\theta_0\right). \tag{4}$$

(3) attributes conditional heteroskedasticity (CH) to ϵ_t. CH is necessary but not sufficient for identification of β_0. (4) assigns a particular GARCH functional form to the CH, the diagonal GARCH(1,1) model. Identification of β_0 derives from this particular GARCH functional form.

The univariate version of (4) was introduced by Bollerslev (1986) and extended into the multivariate setting by Bollerslev, Engle, and Wooldridge (1988). By nature of (4) being a diagonal model, exclusionary restrictions are imposed on all past off-diagonal squared errors and cross products of errors. These second moment exclusionary restrictions identify β_0.

2. Identification Source

The identification problem in (1) can be recast in terms of a control function as in Klein and Vella (2010). Doing so provides a heuristic basis for understanding how (3) and (4) solve this problem. Consider the conditional regression

$$A_0\left(\mathcal{F}_{t-1}\right) \equiv \arg\min_A E\left[\epsilon_{1,t} - A_0\epsilon_{2,t} \mid \mathcal{F}_{t-1}\right]^2 = Cov\left[\epsilon_{1,t},\ \epsilon_{2,t} \mid \mathcal{F}_{t-1}\right] / Var\left[\epsilon_{2,t} \mid \mathcal{F}_{t-1}\right].$$

In this case, $U_t \equiv \epsilon_{1,t} - A_0\left(\mathcal{F}_{t-1}\right)\epsilon_{2,t}$ is uncorrelated with $\epsilon_{2,t}$ conditional on $\mathcal{F}_{t-1}$ and forms the basis for the controlled regression

$$Y_{1,t} = Z_t\beta_0 + A_0\left(\mathcal{F}_{t-1}\right)\epsilon_{2,t} + U_t. \tag{5}$$

Let $V_t = \left[Z_t,\ \epsilon_{2,t}\right]$. Then, if ϵ_t is homoskedastic so that $A_0\left(\mathcal{F}_{t-1}\right)$ is constant, we have the usual identification problem, since (absent exclusionary restrictions for $\beta_{1,0}$) $E\left[V_t'V_t\right]$ is singular.[3] Now suppose, instead, that ϵ_t is CH, and let $W_t = \left[Z_t,\ A_0\left(\mathcal{F}_{t-1}\right)\epsilon_{2,t}\right]$. Then, $E\left[W_t'W_t\right]$ is nonsingular, and the identification problem is solved, provided that $A_0\left(\mathcal{F}_{t-1}\right)$ can be consistently estimated. This latter requirement necessitates (4) and illustrates why CH alone is not sufficient for identifying β_0.

One approach to make estimation of $A_0\left(\mathcal{F}_{t-1}\right)$ feasible is to assume a constant conditional covariance. Specifically, since $A_0\left(\mathcal{F}_{t-1}\right) = h_{12,t}\left(\theta_0\right)/h_{22,t}\left(\theta_0\right)$ given (4), if $h_{12,t}\left(\theta_0\right) = \omega_{12,0}$, then $A_0\left(\mathcal{F}_{t-1}\right)$ can be consistently estimated because $h_{22,t}\left(\theta_0\right)$ is parameterized as a univariate GARCH(1,1) model, and $\epsilon_{2,t}$ is identified provided that $E\left[X_tX_t'\right]$ is nonsingular. Sentana and Fiorentini (2001) employ this precise covariance restriction to identify a latent factor model, where univariate GARCH(1,1) processes characterize the conditional variances of the factors. Lewbel (2010) also relies upon a constant conditional covariance restriction for identifying triangular and simultaneous models. In a similar vein, Vella and Verbeek (1997) and Rummery et al. (1999), too, rely on a covariance restriction for identification by proposing rank order as an instrumental variable.

The contribution of this note is to allow $h_{12,t}\left(\theta_0\right)$ to be time-varying, parameterizing it as an ARMA(1,1) process, analogous to the specification of each conditional variance. Doing so complicates estimation of $A_0\left(\mathcal{F}_{t-1}\right)$ by requiring the control function to be treated simultaneously along with (5), since $h_{12,t}\left(\theta_0\right)$ now depends on past values of $\epsilon_{1,t}$. The functional form in (4) allows for this simultaneous estimation by permitting $\beta_{2,0}$ to be identified from the reduced form of $h_{12,t}\left(\theta_0\right)$. As is the case with traditional exclusionary restrictions imposed on $\beta_{1,0}$, identification from the reduced form of $h_{12,t}\left(\theta_0\right)$ results because of restrictions imposed on the structural form; specifically, the exclusion of past values of $\epsilon_{1,t}^2$ and $\epsilon_{2,t}^2$ from the parameterization of $h_{12,t}\left(\theta_0\right)$.

Klein and Vella (2010) is a work closely related to this one. They show identification of the triangular model given heteroskedastic errors of a semi-parametric functional form. Their estimator is more complicated to implement than the ones I propose, owing to the more general heteroskedastic specification. In many applications of financial economics, the more restrictive CH specification of (3) and (4) proves warranted (see, for example, Hansen

[3]Singularity follows from $\epsilon_{2,t}$ being a linear combination of $Y_{2,t}$ and X_t.

and Lunde 2005). Moreover, the Klein and Vella approach links the conditional covariance between errors directly to each conditional variance. In this note, by contrast, $h_{12,t}(\theta_0)$ is not a direct function of either $h_{11,t}(\theta_0)$ or $h_{22,t}(\theta_0)$.[4]

Other papers that exploit heteroskedasticity for identification include Rigobon (2003) and Rigobon and Sack (2003), where multiple unconditional variance regimes act as probabilistic instruments, and the correlation between structural errors is sourced to common, unobserved, shocks.

The estimators I propose in the next two sections simultaneously estimate (1) and (2) along with the specification for H_t given in (4) (or select autocovariances from that specification). They do not estimate (5). Estimators based on (1)–(4) versus ones based on (5) are equivalent in terms of their requirements for identification.

3. QML Estimation

For the model of (1)–(4), consider the following additional assumptions:

ASSUMPTION A1: $E[X_t X_t']$ and $E[X_t Y_t]$ are finite and identified from the data. $E[X_t X_t']$ is nonsingular.

ASSUMPTION A2: Let $H_t(\theta) = \left[h_{ij,t}(\theta)\right]$. $H_t(\theta)$ is positive definite almost surely.

ASSUMPTION A3(i): $\left\{\left(a_{ij},\ b_{ij}\right):\ a_{ij} > 0,\ b_{ij} \geq 0,\ a_{ij} + b_{ij} < 1\right\}$.

ASSUMPTION A3(ii): $\{(a_{12},\ a_{22}):\ a_{12} \neq a_{22}\}$.

In practice, A2 can be satisfied using the BEKK parameterization of (4) introduced by Engle and Kroner (1995).[5] A3(i) restricts ϵ_t to be covariance stationary. The condition $a_{ij} > 0$ ensures that $h_{ij,t}(\theta_0)$ is identified.[6] Allowing $b_{ij} = 0$ permits $H_t(\theta_0)$ to follow a diagonal ARCH(1) process. A3(ii) is an inequality restriction imposed on $H_t(\theta_0)$ that is necessary for the identification of $\beta_{2,0}$ (see the proof to Proposition 1 in the Appendix) and

[4] An example where $h_{12,t}(\theta_0)$ is a direct function of $h_{11,t}(\theta_0)$ and $h_{22,t}(\theta_0)$ is the CCC model of Bollerslev (1990).

[5] See Proposition 2.6 of the aforementioned work.

[6] If $a_{ij,0} = 0$, then $h_{ij,t}(\theta_0)$ is completely deterministic, and $\omega_{ij,0}$ and $b_{ij,0}$ are not separately identified.

generally illustrative of how parameter restrictions on the heteroskedastic process of $H_t(\theta_0)$ are necessary for identification of the triangular model.

For the sequence $\{Y_t,\ X_t\}_{t=1}^{T}$, let $\epsilon_{1,t}(\theta) = Y_{1,t} - Z_t\beta$, and $\epsilon_{2,t}(\theta) = Y_{2,t} - X_t'\delta$. For $l_t(\theta) \equiv l\left(Y_t,\ \mathcal{F}_{t-1};\ \theta\right)$, where

$$l\left(Y_t,\ \mathcal{F}_{t-1};\ \theta\right) = -1/2\log\left|H_t(\theta)\right| - 1/2\epsilon_t(\theta)' H_t(\theta)^{-1}\epsilon_t(\theta),$$

let $L_T(\theta) = \sum_{t=1}^{T} l_t(\theta)$. Consider the estimator

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} L_T(\theta) \tag{6}$$

PROPOSITION 1. *For the estimator in (6) of the model described by (1)–(4), let Assumptions A1–A3(ii) hold. Then θ_0 is identified.*

Proofs are in the Appendix. Let $e_t \equiv vech\left[\epsilon_t\epsilon_t'\right] = \left[\epsilon_{1,t}^2,\ \epsilon_{1,t}\epsilon_{2,t},\ \epsilon_{2,t}^2\right]'$.[7] Conditional on past values of e_t, $h_{11,t}$ is parameterized only to depend on past values of $\epsilon_{1,t}^2$, $h_{12,t}$ only to depend on past values of $\epsilon_{1,t}\epsilon_{2,t}$, and $h_{22,t}$ only to depend on past values of $\epsilon_{2,t}^2$. It is from these restrictions on the dynamics of $H_t(\theta_0)$ that identification follows, much in the same way that traditional identification of (1) follows from at least one element of the parameter vector $\beta_{1,0}$ being zero. Suppose that instead of the parameterization in (4), $h_{ij,t}$ were parameterized to depend on past values of every element in e_t. In this case, the matrices C_0 and D_0 in (12) would each relate nine reduced form parameters to ten structural unknowns (the nine ARCH (GARCH) parameters plus $\beta_{2,0}$), and (6) would not be identified because a necessary order condition would not be satisfied. In the language of section 2, this example is a case where identification is not achieved because the control function cannot be consistently estimated.

Proposition 3.1 of Engle and Kroner (1995) states that if a model's structural errors follow a GARCH process then so, too, will its reduced form errors. While the lag order of these two processes will coincide, their parametric forms, generally, will not (see Proposition 2.1 of Iglesias and Phillips 2004). Proposition 1 leverages off of this difference between structural

[7] The $vech(\cdot)$ operator stacks the lower triangle, including the diagonal, of a symmetric matrix into a column vector.

and reduced forms. Specifically, while $H_t(\theta_0)$ in (4) is a diagonal model, its reduced form in (12) contains nonzero off-diagonal terms. Identification of $\beta_{2,0}$ depends on these off-diagonal reduced form parameters. In discussing how the relationship between structural and reduced form GARCH models can identify simultaneous systems, Rigobon (2002) states that "the model of heteroskedasticity of the structural residuals impose[s] important constraints on how the reduced form heteroskedasticity can evolve" (p.433). In the context of Proposition 1, the "important constraints" are the exclusionary restrictions imposed on past values of e_t by the diagonal model in (4).

Under Proposition 1, the key identifying assumption is that $H_t(\theta_0)$ follows a diagonal GARCH process. The precise lag order of this diagonal process is unimportant. For instance, $H_t(\theta_0)$ can be specified as a diagonal GARCH(p, q) process with $p, q \geq 1$, and identification still follows. To aid in determining the lag order, the robust lagrange multiplier tests of Bollerslev and Wooldridge (1992) are applicable.

Given Proposition 1, consistency and asymptotic normality of (6) is established by Theorem 2.1 of Bollerslev and Wooldridge (1992). A standard regulatory condition for these results is compactness of Θ. This condition needs to be reconciled with A3(ii). One such reconciliation would be to redefine Θ so that a_{12}/a_{22} is exclusive of an open neighborhood of one.

4. GMM Estimation

Consider, again, the model of (1)–(4). For

$$\begin{aligned} h_t(\theta_0) &\equiv vech\left[H_t(\theta_0)\right] \\ &= \left[h_{11,t}(\theta_0),\ h_{12,t}(\theta_0),\ h_{22,t}(\theta_0)\right]' \end{aligned}$$

note that

$$e_t = h_t(\theta_0) + \eta_t, \tag{7}$$

where $E\left[\eta_t \mid \mathcal{F}_{t-1}\right] = 0$ and $E\left[\eta_t \eta_s'\right] = 0\ \forall\ t \neq s$. Let $\overline{e}_t = \left[\epsilon_{1,t}\epsilon_{2,t},\ \epsilon_{2,t}^2\right]'$, and consider analogous definitions for $\overline{h}_t(\theta_0)$ and $\overline{\eta}_t$, respectively. In addition, let $Z_{t-2} = \left[\overline{e}_{t-2}' \cdots \overline{e}_{t-L}'\right]'$

for a finite $L \geq 2$, where Z_{t-2} can be thought of as a vector of instruments for $\overline{h}_t(\theta_0)$, 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 A3(iii): Let $p_{11} = a_{12} + b_{12}$, and $p_{22} = a_{22} + b_{22}$. $\{(p_{11},\ p_{22}) :\ p_{11} \neq p_{22}\}$.

ASSUMPTION A4(i): $E\left[\overline{\eta}_t \overline{\eta}_t'\right] = \Sigma_{\overline{\eta}} < \infty$.

ASSUMPTION A4(ii) $Cov\left[\overline{e}_t,\ Z_{t-1}\right]$ has full row rank.

LEMMA. *Define A_0 (B_0) as a 2×2 diagonal matrix with $a_{12,0}$ and $a_{22,0}$ $(b_{12,0}$ and $b_{22,0})$ as the diagonal entrees. For the model of (3) and (4), let Assumptions A3(i) and A4(i) hold. Then $\overline{e}_t$ is covariance stationary and*

$$Cov\left[\overline{e}_t,\ \overline{e}_{t-\tau}\right] = (A_0 + B_0)\, Cov\left[\overline{e}_t,\ \overline{e}_{t-(\tau-1)}\right], \tag{8}$$

where $\tau \geq 1$.

This lemma is closely related to Theorem 3 of Hafner (2003) and establishes a subset of the autocovariances of squares and cross products of errors implied by (3) and (4). This subset of autocovariances is shown to provide the additional moment conditions necessary for identifying a GMM estimator of (1) and (2).

Let $\psi = \{\beta,\ \delta,\ \overline{\omega},\ P\}$, where $\overline{\omega} = [\omega_{12},\ \omega_{22}]'$, and $P = A + B$. Define Ψ as the set of all possible values for ψ. In addition, $\overline{\sigma} = [I - P]^{-1}\,\overline{\omega}$, where I is the identity matrix, and $z_{t-2}(\psi) = \left[\left(\overline{e}_{t-2}(\psi) - \overline{\sigma}\right)',\cdots,\ \left(\overline{e}_{t-L}(\psi) - \overline{\sigma}\right)'\right]'$. Consider the following vector valued functions

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

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

$$U_3\left(Y_t,\ \mathcal{F}_{t-1};\ \psi\right) = vec\left(\left(\overline{e}_t(\psi) - \overline{\sigma}\right) z_{t-2}'(\psi) - P\left(\overline{e}_t(\psi) - \overline{\sigma}\right) z_{t-1}'(\psi)\right),$$

where $\otimes$ is the Kronecker product, and $vec(\cdot)$ stacks the columns of a matrix into a column vector. Stack these functions into a single column vector $U\left(Y_t,\ \mathcal{F}_{t-1};\ \psi\right)$. With $U_t(\psi) \equiv$

$U\left(Y_t,\ \mathcal{F}_{t-1};\ \psi\right)$, one can construct Hansen's (1982) GMM estimator

$$\widehat{\psi} = \underset{\psi\in\Psi}{\arg\min}\ Q_T(\psi) = \left[T^{-1}\sum_{t=1}^{T}U_t(\psi)\right]' W_T \left[T^{-1}\sum_{t=1}^{T}U_t(\psi)\right], \tag{9}$$

for some sequence of positive definite W_T, where $T^{-1}\sum_{t=1}^{T}U_{3,t}(\psi)$ is a column vector of the sample autocovariances from (8).

The estimator in (6) estimates each element of the control function $A_0\left(\mathcal{F}_{t-1}\right)$. By contrast, the estimator in (9) estimates the autocovariances implied by each element in $A_0\left(\mathcal{F}_{t-1}\right)$. As seen in Proposition 2, however, identification of the triangular system remains the product of both the parameterizations of $h_{12,t}(\theta_0)$ and $h_{22,t}(\theta_0)$.

PROPOSITION 2. *For the estimator in (9) of the model described by (1), (2), and (8), let Assumptions A1–A3(i) and A3(iii)–A4(ii) hold. Then the only $\psi \in \Psi$ that satisfies $E[U_t(\psi)] = 0$ is $\psi = \psi_0$.*

If (a) $U_t(\psi)$ satisfies the UWLLN of Wooldridge (1990, Definition A.1), (b) $W_T \xrightarrow{p} W_0$, and (c) $\psi_0 \in \operatorname{int}\Psi$, a compact parameter space, then (9) can be shown to be weakly consistent given Proposition 2. Compactness under (c) needs to be reconciled with A3(iii). One possibility is to redefine Ψ so that p_{11}/p_{22} is exclusive of an open neighborhood of one.

(9) can also be shown to be asymptotically normal; however, $E\left[\|U_t(\psi_0)\|^2\right] < \infty$ is necessary. If such moment existence criteria prove overly restrictive, then bootstrap standard errors for $\widehat{\psi}$ are available through an application of the nonoverlapping block bootstrap method 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).

The autocovariance process in (8) is the key identifying assumption for (9). Since this process applies across all lags of $\overline{e}_t$, the vector of instruments Z_{t-2} used in defining the moment conditions $T^{-1}\sum_{t=1}^{T}U_{3,t}(\psi)$ can be quite large. As a consequence, ψ_0 is overidentified, and the standard test of overidentifying restrictions based on the GMM objective function is available. A non-parametric test of these overidentifying restrictions is also possible given the bootstrap method in Brown and Newey (2002).

The principal contribution of this section is the moment conditions in (8) used for identifying the triangular model. An estimator based on these moment conditions is (9). For large values of L, the resulting instrument vector Z_{t-2} produces many (potentially weak) moment conditions. In the case of many (weak) moments, Newey and Smith (2004) show that (9) can be biased. An alternative estimator,

$$\breve{\psi} = \arg\min_{\psi \in \Psi} Q_T(\psi) - T^{-1} tr \left[W_T \left(T^{-1} \sum_{t=1}^{T} U_t(\psi) U_t(\psi)' \right) \right], \tag{10}$$

which is the jackknife GMM (JGMM) estimator of Newey and Windmeijer (2009), remains consistent under many (potentially weak) moments by deleting the term responsible for the bias.[8] This estimator is likely to be preferable to (9) in instances where high values of L lead to large reductions in standard errors.

5. Monte Carlo

This section analyzes the finite sample performance of (9) and (10) benchmarked against the OLS estimator by considering the following simulation design:

$$\begin{aligned} Y_{1,t} &= X_{1,t} + Y_{2,t} + \epsilon_{1,t} \\ Y_{2,t} &= X_{1,t} + \epsilon_{2,t} \\ H_t(\theta_0)^{-1} \epsilon_t &= \zeta_t \sim N(0,\ I), \end{aligned}$$

where $a_{11,0} = a_{12,0} = 0.05$, $a_{22,0} = 0.10$, $b_{11,0} = 0.93$, $b_{12,0} = 0.80$, and $b_{22,0} = 0.85$. Conditional on these $a_{ij,0}$ and $b_{ij,0}$, the constants $\omega_{ij,0}$ are set so that $Var\left[\epsilon_{1,t}\right] = Var\left[\epsilon_{2,t}\right] = 1$, and $Cov\left[\epsilon_{1,t},\ \epsilon_{2,t}\right] = 0.20$. All simulations are conducted with 1,000 observations across 1,000 trials after dropping the first 200 observations to avoid initialization effects. For each trial using (9) and (10), the starting values are the true parameter values. Both (9) and (10) set $W_T = \left(T^{-1} \sum_{t=1}^{T} U_t\left(\dot{\psi}\right) U_t\left(\dot{\psi}\right)' \right)^{-1}$, where $\dot{\psi}$ is a preliminary estimator, and $L = 10$.[9]

[8] This JGMM estimator assumes that $U_t(\psi)$ follows a 1^{st} order Markov process. A generalization of (10) that allows $U_t(\psi)$ to follow higher order Markov processes is discussed in Prono (2010).

[9] Given the simulation design, ϵ_t is eighth moment stationary according to figure 1 of Bollerslev (1986).

Table 1 summarizes the results. The OLS estimator of the model for $Y_{1,t}$ is about 20% biased. The bias drops to about 9% for the GMM estimator. The JGMM estimator is unbiased. The GMM estimates are more dispersed than their OLS counterparts with smaller median absolute errors. The JGMM estimates are less dispersed and have smaller median absolute errors than OLS. In general, these simulation results provide evidence that (9) and (10) remedy the endogeneity bias of the triangular model. In addition, the JGMM estimator is shown to display less bias and higher efficiency than its GMM counterpart for a moderately large set of moment conditions.[10]

[10] For $L = 10$, $U_t(\psi)$ is composed of 40 moment conditions.

Appendix

PROOF OF PROPOSITION 1: Let $\pi_0 = \left\{\pi_{1,0},\ \pi_{2,0},\ \varpi_0,\ C_0,\ D_0\right\}$, where $\pi_{1,0}$ and $\pi_{2,0}$ are the reduced form parameter vectors to (1) and (2), respectively. The reduced form errors $R_{i,t}(\pi_0)$ are then

$$R_{i,t}(\pi_0) = Y_{i,t} - X_t' E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{i,t}\right],$$

which are identified given A1. Substitution of (1) and (2) for $Y_{1,t}$ and $Y_{2,t}$ into the definitions for $R_{1,t}(\pi_0)$ and $R_{2,t}(\pi_0)$ shows that

$$R_{1,t}(\pi_0) = \epsilon_{1,t} - \epsilon_{2,t}\beta_{2,0}, \quad R_{2,t}(\pi_0) = \epsilon_{2,t}. \tag{11}$$

Substitution of (11) into $H_t(\theta_0)$ shows that for $E\left[R_t(\pi_0) R_t(\pi_0)' \mid F_{t-1}\right] = H_t^{(r)}(\pi_0)$, the reduced form conditional variance-covariance matrix,

$$\begin{aligned} h_t^{(r)}(\pi_0) &\equiv vech\left[H_t^{(r)}(\pi_0)\right] = \left[h_{11,t}^{(r)}(\pi_0),\ h_{12,t}^{(r)}(\pi_0),\ h_{22,t}^{(r)}(\pi_0)\right]' \\ &= \varpi_0 + C_0 vech\left[R_{t-1}(\pi_0) R_{t-1}(\pi_0)'\right] + D_0 h_{t-1}^{(r)}(\pi_0), \end{aligned} \tag{12}$$

where $C_0 = \left[c_{kl,0}\right]$ and $D_0 = \left[d_{kl,0}\right]$ for $k, l = 1, 2, 3$. Consider

$$l_t(\pi) = -1/2 \log\left|H_t^{(r)}(\pi)\right| - 1/2 R_t(\pi)' H_t^{(r)}(\pi)^{-1} R_t(\pi),$$

and $L_T(\pi) = \sum_{t=1}^{T} l_t(\pi)$. Given A1, A2, and A3(i), π_0 is a maximizer of $E\left[L_T(\pi)\right]$ that is identifiably unique according to Lemma A.2 and condition A.1(iii)(b) in Bollerslev and Wooldridge (1992). From (12),

$$\begin{aligned} h_{12,t}^{(r)}(\pi_0) &= \varpi_{21,0} + c_{22,0} R_{1,t-1}(\pi_0) R_{2,t-1}(\pi_0) + c_{23,0} R_{2,t-1}^2(\pi_0) \\ &\quad + d_{22,0}\ h_{12,t-1}^{(r)}(\pi_0) + d_{23,0} h_{22,t-1}^{(r)}(\pi_0) \end{aligned}$$

where $c_{22,0} = a_{12,0}$, $c_{23,0} = \left(a_{22,0} - a_{12,0}\right)\beta_{2,0}$, $d_{22,0} = b_{12,0}$, and $d_{23,0} = \left(b_{22,0} - b_{12,0}\right)\beta_{2,0}$. Since $h_{22,t}^{(r)}(\pi_0) = h_{22,t}(\theta_0)$ given (11), $c_{33,0} = a_{22,0}$ and $d_{33,0} = b_{22,0}$. As a result, $\beta_{2,0}$ is identified as

$$\beta_{2,0} = \frac{c_{23,0}\left(c_{33,0} - c_{22,0}\right) + d_{23,0}\left(d_{33,0} - d_{22,0}\right)}{\left(c_{33,0} - c_{22,0}\right)^2 + \left(d_{33,0} - d_{22,0}\right)^2}$$

given A3(ii). Since $\pi_{2,0} = \delta_0$, $\beta_{1,0}$ is identified conditional on $\beta_{2,0}$. Since $\varpi_{31,0} = \omega_{22,0}$, $\omega_{12,0}$ is also identified conditional on $\beta_{2,0}$. The structural parameters to $h_{11,t}(\theta_0)$ are then identified conditional on $\beta_{2,0}$, $\omega_{12,0}$, $\omega_{22,0}$, $a_{12,0}$, $a_{22,0}$, $b_{12,0}$, and $b_{22,0}$.■

PROOF OF THE LEMMA: Let $\overline{h}_t = \overline{h}_t(\theta_0)$. Given (4) and the definitions of $\overline{e}_t$ and $\overline{h}_t$, it follows that

$$\overline{h}_t = \overline{\omega}_0 + A_0 \overline{e}_{t-1} + B_0 \overline{h}_{t-1}. \tag{13}$$

Recursive substitution into (13) produces

$$\overline{h}_t = \sum_{i=1}^{\infty} B_0^{i-1} \left(\overline{\omega}_0 + A_0 \overline{e}_{t-i} \right). \tag{14}$$

Following the steps outlined in the proof to Proposition 2.7 of Engle and Kroner (1995), (14) can be used to show that

$$E\left[\overline{e}_t \mid F_{t-\tau}\right] = \left[I + (A_0 + B_0) + \cdots + (A_0 + B_0)^{\tau-2}\right] \overline{\omega}_0 + (A_0 + B_0)^{\tau-1} \sum_{i=1}^{\infty} B_0^{i-1} \left(\overline{\omega}_0 + A_0 \overline{e}_{t-i-\tau+1} \right).$$

For a square matrix Z, it is well known that $(I + Z + \cdots + Z^{\tau-1}) \to (I - Z)^{-1}$ as $\tau \to \infty$ if and only if the eigenvalues of Z are less than one in modulus. Therefore, $E\left[\overline{e}_t \mid F_{t-\tau}\right] \xrightarrow{p} [I - (A_0 + B_0)]^{-1} \overline{\omega}_0$ (as $\tau \to \infty$) given A3(i).

From (7),

$$E\left[\overline{e}_t \overline{e}_t^{\prime}\right] = E\left[\overline{h}_t \overline{h}_t^{\prime}\right] + \Sigma_{\overline{\eta}}$$

given A4(i). Let $\overline{\sigma}_0 = [I - (A_0 + B_0)]^{-1} \overline{\omega}_0$.

$$\begin{aligned} E\left[\overline{h}_t \overline{h}_t^{\prime}\right] &= \kappa_0 + A_0 E\left[\overline{h}_{t-1} \overline{h}_{t-1}^{\prime}\right] A_0 + A_0 \Sigma_{\overline{\eta}} A_0 + A_0 E\left[\overline{h}_{t-1} \overline{h}_{t-1}^{\prime}\right] B_0 \\ &\quad + B_0 E\left[\overline{h}_{t-1} \overline{h}_{t-1}^{\prime}\right] B_0 + B_0 E\left[\overline{h}_{t-1} \overline{h}_{t-1}^{\prime}\right] B_0 \end{aligned} \tag{15}$$

where $\kappa_0 = \overline{\omega}_0 \overline{\omega}_0^{\prime} + (A_0 + B_0) \overline{\sigma}_0 \overline{\omega}_0^{\prime} + \overline{\omega}_0 \overline{\sigma}_0^{\prime} (A_0 + B_0)$. Applying the $vec(\cdot)$ operator, which stacks the columns of a matrix into a column vector, to (15) and simplifying yields

$$
\begin{aligned}
vec\left(E\left[\overline{h}_t\overline{h}_t^{'}\right]\right) &= \kappa_0 + G_0 vec\left(E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\right) + (A_0 \otimes A_0)\, vec\left(\Sigma_{\overline{\eta}}\right) \\
&= [I + G_0]\left(\kappa_0 + (A_0 \otimes A_0)\, vec\left(\Sigma_{\overline{\eta}}\right)\right) + \left(G_0^2\right) vec\left(E\left[\overline{h}_{t-2}\overline{h}_{t-2}^{'}\right]\right) \\
&= \left[I + G_0 + G_0^2\right]\left(\kappa_0 + (A_0 \otimes A_0)\, vec\left(\Sigma_{\overline{\eta}}\right)\right) + \left(G_0^3\right) vec\left(E\left[\overline{h}_{t-3}\overline{h}_{t-3}^{'}\right]\right) \\
&= \ldots \\
&= \left[I + G_0 + \cdots + G_0^{\tau-1}\right]\left(\kappa_0 + (A_0 \otimes A_0)\, vec\left(\Sigma_{\overline{\eta}}\right)\right) + \left(G_0^{\tau}\right) vec\left(E\left[\overline{h}_{t-\tau}\overline{h}_{t-\tau}^{'}\right]\right)
\end{aligned}
$$

where $G_0 = (A_0 + B_0) \otimes (A_0 + B_0)$, and $\otimes$ is the Kronecker product. Therefore, $vec\left(E\left[\overline{h}_t\overline{h}_t^{'}\right]\right)$ converges to $[I - G_0]^{-1}\left(\kappa_0 + (A_0 \otimes A_0)\, vec\left(\Sigma_{\overline{\eta}}\right)\right)$ as $\tau \to \infty$ given A3(i).

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 $\tau = 1$.

$$
E\left[\overline{e}_t\overline{e}_{t-1}^{'} \mid F_{t-1}\right] = \overline{\omega}_0\overline{e}_{t-1}^{'} + A_0\overline{e}_{t-1}\overline{e}_{t-1}^{'} + B_0\overline{h}_{t-1}\overline{e}_{t-1}^{'}.
$$

By iterated expectations,

$$
E\left[\overline{e}_t\overline{e}_{t-1}^{'}\right] = \overline{\omega}_0\overline{\sigma}_0' + (A_0 + B_0)\,\Sigma_{\overline{h}} + A_0\Sigma_{\overline{\eta}}
$$

and, as a result,

$$
Cov\left[\overline{e}_t,\ \overline{e}_{t-1}\right] = (\overline{\omega}_0 - \overline{\sigma}_0)\,\overline{\sigma}_0' + (A_0 + B_0)\,\Sigma_{\overline{h}} + A_0\Sigma_{\overline{\eta}}
$$

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

$$
\begin{aligned}
E\left[\overline{h}_t \mid S_{t-\tau}\right] &= E\left[\overline{\omega}_0 + A_0 \overline{e}_{t-1} + B_0 \overline{h}_{t-1} \mid \digamma_{t-\tau}\right] \\
&= \overline{\omega}_0 + (A_0 + B_0) E\left[\overline{h}_{t-1} \mid \digamma_{t-\tau}\right] \\
&= \left[I + (A_0 + B_0)\right] \overline{\omega}_0 + (A_0 + B_0)^2 E\left[\overline{h}_{t-2} \mid \digamma_{t-\tau}\right] \\
&= \ldots \\
&= \left[I + (A_0 + B_0) + \ldots + (A_0 + B_0)^{\tau-1}\right] \overline{\omega}_0 + (A_0 + B_0)^{\tau-1}\left[A_0 \overline{e}_{t-\tau} + B_0 \overline{h}_{t-\tau}\right] \\
&= \left[I - (A_0 + B_0)^{\tau}\right] \overline{\sigma}_0 + (A_0 + B_0)^{\tau-1}\left[A_0 \overline{e}_{t-\tau} + B_0 \overline{h}_{t-\tau}\right].
\end{aligned}
$$

By iterated expectations,

$$
\begin{aligned}
E\left[\overline{e}_t \overline{e}_{t-\tau}^{'}\right] &= E\left[E\left[\overline{e}_t \overline{e}_{t-\tau}^{'} \mid \digamma_{t-\tau}\right]\right] \\
&= E\left[E\left[\overline{h}_t \mid \digamma_{t-\tau}\right] \overline{e}_{t-\tau}^{'}\right] \\
&= \left[I - (A_0 + B_0)^{\tau}\right] \overline{\sigma}_0 \overline{\sigma}_0^{\prime} + (A_0 + B_0)^{\tau-1}\left[(A_0 + B_0) E\left[\overline{h}_{t-\tau} \overline{h}_{t-\tau}^{'}\right] + A_0 E\left[\overline{\eta}_{t-\tau} \overline{\eta}_{t-\tau}^{'}\right]\right].
\end{aligned}
$$

As a result,

$$
Cov\left[\overline{e}_t,\ \overline{e}_{t-\tau}\right] = (A_0 + B_0)^{\tau-1}\left[(A_0 + B_0)\left(\Sigma_{\overline{h}} - \overline{\sigma}_0 \overline{\sigma}_0^{\prime}\right) + A_0 \Sigma_{\overline{\eta}}\right], \tag{16}
$$

from which (8) follows.■

PROOF OF PROPOSITION 2: Given (8),

$$
Cov\left[\overline{e}_t,\ Z_{t-2}\right] = (A_0 + B_0) \, Cov\left[\overline{e}_t,\ Z_{t-1}\right]. \tag{17}
$$

Substitution of (11) into (17) produces the reduced form autocovariance relation

$$
Cov\left[\overline{r}_t, Z_{t-2}^{(r)}\right] = \left(\overline{C}_0 + \overline{D}_0\right) Cov\left[\overline{r}_t,\ Z_{t-1}^{(r)}\right],
$$

where $\overline{r}_t = \left[R_{1,t} R_{2,t},\ R_{2,t}^2\right]'$, $Z_{t-2}^{(r)} = \left[\overline{r}_{t-2}^{\prime},\ \ldots,\ \overline{r}_{t-L}^{\prime}\right]'$, $\overline{C}_0 = \left[c_{kl,0}\right]$ and $\overline{D}_0 = \left[d_{kl,0}\right]$

for $k, l = 2, 3$. Define $\Gamma^{(r)}(\tau) \equiv Cov\left[\overline{r}_t, Z^{(r)}_{t-\tau}\right]$. Then

$$\left(\overline{C}_0 + \overline{D}_0\right) = \Gamma^{(r)}(2)\,\Gamma^{(r)}(1)'\left(\Gamma^{(r)}(1)\,\Gamma^{(r)}(1)'\right)^{-1} \tag{18}$$

given A4(ii). From (18), $c_{22,0} + d_{22,0} = a_{12,0} + b_{12,0}$, $c_{33,0} + d_{33,0} = a_{22,0} + b_{22,0}$, and

$$\beta_{2,0} = \frac{c_{23,0} + d_{23,0}}{\left(c_{33,0} + d_{33,0}\right) - \left(c_{22,0} + d_{22,0}\right)}$$

given A3(iii). Conditional on $\beta_{2,0}$,

$$\beta_{1,0} = E\left[X_t X_t'\right]^{-1} E\left[X_t\left(Y_{1,t} - Y_{2,t}\beta_{2,0}\right)\right] \tag{19}$$

given A1. Conditional on $a_{12,0} + b_{12,0}$ and $\beta_{2,0}$, $\omega_{12,0} = \left(a_{12,0} + b_{12,0}\right)^{-1} E\left[\epsilon_{1,t}\epsilon_{2,t}\right]$. Conditional on $a_{22,0} + b_{22,0}$ and $\beta_{2,0}$, $\omega_{22,0} = \left(a_{22,0} + b_{22,0}\right)^{-1} E\left[\epsilon^2_{2,t}\right]$.■

References

[1] Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 307–327.

[2] 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.

[3] 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.

[4] 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.

[5] Brown, B.W. and W.K. Newey, 2002, Generalized method of moments, efficient bootstrapping, and improved inference, Journal of Business and Economic Statistics, 20, 507-571.

[6] 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.

[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.P, 1982, Large sample properties of generalized method of moments estimators, Econometrica, 50, 1029-1054.

[11] Hansen, P.R. and A. Lunde, 2005, A forecast comparison of volatility models: does anything beat a GARCH(1,1)?, Journal of Applied Econometrics, 20, 873-889.

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

[13] Klein, R. and F. Vella, 2010, Estimating a class of triangular simultaneous equations models without exclusion restrictions, Journal of Econometrics, 154, 154-164.

[14] Lewbel, A., 2010, Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models, unpublished manuscript.

[15] Newey, W.K. and R.J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators, Econometrica, 72, 219-255.

[16] Newey, W.K and F. Windmeijer, 2009, Generalized method of moments with many weak moment conditions, Econometrica, 77, 687-719.

[17] Prono, T., 2010, Simple GMM estimation of the semi-strong GARCH(1,1) model, unpublished manuscript.

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

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

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

[21] Rummery, S., F. Vella and M. Verbeek, 1999, Estimating the returns to education for Australian youth via rank-order instrumental variables, Labour Economics, 6, 491-507.

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

[23] Vella, F. and M. Verbeek, 1997, Rank order as an instrumental variable, unpublished manuscript.

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

TABLE 1

SIMULATION RESULTS

		Estimator		
Para.	Stat.	OLS	GMM	JGMM
β_1	Med. Bias	-0.208	-0.090	-0.002
	MDAE	0.208	0.145	0.007
	Dec. Rge.	0.145	0.527	0.033
	SD	0.058	0.219	0.026
β_2	Med. Bias	0.206	0.082	0.011
	MDAE	0.206	0.135	0.015
	Dec. Rge.	0.122	0.517	0.055
	SD	0.048	0.216	0.207
δ	Med. bias	0.002	-0.001	0.001
	MDAE	0.020	0.024	0.012
	Dec. Rge.	0.075	0.092	0.074
	SD	0.030	0.046	0.091

Notes: The true parameter vector is $\beta_{10} = \beta_{20} = \delta_0 = 1$. (J)GMM is the (jackknife) two-step generalized method of moments estimator with $L = 10$ and the optimal weighting matrix. Med. Bias is the median bias, MDAE the median absolute error, and SD the standard deviation of the estimates. DR is the decile range of the estimates, measured as the difference between the 90th and 10th percentiles.