Non-Nested Testing in Models Estimated
via Generalized Method of Moments[1]

Alastair R. Hall

University of Manchester[2]

and

Denis Pelletier

North Carolina State University[3]

May 20, 2015

[1]xxxxxxxx.

[2]Corresponding author: Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK. Email: alastair.hall@manchester.ac.uk

[3]Department of Economics, Box 8110, North Carolina State University, Raleigh, NC 27695-8110, USA. Email: denis_pelletier@ncsu.edu

Abstract

We analyze the limiting distribution of **?** statistic for choosing between two competing dynamic models based on a comparison of GMM minimands. It is shown that: (i) if both models are misspecified then the statistic has a standard normal distribution under the null hypothesis of equal fit but the ranking could be determined by the choice of the weighting matrix; (ii) if both models are correctly specified or locally misspecified then the limiting distribution of the test statistic is non-standard under the null.

1 Introduction

Competing economics theories often lead to econometric models that are non-nested in the sense that one model is not obtained as a special case of the other. It is, therefore, of interest to develop statistical procedures that discriminate between non-nested models. In these circumstances, it may be considered attractive to have some method that allows the researcher to determine which - if either - of the two models is closer to the truth in some sense. **?** provides such a test for models estimated by Quasi Maximum Likelihood (QML). More recently, **?** extend **?** approach to provide a very general framework for the comparison of two competing dynamic models. In this more general context, inference is based on a test statistic that compares measures of goodness of fit for the two models; one model is preferred if its goodness of fit is statistically significantly smaller than its competitor. **?** provide generic conditions under which the statistic has a limiting standard normal distribution under the null hypothesis that both models are "equally good", a concept that is defined below. These generic conditions are very general and it is argued that they cover the situation in which the competing models are estimated via GMM and then compared using either the GMM minimands employed in the estimations or GMM type minimands that are different from those used in the estimation.[1] In spite of this seeming generality, **?** show that the aforementioned distributional result rests crucially on the assumption that a certain variance is non-zero.

In this paper, we investigate whether these generic conditions in fact cover GMM estimators and minimands. It turns out that the analysis depends crucially on whether the models in question are correctly specified, locally misspecified or non-locally misspecified. It is shown that if both models are correctly specified or locally misspecified then **?** generic conditions are not satisfied and the statistic does not converge to a limiting normal distribution but to a non-standard distribution that is a function of nuisance parameters, some of which may not be consistently estimable. However, if both models are non-locally misspecified then the generic conditions are satisfied and the **?** statistic does converge to the limiting standard normal distribution. The latter result indicates that there is scope for using the Rivers and Vuong statistic to compare two misspecified models estimated via GMM although caution may need to be exercised in its use because the outcome can depend on the choice of weighting matrix. Thus our results reveal important limitations to the use of GMM minimands for model selection in this way, which in turn contribute to the wider literature on model selection via goodness of fit in econometrics; see *e.g.* **?**, **?** and **?**.

---

[1]See **?**[p.3 and p.13].

2 Model Analysis

In this section we introduce the Realized GARCH model. They key interest resides on the conditional variance, $H_t = var(r_t|\mathcal{F}_{t-1})$, where r_t is a vector of time series of returns, $\mathcal{F}_{t-1}$ is the filtration till time t-1. In the GARCH (1, 1) model the conditional variance is a function of lagged itself and squared daily return. In the present framework, Ht will depend on X_{t-1}, which represents a realized measure of volatility, such as matrix of realized covariance, bipower variation, realized kernel. The measurement equation link the realized measure to the latent make the model complete. Thus the Realized GARCH model fully specifies the dynamic system of daily returns and realized measure. Assuming we have K assets, r_t the $K \times 1$ vector of returns for the K assets in day t. H_t is the covariance matrix of returns. We will assume $E(r_t|\mathcal{F}_{t-1}) = 0$, which is a legitimate assumption from the empirical work. Otherwise we could always reinterpret rt as the return minus its constant mean to meet this assumption.

The main challenges in multivariate volatility modelling is to ensure that the conditional covariance matrix is positive semidefinite. The BEKK model proposed by Engle and Kroner(1995) could approach this challenge . We could specify a particular BEKK-type system as follows:

$$P_t = H_t^{1/2} \xi_t H_t^{1/2} \tag{1}$$

where

$$\log H_t = CC' + B \log H_{t-1} B + A \log X_{t-1} A \tag{2}$$

and

$$\log x_t = D + \Gamma \log h_t + u_t \tag{3}$$

Suppose we have K assets. We refer to first two equation as the return equation and the Realized GARCH equation. Since the realized measure X_t could be interpreted as a measurement of conditional covariance H_t, the last equation is referred as the measurement equation. In (1), $\xi_t = z_t z_t'$ is $K \times K$ symmetric innovation matrix which is outer product of innovation vector z_t,satisfying $E_{t-1}[\xi_t] = I_k$, where I_k is an identity matrix. The innovation vector $z_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. In (2), we define the symmetric square root of a positive semidefinite matrix H, denoted by $H^{1/2}$, using the spectral decomposition such that $H^{1/2} = U\Lambda^{1/2}U$ where U is a matrix containing the eigenvectors of H, and $\Lambda^{1/2}$ is a diagonal matrix containing the square root of the eigenvalues of H.

We use $\log$ before matrix to represent the principal matrix logarithm. For most matrices,$\log(\exp(A)) =$

$A = \exp(\log(A))$. The $K \times K$ matrices B, A each have K^2 free parameters, while C is $K \times K$ a lower triangular matrix with $K^* = K(K+1)/2$ free parameters. The parameterization in (2) guarantee that $\log H_t$ is positive semidefinite for all t assuming the starting value $\log H_0$ is positive semidefinite. Moreover, if C is full rank matrix, then $\log H_t$ is positive definite for all t. The coefficient matrix B determines the persistence of the Realized MGARCH model. To make sure the covariance stationarity, the eigenvalues of the coefficient matrix B must be less than one in modulus. We define $x_t = vec(\log X_t)$ and similarly $h_t = vec(\log H_t)$ where the vec oprator stacks the matrix into a column vector. Since $\log Xt$ and $\log Ht$ are supposed to be symmetric matrix, the vec operator vectorize the low triangular part including the main diagonal of a $K \times K$ symmetric matrix into a $K^* \times 1$ vector. Hence, x_t, D, h_t and u_t are all $K^* \times 1$ vectors . Moreover, the innovation $u_t \overset{i.i.d}{\sim} \mathcal{N}(\mu, \delta^2)$. The coefficient matrix Γ is a $K^* \times K^*$ matrix. This matrix characterize the correlation between realized measure and all the elements of conditional covariance matrix. Specifically it forms a seemingly unrelated regression between the two variables. As a result, the equation (3) will have $O(K^4)$ parameters, which will suffer the curse-of-dimensionality problem. Since the measurement equation could be estimated with linear regression, which is an easy task in computer. So the problem of high dimensionality is not about burden of computation but the prediction variation produced by too many parameters.

Even the BEKK-type parameterization in (2) has $O(K^2)$ parameters. To avoid the curse-of-dimensionality problem here we could impose that A and B are diagonal matrices, which yields the diagonal Realized MGARCH model. In this case, the equations for each element of $\log H$ would characterize a bunch of univariate Realized GARCH models in which the conditional variance or covariance is driven by their own lags and the corresponding realized measure. If the elements of A, B are unrestricted which is a full parameterization, the evolution of every element in $\log Ht$ will be influenced by their own lags as well as every cross-asset effects Realized MGARCH model.

For demonstration, we write out the 2 by 2 case of (2)

$$
\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{pmatrix} = \begin{pmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{pmatrix}'
$$

$$
+ \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} \begin{pmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{pmatrix} \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}
$$

$$
+ \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} x_{11,t-1} & x_{12,t-1} \\ x_{21,t-1} & x_{22,t-1} \end{pmatrix} \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}
$$

We could write out (3) of 2 by 2 case in vec form

$$
\begin{pmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} d_{11,t} \\ d_{21,t} \\ d_{31,t} \end{pmatrix} + \begin{pmatrix} \gamma_{11,t} & \gamma_{12,t} & \gamma_{13,t} \\ \gamma_{21,t} & \gamma_{22,t} & \gamma_{23,t} \\ \gamma_{21,t} & \gamma_{22,t} & \gamma_{33,t} \end{pmatrix} \begin{pmatrix} h_{11,t} \\ h_{21,t} \\ h_{31,t} \end{pmatrix} + \begin{pmatrix} u_{11,t} \\ u_{21,t} \\ u_{31,t} \end{pmatrix}
$$

We could tell from the example that in the measurement equation, each element in the conditional covariance matrix in function of every component in the realized covariance. Clearly the measurement equation is over-parameterized. We could use shrinkage and variable selection method in the estimation to reduce elements in coefficient matrix Γ.

Note that, the (2)could easily written in its vec form

$$
\log h_t = \bar{C} + \bar{B} \log h_{t-1} + \bar{A} \log x_{t-1} \tag{4}
$$

Which $\bar{B} = L_k(B \otimes B)D_k$ and $\bar{A} = L_k(A \otimes A)D_k$,C is $(K^* \times 1)$ vector, while B and A are $(K^* \times K^*)$ matrices. The elimination and duplication matrices, L_k and D_k are operational matrices of zeros and ones, so the parameters in (4) are just linear transformation from the equation (2). If we insert (3) into the (4), we could observed that the $log(h_t) \sim AR(1)$ and $log(h_t) \sim ARMA(1,1)$ process. Specifically,

$$
\log h_t = C + \bar{A}D + (\bar{B} + \bar{A}\Gamma) \log h_{t-1} + \Gamma u_{t-1} \tag{5}
$$

$$
\log x_t = D + \Gamma\bar{C} - \bar{B}D + (\bar{B} + \bar{A}\Gamma) \log x_{t-1} + u_t - \bar{B}u_{t-1} \tag{6}
$$

We could easily generalize this to AR(p) and ARMA(p,q) for any p,q are positive integer.

Note the $\bar{B}u_{t-1}$ induce an additional stochastic component that, so the Realized MGARCH model would form a stochastic volatility structure. In the multiple-step-ahead predictions, the Realized M-GARCH is like a stochastic volatility since we don't know the future values of u_t. However, in the one-step-ahead predictions, u_{t-1} is $\mathcal{F}_{t-1}$ measurable, so it is simply a GARCH structure model. The coefficient matrix$(\bar{B} + \bar{A}\Gamma)$ determines the persistence of the Realized MGARCH system. To ensure the stationary condition of covariance matrix, the eigenvalues of this matrix must be less than one in modulus. In the following assumption we explicitly state this covariance stationarity condition, where for any $K \times K$ matrix A with eigenvalues $\lambda_1, \ldots, \lambda_k$ for all $k \in \mathbb{Z}^+, \rho(A) := \max_i |\lambda_i|$

Assumption 1 *In the Realized MGARCH model given by (5) and (6), $\rho(\bar{B} + \bar{A}\Gamma) \leq 1$.*

3 Estimation and Inference

3.1 Estimation for MGARCH Equation

In this section we discuss the asymptotic properties of the quasi-maximum likelihood estimator within (2). In GARCH MODELS, the vector of daily returns is usually modeled as $R_t = Ht^{1/2}z_t$, with $z_t \stackrel{i.i.d}{\sim} \mathcal{N}(0,1)$, which motivates quasi-maximum likelihood estimation (QMLE).In our model, since $\xi_t = z_t z_t'$, the assumption that $z_t \stackrel{i.i.d}{\sim} \mathcal{N}(0,1)$ implies that ξ follows a Wishart distribution. The choice of Wishart distribution is suitable for the model where the support of the random variable of interest is confined to the space of positive semidefinite matrices.

The Realized MGARCH model is parameterized with a finite-dimensional $(\delta \times 1)$ parameter vector $\theta \in \Theta \subset \mathbb{R}^\delta$. Decompose $\theta = (\theta'_H, \theta'_x)'$, where θ'_H the denote the parameters in Realized MGARCH equation, θ'_X denote the parameters in measurement equation. We will estimate θ'_H with QMLE. The log-likelihood for the tth observation will be denoted by $l_{H,t}(\theta_H)$. Inference for the Realized MGARCH equation will be based on QMLE of the following log-likelihood function of Wishart density which is similar with HEAVY model in Noureldin, etc (2011).

$$l_{H,t}(\theta_H) = c_H - \frac{1}{2}(\log |H_t| + tr(H_t^{-1}P_t)) \tag{7}$$

where C_H is a constant with respect to θ_H We assume the initial values H_0 is known and is positive semidefinite. We also assume that θ_H is variation free in the sense of Engle et al. (1983), which allows for equation-by-equation estimation. This assumption will simplify estimation and inference. The QML estimator is $\hat{\theta}_H$, where

$$\hat{\theta}_H = \arg\max_{\theta_H \in \Theta} L_{H,t}(\theta_H)$$

where $L_{H,t}(\theta_H) = \sum_{t=1}^T l_{H,t}(\theta_H)$

For the BEKK model, Comte and Lieberman (2003) show strong consistency of QMLE by verifying the conditions given in Jeantheau (1998). To establish the strong consistency results, we assume the Realized MGARCH model admit a strictly stationary and ergodic solution.