Statistica Sinica **15**(2005), 303-323

ON CANONICAL ANALYSIS OF MULTIVARIATE TIME SERIES

Wanli Min and Ruey S. Tsay

IBM and University of Chicago

Abstract: Canonical correlation analysis has been widely used in the literature to identify the underlying structure of a multivariate linear time series. Most of the studies assume that the innovations to the multivariate system are Gaussian. On the other hand, there are many applications in which the normality assumption is either questionable or clearly inadequate. For example, most empirical time series in business and finance exhibit conditional heteroscedasticity and have high excess kurtosis. In this paper, we establish some asymptotic results for canonical correlation analysis of multivariate linear time series when the data possess conditional heteroscedasticity. We show that for correct identification of a multivariate time series model, it is essential to use a modification, which we prescribe, to a commonly used test statistic for testing zero canonical correlations. We also use simulation to study the efficacy of the modification, and apply the modified test statistics to analyze daily log returns of three assets.

Key words and phrases: Canonical correlation, Central Limit Theorem, Hankel matrix, Kronecker index.

1. Introduction

Many statistical applications involve analysis of multivariate time series data, and dynamic linear vector models are often used to model the underlying structure of the series. The models used include the state-space model and the vector autoregressive moving-average (VARMA) model. In building such models, an analyst often encounters the problems of "curse of dimensionality" and "explosion in the number of parameters". In addition, use of un-constrained dynamic linear models may encounter the difficulty of "identifiability", such as the existence of exchangeable models and redundant parameters discussed in Tiao and Tsay (1989). To overcome the difficulties, various methods of structural specification have been proposed in the literature. Two useful methods to achieve structural simplification of a linear vector process are Kronecker indexes and scalar component models; see Hannan and Deistler (1988) and Tiao and Tsay (1989). For both methods, canonical correlation analysis is often the statistical tool used to identify the simplifying structure of the data.

Since proposed by Hotelling (1936), canonical correlation analysis has been widely applied in many statistical areas, especially in multivariate analysis. Time series analysis is no exception. Box and Tiao (1977) proposed a canonical analysis of vector time series that can reveal the underlying structure of the data to aid model interpretation. In particular, they showed that linear combinations of several unit-root non-stationary time series can become stationary. This is the idea of co-integration that was popular among econometricians in the 1990s after the publication of Engle and Granger (1987). Tsay and Tiao (1985) applied canonical correlation analysis to develop the smallest canonical correlation (SCAN) method for identifying univariate ARMA model for a stationary and/or non-stationary time series. Tiao and Tsay (1989) introduced the concept of scalar component models to build a parsimonious VARMA model for a given multivariate time series. Again, canonical correlation analysis was used extensively to search for scalar component models. Many other authors also used canonical analysis in time series analysis. See, for instance, Quenouille (1957), Robinson (1973), Akaike (1976), Cooper and Wood (1982) and Jewell and Bloomfield (1983).

To build a model for a k-dimensional linear process, it suffices to identify the k Kronecker indexes or k linearly independent scalar component models, because we can use such information to identify those parameters that require estimation and those that can be set to zero within a dynamic linear vector model. Simply put, the Kronecker indexes and scalar component models can overcome the difficulties of curse of dimensionality, parameter explosion, exchangeable models, and redundant parameters in modeling a linear vector time series. For simplicity, we consider the problem of specifying Kronecker indexes in this paper. The issue discussed, however, is equally applicable to specification of scalar component models.

The method of determining Kronecker indexes of a linear vector process with Gaussian innovations has been studied by Akaike (1976), Cooper and Wood (1982) and Tsay (1989a, 1991), among others. These studies show that canonical correlation analysis is useful in specifying the Kronecker indexes under normality. On the other hand, the assumption of Gaussian innovations is questionable in many applications, especially in analysis of economic and financial data that often exhibit conditional heteroscedasticity. See, for instance, the summary statistics of asset returns in Tsay (2002, Chap.1). In the literature, a simple approach to model conditional heteroscedasticity is to apply the generalized autoregressive conditional heteroscedastic (GARCH) model of Engle (1982) and Bollerslev (1986). We adopt such a model for the innovation series of multivariate time series data.

In this paper, we continue to employ canonical analysis in multivariate time series. However, we focus on statistical inference concerning canonical correlation coefficients when the distribution of the innovations is not Gaussian. Our main objective is to identify a vector model with structural specification for a given time series that exhibits conditional heteroscedasticity and has high kurtosis. Specifically, we study canonical correlation analysis when the innovations of the vector time series follow a multivariate GARCH model. In other words, we consider the problem of specifying Kronecker indexes of a vector linear time series when the innovations follow a GARCH model. We provide theoretical justifications for using the analysis and propose a modified test statistic for testing zero canonical correlation. We use simulations to investigate the performance of the proposed analysis in the finite sample case and apply the analysis to a financial vector time series.

1.1. Preliminaries

Based on the Wold decomposition, a k-dimensional stationary time series $\boldsymbol{Z}_t = (z_{1t}, z_{2t}, \ldots, z_{kt})'$ can be written as

$$\boldsymbol{Z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{a}_{t-i}, \tag{1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ is a constant vector, $\boldsymbol{\psi}_i$ are $k \times k$ coefficient matrices with $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$ being the identity matrix, and $\{\boldsymbol{a}_t = (a_{1t}, \ldots, a_{kt})'\}$ is a sequence of k-dimensional uncorrelated random vectors with mean zero and positive-definite covariance matrix $\boldsymbol{\Sigma}$. That is, $E(\boldsymbol{a}_t) = \boldsymbol{0}$, $E(\boldsymbol{a}_t \boldsymbol{a}_{t-i}') = \boldsymbol{0}$ if $i \neq 0$ and $E(\boldsymbol{a}_t \boldsymbol{a}_t') = \boldsymbol{\Sigma}$. The $\boldsymbol{a}_t$ process is referred to as the innovation series of $\boldsymbol{Z}_t$. If $\sum_{i=0}^{\infty} \|\boldsymbol{\psi}_i\| < \infty$, then $\boldsymbol{Z}_t$ is (asymptotically) weakly stationary, where $\|\boldsymbol{A}\|$ is a matrix norm, e.g., $\|\boldsymbol{A}\| = \sqrt{\text{trace}(\boldsymbol{A}\boldsymbol{A}')}$. Often one further assumes that $\boldsymbol{a}_t$ is Gaussian. In this paper, we assume that

$$\sup_{i,t} E(|a_{it}|^{\eta} | F_{t-1}) < \infty \quad \text{almost surely for some } \eta > 2, \tag{2}$$

where $F_{t-1} = \sigma\{\boldsymbol{a}_{t-1}, \boldsymbol{a}_{t-2}, \ldots\}$ denotes information available at time $t-1$. Writing $\boldsymbol{\psi}(B) = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i B^i$, where B is the backshift operator such that $B\boldsymbol{Z}_t = \boldsymbol{Z}_{t-1}$, then $\boldsymbol{Z}_t = \boldsymbol{\mu} + \boldsymbol{\psi}(B)\boldsymbol{a}_t$. If $\boldsymbol{\psi}(B)$ is rational, then $\boldsymbol{Z}_t$ has a VARMA representation

$$\boldsymbol{\Phi}(B)(\boldsymbol{Z}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(B)\boldsymbol{a}_t, \tag{3}$$

where $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \sum_{i=1}^{p} \boldsymbol{\Phi}_i B^i$ and $\boldsymbol{\Theta}(B) = \boldsymbol{I} - \sum_{j=1}^{q} \boldsymbol{\Theta}_j B^j$ are two matrix polynomials of order p and q, respectively, and have no common left factors. To ensure identifiability, further conditions are required such as $\text{rank}[\boldsymbol{\Phi}_p : \boldsymbol{\Theta}_q] = \dim(\boldsymbol{Z}_t)$. See Dunsmuir and Hannan (1976) for more details. The stationarity condition of $\boldsymbol{Z}_t$ is equivalent to that all zeros of the polynomial $|\boldsymbol{\Phi}(B)|$ are outside the unit circle. The matrix polynomials of the model are related by $\boldsymbol{\psi}(B) = [\boldsymbol{\Phi}(B)]^{-1}\boldsymbol{\Theta}(B)$.

The number of parameters of the VARMA model in (3) could reach $(p+q)k^2 + k + k(k+1)/2$ if no constraint is applied. This could make parameter estimation unnecessarily difficult. Some procedures were proposed to reduce the number of parameters in fitting VARMA(p,q) models, e.g., Koreisha and Pukkila (1987). On the other hand, several methods are available in the literature that can simplify the use of VARMA models when the innovations $\{\boldsymbol{a}_t\}$ are Gaussian. For instance, specification of Kronecker indexes of a Gaussian vector time series can lead to a parsimonious parametrization of VARMA representation, see Tsay (1989b).

In many situations, the innovational process $\boldsymbol{a}_t$ has conditional heteroscedasticity. In the univariate case, Bollerslev(1986) proposed a GARCH(r_1, r_2) model to handle conditional heteroscedasticity. The model can be written as

$$a_t = \sqrt{g_t}\epsilon_t, \qquad g_t = \alpha_0 + \sum_{i=1}^{r_1} \alpha_i a_{t-i}^2 + \sum_{j=1}^{r_2} \beta_j g_{t-j}, \tag{4}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance 1. It is well known that a_t is asymptotically second order stationary if $\sum_{i=1}^{r_1} \alpha_i + \sum_{j=1}^{r_2} \beta_j < 1$. Generalization of the GARCH models to the multivariate case introduces additional complexity to the modelling procedure because the covariance matrix of $\boldsymbol{a}_t$ has $k(k+1)/2$ elements. Writing the conditional variance-covariance matrix of $\boldsymbol{a}_t$ given the past information as $\boldsymbol{\Sigma}_t = E(\boldsymbol{a}_t\boldsymbol{a}_t'|F_{t-1})$, where F_{t-1} is defined in (2), we have $\boldsymbol{a}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\epsilon}_t$, where $\boldsymbol{\Sigma}_t^{1/2}$ is the symmetric square-root of the matrix $\boldsymbol{\Sigma}_t$ and $\{\boldsymbol{\epsilon}_t\}$ is a sequence of independent and identically distributed random vectors with mean zero and identity covariance matrix. Often $\boldsymbol{\epsilon}_t$ is assumed to follow a multivariate normal or Student-t distribution. To ensure the positive definiteness of $\boldsymbol{\Sigma}_t$, several models have been proposed in the literature. For example, consider the simple case of order (1,1). Engle and Kroner (1995) consider the BEKK model

$$\boldsymbol{\Sigma}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{A}\boldsymbol{a}_{t-1}\boldsymbol{a}_{t-1}'\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{\Sigma}_{t-1}\boldsymbol{B}',$$

where $\boldsymbol{C}$ is a lower triangular matrix and $\boldsymbol{A}$ and $\boldsymbol{B}$ are $k \times k$ matrices. Ding (1994) and Bollerslev, Engle and Nelson (1994) discuss the diagonal model

$$\boldsymbol{\Sigma}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{A}\boldsymbol{A}' \otimes (\boldsymbol{a}_{t-1}\boldsymbol{a}_{t-1}') + \boldsymbol{B}\boldsymbol{B}' \otimes \boldsymbol{\Sigma}_{t-1},$$

where $\otimes$ stands for matrix Hadamard product (element-wise product).

In the presence of GARCH effects, the time series $\boldsymbol{Z}_t$ is no longer Gaussian. Its innovations become a sequence of uncorrelated, but serially dependent random vectors. It is well-known that such innovations tend to have heavy tails, see Engle (1982) and Tsay (2002), among others. The performance of canonical correlation analysis under such innovations has yet to be investigated. This is the main objective of our paper.

The paper is organized as follows. Section 2 reviews Kronecker indexes and VARMA model specification when the Kronecker indexes are known. Section 3 reviews the specification of Kronecker indexes for Gaussian linear time series. Section 4 establishes the statistics to specify Kronecker indexes for VARMA +GARCH process by a central limit theorem for the sample autocovariance of a non-Gaussian process. Section 5 presents some simulation results, and Section 6 applies the analysis to a financial time series.

2. The Kronecker Index and Vector ARMA Representation

2.1. Vector ARMA model implied by Kronecker index

For simplicity, we assume that $E(\boldsymbol{Z}_t) = \boldsymbol{\mu} = \boldsymbol{0}$. Given a time point t, define the past and future vectors $\boldsymbol{P}_t$ and $\boldsymbol{F}_t$ of the process $\boldsymbol{Z}_t$ as $\boldsymbol{P}_t = (\boldsymbol{Z}'_{t-1}, \boldsymbol{Z}'_{t-2}, \ldots)'$ and $\boldsymbol{F}_t = (\boldsymbol{Z}'_t, \boldsymbol{Z}'_{t+1}, \ldots)'$. The Hankel Matrix of $\boldsymbol{Z}_t$ is defined as $\boldsymbol{H} = E(\boldsymbol{F}_t \boldsymbol{P}'_t)$. It is obvious that for a VARMA model in (3), the Hankel matrix $\boldsymbol{H}$ is of finite rank. In fact, it can be shown that Rank($\boldsymbol{H}$) is finite if and only if $\boldsymbol{Z}_t$ has a VARMA model representation, see Hannan and Deistler (1988) and Tsay (1991).

The Kronecker indexes of $\boldsymbol{Z}_t$ consist of a set of non-negative integers $\{K_i \mid i = 1, \ldots, k\}$ such that for each i, K_i is the smallest non-negative integer that the $(k \times K_i + i)$th row of $\boldsymbol{H}$ is either a null vector or is a linear combination of the previous rows of $\boldsymbol{H}$. It turns out that $\sum_{i=1}^{k} K_i$ is the rank of $\boldsymbol{H}$, which is invariant under different VARMA presentations of $\boldsymbol{Z}_t$. In fact, the set of Kronecker indexes, $\{K_i\}_{i=1}^{k}$, of a given VARMA process is invariant under various forms of model representation. Tsay(1991) illustrates how to construct an Echelon VARMA form for $\boldsymbol{Z}_t$ using the Kronecker indexes $\{K_i\}_{i=1}^{k}$. Specifically, for a stationary process $\boldsymbol{Z}_t$ with specified Kronecker index $\{K_1, \ldots, K_k\}$, let $p = \max\{K_i \mid i = 1, \ldots, k\}$. Then $\boldsymbol{Z}_t$ follows a VARMA(p, p) model

$$\boldsymbol{\Phi}_0 \boldsymbol{Z}_t - \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{Z}_{t-i} = \boldsymbol{\delta} + \boldsymbol{\Phi}_0 \boldsymbol{e}_t - \sum_{j=1}^{p} \boldsymbol{\Theta}_j \boldsymbol{e}_{t-j}, \tag{5}$$

where $\boldsymbol{\delta}$ is a constant vector, the ith row of $\boldsymbol{\Phi}_j$ and $\boldsymbol{\Theta}_j$ are zero for $j > K_i$, and $\boldsymbol{\Phi}_0$ is a lower triangular matrix with ones on the diagonal. Furthermore, some elements of $\boldsymbol{\Phi}_i$ can be set to zero based on the Kronecker indexes. A VARMA model in (5) provides a unique ARMA representation for $\boldsymbol{Z}_t$, see Hannan and Deistler (1988, Theorem 2.5.1).

2.2. Specification of the Kronecker index

Under normality, canonical correlation analysis has been applied to identify the Kronecker index by Akaike (1976), Cooper and Wood (1982) and Tsay (1989a), among others. A canonical correlation $\rho \geq 0$ between two random vectors $\boldsymbol{P}$ and $\boldsymbol{F}$ can be obtained from:

$$\begin{aligned} \boldsymbol{\Sigma}_{pp}^{-1}\boldsymbol{\Sigma}_{pf}\boldsymbol{\Sigma}_{ff}^{-1}\boldsymbol{\Sigma}_{fp}\boldsymbol{V}_p = \rho^2\boldsymbol{V}_p, \\ \boldsymbol{\Sigma}_{ff}^{-1}\boldsymbol{\Sigma}_{fp}\boldsymbol{\Sigma}_{pp}^{-1}\boldsymbol{\Sigma}_{pf}\boldsymbol{V}_f = \rho^2\boldsymbol{V}_f, \end{aligned} \tag{6}$$

where $\boldsymbol{\Sigma}_{fp} = \mathrm{Cov}(\boldsymbol{F}, \boldsymbol{P})$, e.g., Reinsel (1997). The variables $X = \boldsymbol{V'}_f\boldsymbol{F}$ and $Y = \boldsymbol{V}'_p\boldsymbol{P}$ are the corresponding canonical variates. The canonical correlation is the cross correlation between X and Y, ie., $\rho = \mathrm{corr}(X, Y)$. Sample canonical correlations are constructed from sample covariance matrices of $\boldsymbol{P}$ and $\boldsymbol{F}$.

If the smallest canonical correlation between the future and past vectors $\boldsymbol{F}_t$ and $\boldsymbol{P}_t$ is zero, then $X_t = \boldsymbol{V'}_f\boldsymbol{F}_t$ is uncorrelated with $\boldsymbol{P}_t$, i.e., $\mathrm{Cov}(X_t, \boldsymbol{P}_t) = \boldsymbol{V'}_f E(\boldsymbol{F}_t\boldsymbol{P}'_t) = \boldsymbol{V'}_f\boldsymbol{H} = \boldsymbol{0}$. This leads to a row dependency of the Hankel matrix so that the analysis is directly related to the Kronecker index. Testing for zero canonical correlation thus plays an important role in specifying Kronecker indexes. Cooper and Wood (1982) used the traditional χ^2 test to propose a modelling procedure:

Step 1: Select a sufficiently large lag s so that the vector $\boldsymbol{P}_t = (\boldsymbol{Z}'_{t-1}, \ldots, \boldsymbol{Z}'_{t-s})'$ is a good approximation of the past vector, and choose an initial future sub-vector $\boldsymbol{F}^*_t = \{Z_{1t}\}$. If a vector AR approximation is used, then s can be selected by an information criterion such as AIC or BIC.

Step 2: Let $\hat{\rho}$ be the smallest sample canonical correlation in modulus between $\boldsymbol{F}^*_t$ and $\boldsymbol{P}_t$. Denote the canonical variates by $X_t = \boldsymbol{V'}_f\boldsymbol{F}^*_t$ and $Y_t = \boldsymbol{V}'_p\boldsymbol{P}_t$, and compute the test statistics

$$S = -n\log(1 - \hat{\rho}^2) \sim \chi^2_{ks-f+1}, \tag{7}$$

where n is the number of observations, f and ks are the dimension of $\boldsymbol{F}^*_t$ and $\boldsymbol{P}_t$, respectively.

Step 3: Denote the last element of $\boldsymbol{F}^*_t$ as $Z_{i,t+h}$. If $H_0 : \rho = 0$ is not rejected, then the Kronecker index for the ith component Z_{it} of $\boldsymbol{Z}_t$ is $K_i = h$. In this case, update the future vector $\boldsymbol{F}_t$ by removing $Z_{i,t+j}$ for $j \geq h$. If all k Kronecker indexes have been found, the procedure is terminated. Otherwise, augment $\boldsymbol{F}^*_t$ by adding the next available element of the updated $\boldsymbol{F}_t$ and return to Step 2.

The asymptotic χ^2 distribution of the S-statistic in (7) of Step 2 is derived under the independence sampling assumption. Tsay (1989a) argued that the canonical correlations cannot be treated as the cross correlation of two white-noise series since the corresponding canonical variates are serially correlated. Suppose $\boldsymbol{F}_t^* = (Z_{1,t}, \ldots, Z_{i,t+h})'$. The smallest sample canonical correlation $\hat{\rho}$ is the lag-$(h+1)$ sample cross-correlation $\hat{\rho}_{xy}(h+1)$ of the corresponding canonical variates $X_t = \boldsymbol{V}_f'\boldsymbol{F}_t^*$ and $Y_t = \boldsymbol{V}_p'\boldsymbol{P}_t$, because Y_t is observable at time $t-1$ whereas X_t is observable at time $t+h$. Under $H_0 : \rho_{xy}(m) = 0$, the asymptotic variance of $\hat{\rho}_{xy}(m)$ is (Box and Jenkins (1976, p.736))

$$\mathrm{var}[\hat{\rho}_{xy}(m)] \approx n^{-1} \sum_{\nu=-\infty}^{\infty} \{\rho_{xx}(\nu)\rho_{yy}(\nu) + \rho_{xy}(m+\nu)\rho_{yx}(m-\nu)\}. \tag{8}$$

For canonical correlation analysis, under $H_0 : \rho = 0$, we have $\mathrm{cov}(X_t, \boldsymbol{C}'\boldsymbol{P}_t) = \boldsymbol{0}$ $\forall \boldsymbol{C} \in R^g$, where g is the dimension of $\boldsymbol{P}_t$. In particular, $\mathrm{cov}(X_t, \boldsymbol{Z}_{t-i}) = \boldsymbol{0}$ $\forall i > 0$. Therefore, $\mathrm{cov}(X_t, X_{t-j}) = 0$ for $j \geq h+1$ because X_{t-j} is in the sigma-field $\sigma\{\boldsymbol{Z}_{t-1}, \boldsymbol{Z}_{t-2}, \ldots\}$. Consequently, $\rho_{xx}(j) = 0$ for $j \geq h+1$ and X_t is in $\sigma\{a_t, \ldots, a_{t+h}\}$ and is an MA(h) process. Using this fact and (8), the asymptotic variance of the sample canonical correlation $\hat{\rho}$ is $\mathrm{var}(\hat{\rho}) \approx n^{-1}\{1 + 2\sum_{\nu=1}^{h} \rho_{xx}(\nu)\rho_{yy}(\nu)\}$. Tsay (1989a) proposed a modified test statistic

$$T = -(n-s)\log(1 - \frac{\hat{\rho}^2}{\hat{d}}) \sim \chi^2_{ks-f+1}, \tag{9}$$

where $\hat{d} = 1 + 2\sum_{\nu=1}^{h} \hat{\rho}_{xx}(\nu)\hat{\rho}_{yy}(\nu)$. In (9), it is understood that $\hat{d} = 1$ if $h = 0$, $\hat{\rho}_{xx}(\nu)$ and $\hat{\rho}_{yy}(\nu)$ are the lag-ν sample autocorrelations of X_t and Y_t, respectively, and n is the sample size.

Akaike (1976) adopted a criterion function to judge the significance of the smallest canonical correlation in Step 2 of the above procedure: $DIC(f) = -n\log(1 - \hat{\rho}_i{}^2) - 2(ks - f + 1)$, where f is the dimension of $\boldsymbol{F}_t^*$. The canonical correlation ρ is judged to be zero if $DIC(f) < 0$.

The Bartlett's formula in (8) is for independent Gaussian innovations $\{a_t\}$. This is not the case when the innovations follow a GARCH(r_1, r_2) model. For a weakly dependent process, traditional asymptotic variance estimates may be misleading (Romano and Thombs (1996) and Berlinet and Franco (1997)). In the next section, we study the property of sample auto-covariances in the presence of GARCH effects in the innovations.

3. Sample Auto-covariance Functions of a Linear Process

We begin our study with a useful result.

Lemma 1. *Suppose $\{a_t\}$ is a stationary GARCH(r_1, r_2) process of* (4), *with finite fourth moment and ϵ_t symmetrically distributed, then $E(a_i a_k a_j a_l) = 0$ $\forall i \le j \le k \le l$ unless $i = j$ and $k = l$ both hold.*

Proof. $E(a_i a_j a_k a_l) = 0$ holds if $k < l$ or $i < j = k = l$ since $E(a_l|F_{l-1}) = 0, E(a_l^3|F_{l-1}) = 0$. For $i < j < k = l$, observe $E(a_i a_j a_k^2) = E(a_i a_j h_k) = \sum_{m=1}^{(k-j)\wedge r_1} \alpha_m E(a_i a_j h_{k-m}) + \sum_{m=1}^{r_2} \beta_m E(a_i a_j h_{k-m})$. Then $E(a_i a_j a_k^2) = 0$ follows from induction by noticing that $E(a_i a_j h_m) = 0$ $\forall m \le j$.

Proposition 1. *Suppose $\{a_t\}$ is a GARCH(r_1, r_2) process with $E(a_t^2) = \sigma^2$ and $E(a_t^4) < \infty$, and the process X_t is defined as $X_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}$ with $\sum_i |\psi_i| < \infty$, and $\sum_i i\psi_i^2 < \infty$. Let $\gamma_{xx}(0) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$. Then $\sum_{t=1}^{\infty} \|E(X_t^2 - \gamma_{xx}(0)|\boldsymbol{\xi}_0)\| < \infty$, where $\boldsymbol{\xi}_0 = \sigma\{\epsilon_0, \epsilon_{-1}, \ldots\}$ and $\|Y\|$ denotes the L^2-norm of a random variable Y.*

Proof. Let $C > 0$ be some generic constant throughout this proof. Note that

$$\|E(X_t^2 - \gamma_{xx}(0)|\boldsymbol{\xi}_0)\| \le \|\sum_{i \ge t} \psi_i^2 (a_{t-i}^2 - \sigma^2)\| + \|\sum_{i \ne j \ge t} \psi_i \psi_j a_{t-i} a_{t-j}\| + \sum_{i=0}^{t-1} \psi_i^2 \|E(a_{t-i}^2 - \sigma^2|\boldsymbol{\xi}_0)\|.$$

The first term of the right side satisfies $\|\sum_{i \ge t} \psi_i^2 (a_{t-i}^2 - \sigma^2)\| \le C \sum_{i \ge t} \psi_i^2$ by the triangle inequality. The second term satisfies $\|\sum_{i \ne j \ge t} \psi_i \psi_j a_{t-i} a_{t-j}\| = \sqrt{2 \sum_{i \ne j \ge t} \psi_i^2 \psi_j^2 E(a_{t-i}^2 a_{t-j}^2)} \le C \sum_{i \ge t} \psi_i^2$. The third term becomes $\sum_{i=0}^{t-1} \psi_i^2 \|E(a_{t-i}^2 - \sigma^2|\boldsymbol{\xi}_0)\| \le C \sum_{i=0}^{t-1} r^{t-i} \psi_i^2$ because $\|E(a_{t-i}^2 - \sigma^2|\boldsymbol{\xi}_0)\| \le C r^{t-i}$ with $r \in (0, 1)$, which can be shown by using the ARMA representation of GARCH(r_1, r_2) process. These inequalities show that $\sum_{t=1}^{\infty} \|E(X_t^2 - \gamma_{xx}(0)|\boldsymbol{\xi}_0)\| \le \sum_{t=1}^{\infty} C(\sum_{i \ge t} \psi_i^2 + \sum_{i=0}^{t-1} r^{t-i} \psi_i^2) = C \sum_{i=0}^{\infty} \psi_i^2 (i + r/(1-r)) < \infty$.

Defining the norm of a random matrix as $\|A\| := \sqrt{E(trAA')}$, we can generalize Proposition 1 to a linear process with innovational process that follows a multivariate GARCH model.

Proposition 2. *Assume $\boldsymbol{a}_t = (a_{1t}, \ldots, a_{mt})'$, and a_{it} follows a univariate GARCH(r_1, r_2) model and is stationary with finite fourth moment for each $i = 1, \ldots, m$. Consider the process $X_t = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i' \boldsymbol{a}_{t-i}$, where $\boldsymbol{\Psi}_i$ are m-dimensional vectors. Assume further that $\sum_{i=0}^{\infty} \|\boldsymbol{\Psi}_i\| < \infty$ and $\sum_{i=0}^{\infty} i \|\boldsymbol{\Psi}_i\|^2 < \infty$. Let $\boldsymbol{\xi}_0 = \sigma\{\boldsymbol{a}_0, \boldsymbol{a}_{-1}, \ldots\}$. Then $\sum_{t=1}^{\infty} \|E(X_t^2 - \gamma_{xx}(0)|\boldsymbol{\xi}_0)\| < \infty$, where $\gamma_{xx}(0) = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i' \boldsymbol{\Sigma} \boldsymbol{\Psi}_i$ and $\boldsymbol{\Sigma} = E(\boldsymbol{a}_t \boldsymbol{a}_t') = diag(\sigma_1^2, \ldots, \sigma_m^2)$.*

Proof. Since $X_t^2 = \sum_{i,j \ge 0} \boldsymbol{\Psi}_i' \boldsymbol{a}_{t-i} \boldsymbol{a}_{t-j}' \boldsymbol{\Psi}_j$. By the triangle inequality,

$$\|E(X_t^2 - \gamma_{xx}(0)|\boldsymbol{\xi}_0)\| \le \|\sum_{i \ge t} \boldsymbol{\Psi}_i' (\boldsymbol{a}_{t-i} \boldsymbol{a}_{t-i}' - \boldsymbol{\Sigma}) \boldsymbol{\Psi}_i\| + \|\sum_{i \ne j \ge t} \boldsymbol{\Psi}_i' \boldsymbol{a}_{t-i} \boldsymbol{a}_{t-j}' \boldsymbol{\Psi}_j\|$$

$$+\sum_{i=0}^{t-1}\|\boldsymbol{\Psi}_i' E(\boldsymbol{a}_{t-i}\boldsymbol{a}_{t-i}' - \boldsymbol{\Sigma}|\boldsymbol{\xi}_0)\boldsymbol{\Psi}_i\|$$
$$\leq C\sum_{i\geq t}\|\boldsymbol{\Psi}_i\|^2 + C\sum_{i\geq t}\|\boldsymbol{\Psi}_i\|^2 + C\sum_{i=0}^{t-1}\|\boldsymbol{\Psi}_i\|^2 r^{t-i}.$$

The result follows.

From $X_t Y_{t+h} = ((X_t+Y_{t+h})^2 - (X_t - Y_{t+h})^2)/4$ and the triangle inequality, one has the following.

Corollary 1. *Suppose* $X_t = \sum_{i=0}^{\infty}\boldsymbol{\Psi}_i'\boldsymbol{a}_{t-i}$ *and* $Y_t = \sum_{i=0}^{\infty}\boldsymbol{\Phi}_i'\boldsymbol{a}_{t-i}$ *both satisfy the conditions in Proposition* 2*. Let* $\gamma_{xy}(h) = E(X_tY_{t+h})$*, where* h *is an integer. Then* $\sum_{t=1}^{\infty}\|E(X_tY_{t+h} - \gamma_{xy}(h)|\boldsymbol{\xi}_0)\| < \infty$.

To generalize the result to the case that $\boldsymbol{X}_t$ is multivariate, we adopt the matrix vectorization operator Vec defined as: Given a matrix $\boldsymbol{A}_{m\times n}$, $\text{Vec}(\boldsymbol{A}) = (\boldsymbol{A}_1', \ldots, \boldsymbol{A}_n')'$, i.e., $\text{Vec}(\boldsymbol{A})$ is a column vector generated by stacking the columns of $\boldsymbol{A}$.

Proposition 3. *Let* $\boldsymbol{X}_t = (X_{1t},\ldots,X_{kt})' = \sum_{i=0}^{\infty}\boldsymbol{\Psi}_i\boldsymbol{a}_{t-i}$*, where* $\boldsymbol{\Psi}_i$ *are matrices of dimension* $k\times m$ *and* $\boldsymbol{a}_t$ *is* m*-dimensional and follows a pure diagonal stationary GARCH*(r_1,r_2) *model with finite fourth moment. Further, suppose that* $\sum_{i=0}^{\infty}\|\boldsymbol{\Psi}_i\| < \infty, \sum_{i=0}^{\infty} i\|\boldsymbol{\Psi}_i\|^2 < \infty$*. Letting* $\boldsymbol{\Sigma} = E(\boldsymbol{X}_t\boldsymbol{X}_{t+h}')$ *where* h *is an integer, we have* $\sum_{t=1}^{\infty}\|\,Vec(E(\boldsymbol{X}_t\boldsymbol{X}_{t+h}'|\boldsymbol{\xi}_0) - \boldsymbol{\Sigma})\| < \infty$.

Proof. $\|\text{Vec}(E(\boldsymbol{X}_t\boldsymbol{X}_{t+h}'|\boldsymbol{\xi}_0) - \boldsymbol{\Sigma})\| \leq \sum_{i,j=1}^{k}\|E(X_{it}X_{j,t+h}|\boldsymbol{\xi}_0) - \Sigma_{i,j}\|$. Since each component of $\boldsymbol{X}_t$ satisfies the conditions in Corollary 1, the result follows.

The following lemma of Wu (2003) is used to prove the next proposition.

Lemma 2. (Wu (2003)) *Let* $\{\xi_n\}_{n\in\mathbb{Z}}$ *be a stationary and ergodic Markov chain and* $h(\cdot)$ *be a measureable function defined on the state space of the chain with zero mean and finite variance. If* $\sum_{n=1}^{\infty}\|E[h(\xi_n)|\xi_1] - E[h(\xi_n)|\xi_0]\| < \infty$*, then for some* $\sigma^2 < \infty$*,* $(1/\sqrt{n})\sum_{i=1}^{n}h(\boldsymbol{\xi}_i) \longrightarrow N(0,\sigma^2)$.

Proposition 4. *Let* $\boldsymbol{X}_t = (X_{1t},\ldots,X_{kt})' = \sum_{i=0}^{\infty}\boldsymbol{\Psi}_i\boldsymbol{a}_{t-i}$ *and* $\boldsymbol{Y}_t = (Y_{1t},\ldots,Y_{lt})' = \sum_{i=0}^{\infty}\boldsymbol{\Phi}_i\boldsymbol{a}_{t-i}$*, where* $\boldsymbol{\Psi}_i$ *and* $\boldsymbol{\Phi}_i$ *are matrices of dimension* $k\times m$ *and* $l\times m$*, respectively. Suppose both* $\boldsymbol{X}_t$ *and* $\boldsymbol{Y}_t$ *satisfy the conditions in Proposition* 3*. If* $\boldsymbol{\Sigma}_{xy}(h) = E(\boldsymbol{X}_t\boldsymbol{Y}_{t+h}')$*, then*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Vec(\boldsymbol{X}_t\boldsymbol{Y}_{t+h}' - \boldsymbol{\Sigma}_{xy}(h)) \longrightarrow N(\boldsymbol{0},\boldsymbol{\Sigma}),$$

where h *is any integer and* $\boldsymbol{\Sigma} \in R^{kl\times kl}$.

Proof. It suffices to prove the result for the case that $h = 0$ and $\boldsymbol{Y}_t = \boldsymbol{X}_t$, because we can consider $(\boldsymbol{X}'_t, \boldsymbol{Y}'_{t+h})'$ as a new $\boldsymbol{X}_t$ and apply the result we show. We only need to show that $\forall\lambda \in R^{k^2}$, $1/\sqrt{n}(\sum_{t=1}^{n} \lambda'\text{Vec}(\boldsymbol{X}_t\boldsymbol{X}'_t - \boldsymbol{\Sigma}_{xx}(0))) \longrightarrow N(\boldsymbol{0}, \lambda'\boldsymbol{\Sigma}\lambda)$ and then apply Crámer's device. By Proposition 3, $\sum_{t=1}^{\infty} \|E\lambda'$ $\text{Vec}(\boldsymbol{X}_t\boldsymbol{X}'_t - \boldsymbol{\Sigma}_{xx}(0)|\boldsymbol{\xi}_0)\| \le \sum_{t=1}^{\infty} \|\lambda\|.\ \|\text{Vec}(E(\boldsymbol{X}_t\boldsymbol{X}'_t|\boldsymbol{\xi}_0) - \boldsymbol{\Sigma}_{xx}(0))\| < \infty$. The result follows from Lemma 2. To apply the Lemma, let $\boldsymbol{\xi}_k = (\boldsymbol{\epsilon}_k, \boldsymbol{\epsilon}_{k-1}, \ldots)$ and notice that we have proved $\sum_{n=1}^{\infty} \|E(h(\boldsymbol{\xi}_n)|\boldsymbol{\xi}_0)\| < \infty$ and the result $\sum_{n=1}^{\infty} \|E(h(\boldsymbol{\xi}_n)|\boldsymbol{\xi}_1)\| < \infty$ holds by a similar argument.

Remark 1. For a causal, stationary VARMA(p, q) process $\boldsymbol{\Phi}(B)(\boldsymbol{Z}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(B)\boldsymbol{a}_t$, its $MA(\infty)$ representation $\boldsymbol{Z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i\boldsymbol{a}_{t-i}$ satisfies the condition $\sum_{i=0}^{\infty} \|\boldsymbol{\Psi}_i\| < \infty, \sum_{i=0}^{\infty} i\|\boldsymbol{\Psi}_i\|^2 < \infty$, since $\|\boldsymbol{\Psi}_i\| \sim r^i$ with $r \in (0, 1)$ being the largest root (in magnitude) of $\boldsymbol{\Phi}(B^{-1})$. Consequently, if $\boldsymbol{a}_t$ follows a pure diagonal GARCH model with finite fourth moment, the sample auto-covariance matrix of $\boldsymbol{Z}_t$ has an asymptotic joint normal distribution.

We now establish a result similar to Theorem 1 of Tsay (1989a) for VARMA models with GARCH innovations.

Theorem 1. *Suppose that $\boldsymbol{Z}_t$ is a k-dimensional stationary VARMA process in (3), where the innovation series $\boldsymbol{a}_t$ follows a GARCH(r_1, r_2) model with finite fourth moment. Let $\boldsymbol{P}_t = (\boldsymbol{Z}'_{t-1}, \ldots, \boldsymbol{Z}'_{t-s})'$ be a past vector with a prespecified $s > 0$ that contains all the information needed in predicting the future observation of $\boldsymbol{Z}_t$, $\boldsymbol{F}_t = (z_{1,t}, \ldots, z_{i,t+h})'$ be the future subvector of $\boldsymbol{Z}_t$ constructed according to the procedure described in Section 2. Let $\hat{\rho}$ be the smallest sample canonical correlation between $\boldsymbol{P}_t$ and $\boldsymbol{F}_t$. Under the null hypothesis that the smallest canonical correlation ρ between $\boldsymbol{P}_t$ and $\boldsymbol{F}_t$ is zero, but all the other canonical correlations are nonzero, then $\hat{\rho}^2/var(\hat{\rho})$ has an asymptotic χ^2 distribution with $ks - f + 1$ degrees of freedom, where f is the dimension of $\boldsymbol{F}_t$.*

Proof. First, assume the future canonical variate $X_t = \boldsymbol{V}'_f\boldsymbol{F}_t$ is known. Under the null hypothesis that the smallest canonical correlation between $\boldsymbol{F}_t$ and $\boldsymbol{P}_t$ is zero, X_t is uncorrelated with $\boldsymbol{P}_t$. Let $\boldsymbol{P}$ and $\boldsymbol{X}$ be the matrices of $\boldsymbol{P}_t$ and X_t. The sample covariance matrix $\hat{\boldsymbol{\beta}} = \boldsymbol{P}'\boldsymbol{X} \sim N_{ks}(\boldsymbol{0}, \boldsymbol{V})$ by Proposition 4. Thus $\hat{\boldsymbol{\beta}}'\boldsymbol{V}^{-1}\hat{\boldsymbol{\beta}} \sim \chi^2_{ks}$. Observe

$$\hat{\boldsymbol{\beta}}'\boldsymbol{V}^{-1}\hat{\boldsymbol{\beta}} = \sup_{\boldsymbol{\alpha}\in R^{ks}, \neq 0} \frac{(\boldsymbol{\alpha}'\hat{\boldsymbol{\beta}})^2}{\boldsymbol{\alpha}'\boldsymbol{V}\boldsymbol{\alpha}} = \sup_{\boldsymbol{\alpha}\neq 0} \frac{(\boldsymbol{\alpha}'\hat{\boldsymbol{\beta}})^2/(\boldsymbol{\alpha}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{\alpha}\boldsymbol{X}'\boldsymbol{X})}{\boldsymbol{\alpha}'\boldsymbol{V}\boldsymbol{\alpha}/(\boldsymbol{\alpha}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{\alpha}\boldsymbol{X}'\boldsymbol{X})}.$$

Let $\rho = \text{corr}(X_t, \boldsymbol{\alpha}'\boldsymbol{P}_t) = (\boldsymbol{X}'\boldsymbol{P}\boldsymbol{\alpha})/(\sqrt{\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{P}\boldsymbol{\alpha})'(\boldsymbol{P}\boldsymbol{\alpha})})$ be the sample correlation between X_t and a linear combination of the past vectors $\boldsymbol{\alpha}'\boldsymbol{P}_t$. By a Taylor expansion we have

$$\text{Var}(\rho) \approx \frac{\text{Var}(\boldsymbol{X}'\boldsymbol{P}\boldsymbol{\alpha})}{\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{P}\boldsymbol{\alpha})'\boldsymbol{P}\boldsymbol{\alpha}} \approx \frac{\boldsymbol{\alpha}'\boldsymbol{V}\boldsymbol{\alpha}}{\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{P}\boldsymbol{\alpha})'\boldsymbol{P}\boldsymbol{\alpha}}$$

under the null hypothesis that $\mathrm{Cov}(X_t, \boldsymbol{\alpha}'\boldsymbol{P}_t) = 0$. So $\hat{\boldsymbol{\beta}}'\boldsymbol{V}^{-1}\hat{\boldsymbol{\beta}} = \sup_{\boldsymbol{\alpha}\neq 0} \rho^2/\mathrm{Var}(\rho)$ is the maximum normalized correlation between X_t and any linear combination of the past vector $\boldsymbol{P}_t$, which is 0 under the null hypothesis. Notice the canonical variate $\boldsymbol{V}_p'\boldsymbol{P}_t$ of ρ also gives the maximum correlation between X_t and any linear combination of $\boldsymbol{P}_t$, which is 0. This implies $\hat{\boldsymbol{\beta}}'\boldsymbol{V}^{-1}\hat{\boldsymbol{\beta}} \approx \hat{\rho}^2/\mathrm{Var}(\hat{\rho}) \sim \chi^2_{ks}$. Since $\boldsymbol{V}_f$ has to be estimated from the data, the degrees of freedom become $ks - f + 1$.

4. Asymptotic Variance of Sample Cross Correlation

Next we consider the variance of the sample cross-correlation coefficient for the case that gives rise to a zero canonical correlation between the past and future vectors of $\boldsymbol{Z}_t$. To this end, we make use of the following result. When the variation of variables U, V and W about their mean u, v, and w, respectively, becomes small, we have

$$\mathrm{Var}\left(\frac{W}{\sqrt{UV}}\right) \approx \frac{\mathrm{Var}(W)}{uv} + \frac{1}{4}\frac{w^2}{u^2 v}\mathrm{Var}(U) + \frac{1}{4}\frac{w^2}{uv^3}\mathrm{Var}(V) - \frac{w}{u^2 v}\mathrm{Cov}(W,U) - \frac{w}{uv^2}\mathrm{Cov}(W,V) + \frac{1}{2}\frac{w^2}{u^2v^2}\mathrm{Cov}(U,V).$$

Suppose Y_t and X_t are stationary moving-average processes. More specifically, $Y_t = \sum_{i=0}^{h}\phi_i a_{t-i}$ and $X_t = \sum_{i=0}^{\infty}\psi_i a_{t-i}$, with a_t being a GARCH(r_1, r_2) process (4). By Lemma 1, $E(a_i a_j a_k a_l) = 0$, $\forall i \leq j \leq k \leq l$, unless $i = j$ and $k = l$ both hold. Let $U = \hat{\gamma}_{xx}(0)$, $V = \hat{\gamma}_{yy}(0)$ and $W = \hat{\gamma}_{xy}(q) = 1/(n-q)\sum_{t=1}^{n-q} X_t Y_{t+q}$. Given $q > h$, where h corresponds to a Kronecker index, we have $\gamma_{xy}(q) = \gamma_{yy}(q) = 0$ and, by the above formula, the following holds.

$$\begin{aligned}\mathrm{Var}(\hat{\rho}_{xy}(q)) &\approx \frac{1}{n\gamma_{xx}(0)\gamma_{yy}(0)}\sum_{|d|\leq h}\mathrm{Cov}(X_0Y_q, X_dY_{d+q}) \\ &\approx \frac{1}{n\gamma_{xx}(0)\gamma_{yy}(0)}\sum_{|d|\leq h}[\gamma_{xx}(d)\gamma_{yy}(d) + \gamma_{xy}(d+q)\gamma_{xy}(q-d) \\ &\quad +\mathrm{Cum}(X_0, X_d, Y_q, Y_{d+q})] \\ &\approx \frac{1}{n}\sum_{|d|\leq h}\left[\rho_{xx}(d)\rho_{yy}(d) + \frac{\mathrm{Cum}(X_0,X_d,Y_q,Y_{q+d})}{\gamma_{xx}(0)\gamma_{yy}(0)}\right], \end{aligned} \tag{10}$$

where

$$\mathrm{Cum}(X_0, X_d, Y_q, Y_{q+d}) = \sum_{i,j,k,l}\psi_i\psi_j\phi_k\phi_l E a_{t-i}a_{t+d-j}a_{t+q-k}a_{t+q+d-l} - \Big(\sum_{i\geq 0}\psi_i\psi_{i+d}\sigma^2\Big)\Big(\sum_{k=0}^{h-d}\phi_k\phi_{k+d}\sigma^2\Big)$$

$$= \sum_{i\geq 0}\sum_{k=0}^{h-d} \psi_i \psi_{i+d}\phi_k\phi_{k+d}\text{Cov}(a_0^2, a_{q-k+i}^2).$$

Therefore, the fourth order cumulants of $\{X_t\}$ depend on the auto-covariance function of $\{a_t^2\}$. Compared to $\gamma_{xx}(d)\gamma_{yy}(d)$, $\text{Cum}(X_0, X_d, Y_q, Y_{q+d})$ has a non-negligible impact on $\text{Var}(\hat{\rho}_{xy}(p))$ if $\text{Cov}(a_0^2, a_{q-k+i}^2)/E^2(a_0^2)$ is large. For instance, if a_t is a GARCH(1,1) process, then $\text{Cov}(a_0^2, a_1^2)/\sigma^4 = 2\alpha_1 + (6\alpha_1^2(\alpha_1+\beta_1/3))/(1-2\alpha_1^2-(\alpha_1+\beta_1)^2)$. This ratio is 86 given $\alpha_1 = 0.5$ and $\beta_1 = 0.2$. Considering the fourth order cumulant correction term in $\text{Var}(\hat{\rho})$, one can modify the T statistic proposed by Tsay as

$$\begin{aligned} T^* &= -(n-s)\log(1-\frac{\hat{\rho}^2}{\hat{d}^*}) \sim \chi^2_{ks-f+1}, \\ \hat{d}^* &= \sum_{|d|\leq h}\left[\rho_{xx}(d)\rho_{yy}(d) + \frac{\text{Cum}(X_0, X_d, Y_q, Y_{q+d})}{\gamma_{xx}(0)\gamma_{yy}(0)}\right]. \end{aligned} \tag{11}$$

Note that if $\alpha_1 = 0$, the kurtosis of a_t and autocovariance of a_t^2 vanish. Therefore Bartlett's formula continues to hold. Consequently, the test statistics proposed by Tsay should work well. This is confirmed by our simulation study although we did not include the result in the simulation report below.

5. Simulations Study

We conducted some simulations to study the finite sample performance of the modified test statistics. We focused on a two-dimensional ARMA(1,1)+GARCH (1,1) model chosen to have GARCH parameters similar to those commonly seen in empirical asset returns. The model was

$$\boldsymbol{Z}_t - \begin{bmatrix} 0.8 & 0 \\ 0 & 0.3 \end{bmatrix} \boldsymbol{Z}_{t-1} = \boldsymbol{a}_t - \begin{bmatrix} -0.8 & 1.3 \\ -0.3 & 0.8 \end{bmatrix} \boldsymbol{a}_{t-1} \quad t = 1, \ldots, n, \tag{12}$$

where n indicates sample size and

$$\boldsymbol{a}_t = \begin{bmatrix} \sqrt{g_{1t}} & 0 \\ 0 & \sqrt{g_{2t}} \end{bmatrix} \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim i.i.d. \quad N_2(0, I),$$

where the conditional variances satisfy the GARCH(1,1) model

$$\boldsymbol{g}_t = \begin{bmatrix} g_{1t} \\ g_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + 0.2 \begin{bmatrix} a_{1,t-1}^2 \\ a_{2,t-1}^2 \end{bmatrix} + 0.7 \begin{bmatrix} g_{1t-1} \\ g_{2t-1} \end{bmatrix}.$$

For a given sample size n, each realization was obtained by generating $5n$ observations. To reduce the effect of the starting values $\boldsymbol{Z}_0$ and $\boldsymbol{a}_0$, we only used the last n observations. For this model, the two future subvectors, which in theory give a zero canonical correlation, are $\boldsymbol{F}_t(1) = (z_{1t}, z_{2t}, z_{1,t+1})'$ and $\boldsymbol{F}_t(2) = (z_{1t}, z_{2t}, z_{2,t+1})'$. A value of $s = 5$ was selected according to the AIC criterion in a preliminary analysis using pure vector AR models. The corresponding past vector was $\boldsymbol{P}_t = (\boldsymbol{Z}'_{t-1}, \ldots, \boldsymbol{Z}'_{t-5})'$.

Let $S(1)$ and $S(2)$ be the test statistics $S = -n\log(1-\hat{\rho}^2)$ of Cooper and Wood (1982) when the future subvectors are $\boldsymbol{F}_t(1)$ and $\boldsymbol{F}_t(2)$, respectively. Similarly, let $T(1)$ and $T(2)$ be the corresponding test statistics $T = -(n-s)\log(1-\hat{\rho}^2/\hat{d})$ of Tsay (1989a) and $T^*(1)$ and $T^*(2)$ be the test statistics $T^* = -(n-s)\log(1-\hat{\rho}^2/\hat{d}^*)$ proposed in (11). In particular, we adopted the approach of Berlint and Francq (1997) to estimate the variance of sample cross-covariance $\mathrm{Var}[\hat{\gamma}_{xy}(q)]$ by

$$n\ \mathrm{Var}[\hat{\gamma}_{xy}(q)] \approx \hat{\sigma}^*(0) + 2\sum_{i=1}^{n-q}(1-i/n)K(ib_n)\hat{\sigma}^*(i), \tag{13}$$

where $\hat{\sigma}^*(i) = \sum_t X_t Y_{t+q} X_{t+i} Y_{t+i+q}/n - \hat{\gamma}^2_{xy}(q)$, $K(x) = I_{|x|\le 1}$, and $b_n = n^{-1/4}$. However, to improve the robustness of the variance estimate in finite samples, we employed a modified estimate of $\hat{\sigma}^*(i)$. The modification was to use a trimmed sequence $\{X_t Y_{t+q}\}$ by trimming both the lower and upper 0.2 percentiles of $X_t Y_{t+q}$.

As an alternative, we also applied the stationary bootstrap method of Romano and Thombs (1996) to estimate $\mathrm{Var}(\hat{\rho})$. Each bootstrap step was repeated $1,000$ times. Let $B(1)$ and $B(2)$ be the corresponding test statistics $-(n-s)\log(1-\hat{\rho}^2/\hat{d})$, where $\hat{d}$ is obtained from bootstraps.

Tables 1 and 2 compare empirical percentiles and the size of various test statistics discussed above for the model (12). The sample sizes used were 1,000 and 2,000, respectively (these sample sizes are common among financial data). The corresponding quantiles of the asymptotic χ^2_8 are also given in the table. From the tables, we make the following observations. First, the T^* and bootstrap B statistics performed reasonably well when the sample size was sufficiently large. The bootstrap method outperformed the other test statistics, though it requires intensive computation. For instance, it took several hours to compute the bootstrap tests in Table 2 whereas it only took seconds to compute the other tests. Second, the T statistics underestimated the variance of cross-correlation so that the empirical quantiles exceeded their theoretical counterparts. Third, as expected, the S statistics performed poorly for both sample sizes considered. Fourth, the performance of the proposed test statistic T^* indicates that the Berlint and Francq (1997) method to estimate the variance of cross-covariance is reasonable in the presence of GARCH effects provided that robust estimators $\hat{\sigma}^*(i)$ are used.

Table 1. Empirical percentiles and size of various statistics for testing zero canonical correlations. The model used is (12). The results are based on 2,000 iterations, each with 1,000 observations.

Statistic	mean	S.D	**Percentile**					Rej. Rate at $\chi^2_8(0.95)$
			50%	75%	90%	95%	99%	percentage
S(1)	10.89	6.30	9.64	13.88	18.63	22.24	32.66	18.4
S(2)	11.06	7.26	9.58	13.8	19.42	23.39	35.83	19.1
T(1)	10.29	5.82	9.13	13.04	17.49	20.50	30.39	14.5
T(2)	8.73	5.26	7.76	10.99	14.67	17.99	25.60	8.3
χ^2_8	8.00	4.00	7.34	10.21	13.36	15.51	20.10	5.0
$T^*(1)$	8.03	4.16	7.38	10.22	13.29	15.76	20.94	5.3
$T^*(2)$	6.91	3.64	6.24	8.87	11.77	13.65	18.38	2.6
B(1)	8.02	4.17	7.34	10.20	13.67	15.50	21.80	5.0
B(2)	6.77	3.89	5.97	8.71	11.74	14.11	19.00	3.3

Table 2. Empirical Percentiles of Various Test Statistics for Testing Zero Canonical Correlations. The results are based on 2,000 replications and the sample size is 2,000.

Statistic	Mean	S.D	**Percentile**					Rej. Rate at $\chi^2_8(0.95)$
			50 %	75%	90%	95%	99%	percentage
$S(1)$	10.81	5.81	9.79	13.84	18.20	21.67	30.30	20.3
$S(2)$	11.63	8.91	9.72	14.68	20.94	25.94	37.48	22.2
$T(1)$	10.88	6.89	9.52	13.62	18.26	22.04	32.5	17.3
$T(2)$	9.14	6.66	7.82	11.55	15.94	19.27	28.66	10.8
χ^2_8	8	4	7.34	10.21	13.36	15.51	20.10	5.0
$T^*(1)$	8.13	4.29	7.39	10.43	13.82	16.35	21.60	6.5
$T^*(2)$	7.01	3.93	6.35	9.05	11.99	14.01	20.11	3.4
$B(1)$	7.72	4.1	7.04	9.75	13.05	15.32	20.81	4.8
$B(2)$	6.31	3.66	5.54	8.08	11.03	13.65	18.47	4.0

5.1. Case when the fourth moment of GARCH does not exist

The asymptotic results shown in this paper require that the fourth moments of the innovations exist. In this simulation, we show that this condition is indeed necessary. We use the same bivariate ARMA(1,1) model as before but the GARCH(1,1) model becomes

$$\boldsymbol{g}_t = \begin{bmatrix} g_{1t} \\ g_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + 0.5 \begin{bmatrix} a^2_{1,t-1} \\ a^2_{2,t-1} \end{bmatrix} + 0.4 \begin{bmatrix} g_{1t-1} \\ g_{2t-1} \end{bmatrix}.$$

It's easy to verify $E(a^4) = \infty$. Therefore $\mathrm{Var}(\hat{\rho})$, although it exists, cannot be estimated by the *Bartlett* formula. We expect the statistics T to fail in this case. Table 3 gives the results.

The test statistics $S(1)$, $S(2)$, $T(1)$ and $T(2)$ based on Bartlett's formula are completely off the mark here. This is expected, since the asymptotic variance of sample cross-correlations is invalidated when $Ea^4 = \infty$. However, with the variance estimated from stationary bootstrapping, the asymptotic χ^2 test has a Type I error which is close to the nominal value of 5%. The proposed variance estimate (13) overestimates the theoretical asymptotic variance of cross-correlation so that the T^* statistic is too small. The clear violation of the assumption of $Ea^4 < \infty$ for the asymptotic χ^2 test raises an open question of establishing the asymptotic distributions of the smallest canonical correlations when $Ea^4 = \infty$.

Table 3. Simulation results when the innovational series does not have a finite fourth moment. The results are based on 2,000 iterations, each with 1,000 observations.

Statistic	Mean	S.D	**Percentile**					Rej. Rate at $\chi^2_8(0.95)$
			50%	75%	90%	95%	99%	percentage
S(1)	22.08	21.45	16.07	25.68	40.58	56.57	110.51	52.8
S(2)	21.00	23.64	14.71	23.92	39.54	56.67	113.34	47.0
T(1)	21.09	20.96	15.41	24.02	38.82	53.23	110.12	49.6
T(2)	16.94	19.52	11.97	19.04	31.05	44.59	103.98	35.4
χ^2_8	8.00	4.00	7.34	10.21	13.36	15.51	20.10	5.0
$T^*(1)$	6.72	3.58	6.15	8.71	11.57	13.43	17.21	2.0
$T^*(2)$	6.33	3.50	5.73	8.39	11.00	12.83	16.88	1.7
B(1)	8.32	4.93	7.13	10.52	14.55	17.03	26.03	7.5
B(2)	7.17	4.56	6.13	9.11	13.07	15.96	22.62	5.6

6. An Illustrative Example

In this section we apply the proposed test statistics to a three-dimensional financial time series. The data consist of daily log returns, in percentages, of Amoco, IBM, and Merck stocks from February 2, 1984 to December 31, 1991 with 2,000 observations. The series are shown in Figure 1. It is well-known that daily stock return series tend to have weak dynamic dependence, but strong conditional heteroscedasticity, making them suitable for the proposed test; see Figures 2 and 3. Our goal here is to provide an illustration of specifying a vector ARMA model with GARCH innovations rather than a thorough analysis of the term structure of stock returns. As such, we do not seek to simplify the identified model by removing all insignificant parameters in estimation.

Denote the return series by $\mathbf{Z}_t = (Z_{1t}, Z_{2t}, Z_{3t})'$ for Amoco, IBM and Merck stocks, respectively. Following the order specification procedure of Section 2.2, we apply the proposed test (11), denoted by T^*, to the data and summarize the test results in Table 4. We also include the test statistics T from (9) for comparison.

The past vector $\boldsymbol{P}_t$ is determined by the AIC as $\boldsymbol{P}_t = (\boldsymbol{Z}'_{t-1}, \boldsymbol{Z}'_{t-2})'$. The p-value is based on a χ^2_{ks-f+1} test, where $k = 3$, $s = 2$, and $f = \dim(\boldsymbol{F}^*_t)$.

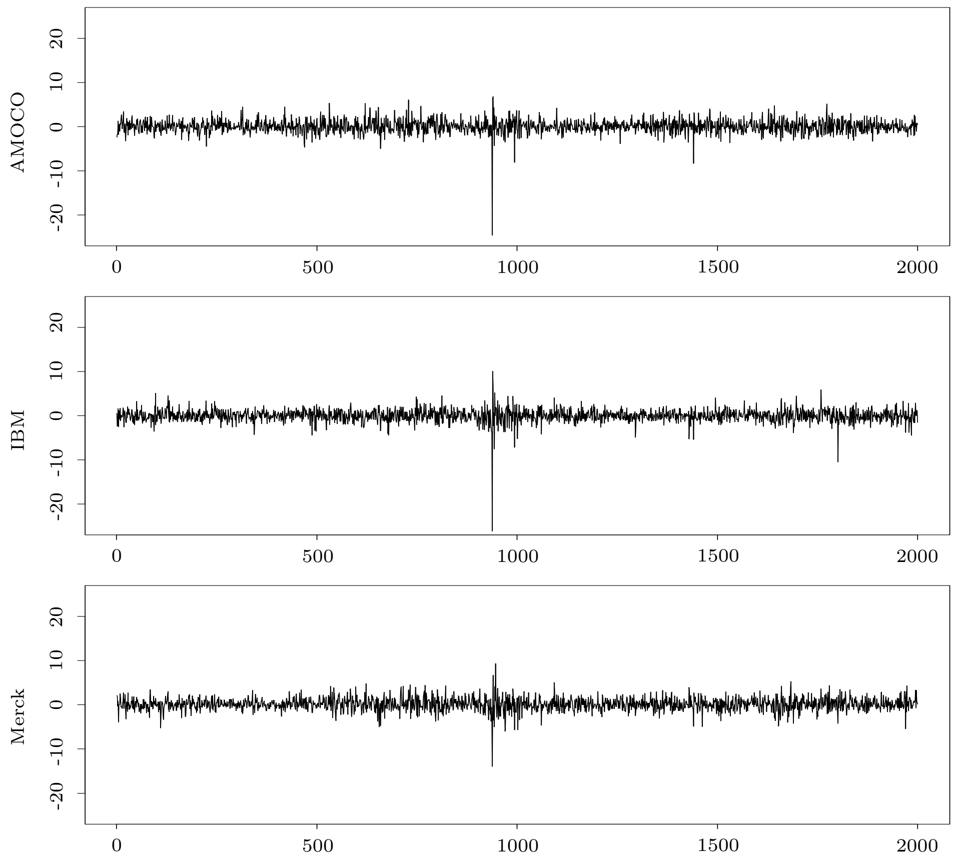

Figure 1. Time series of Amoco, IBM and Merck stocks daily return (2/2/1985 –12/31/1991).

From Table 4, the proposed test statistic T^* identified $\{1, 1, 1\}$ as the Kronecker indexes for the data, i.e., $K_i = 1$ for all i. On the contrary, if one assumes that there are no GARCH effects and uses the test statistic T, then one would identify $\{1, 1, 2\}$ as the Kronecker indexes. More specifically, the T statistic specifies $K_1 = K_2 = 1$, but finds the smallest canonical correlation between $\boldsymbol{F}^*_t = (Z_{1,t}, Z_{2,t}, Z_{3,t}, Z_{3,t+1})$ and $\boldsymbol{P}_t$ to be significant at the usual 5% level. To determine K_3, one needs to consider the canonical correlation analysis between $\boldsymbol{F}^*_t = (Z_{1,t}, Z_{2,t}, Z_{3,t}, Z_{3,t+1}, Z_{3,t+2})'$ and the past vector $\boldsymbol{P}_t$. The corresponding test statistic is $T = 4.05$, which is insignificant with p-value 0.134 under the asymptotic χ^2_2 distribution. Therefore, without considering GARCH effects, the identified Kronecker indexes are $K_1 = 1$, $K_2 = 1$, $K_3 = 2$, resulting in an ARMA(2,2) model for the data. Consequently, by correctly considering the GARCH effect, the proposed test statistic T^* was able to specify a more parsimonious ARMA(1,1) model for the data.

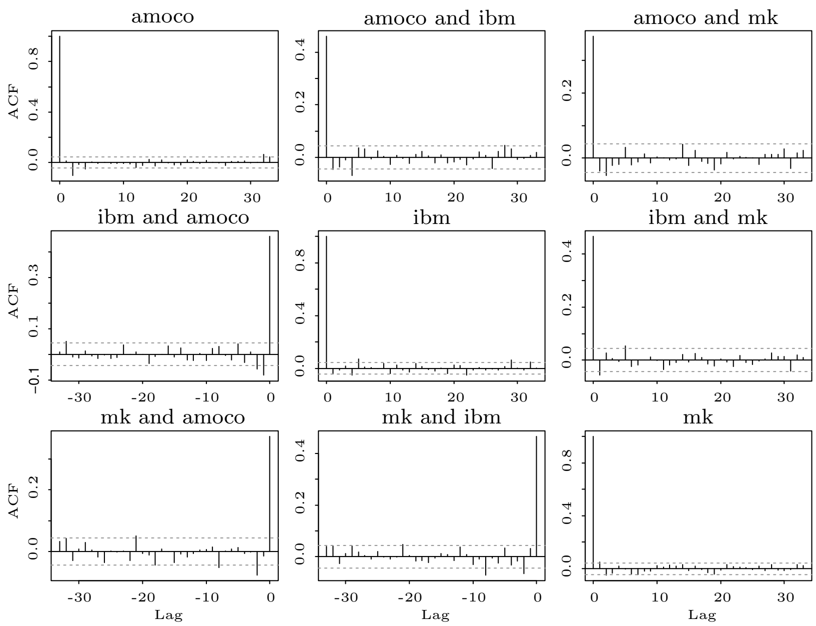

Figure 2. Auto- and cross-correlation plot of original series.

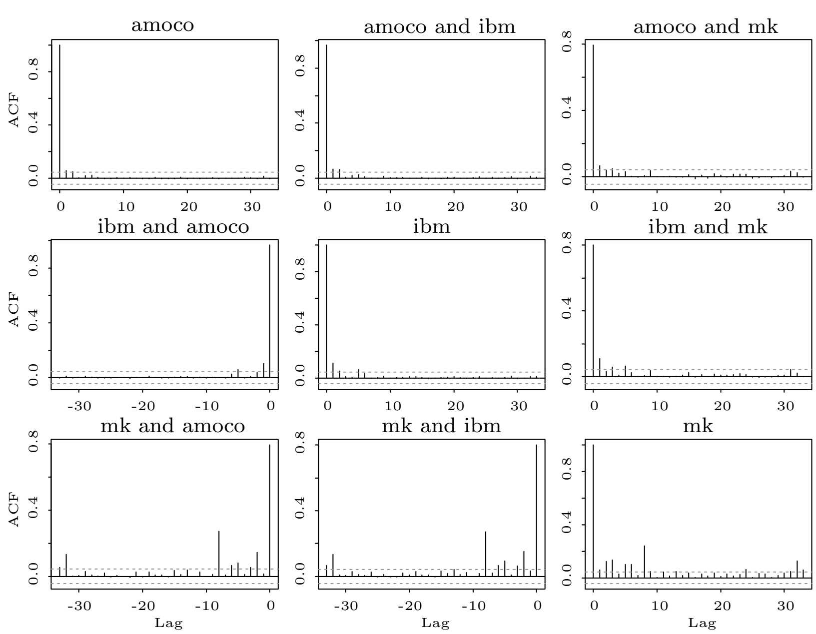

Figure 3. Auto- and cross-correlation functions of squared series.

Table 4. Model Specification of Daily Stock Returns Data (Amoco, IBM, Merck).

future subvector $\boldsymbol{F}_t^*$	can. cor(smallest)	T^*	d.f	p-value of T^*	Remark	T
$(Z_{1,t})$	0.130	33.957	6	0		33.957
$(Z_{1,t}, Z_{2,t})$	0.116	26.972	5	0		26.972
$(Z_{1,t}, Z_{2,t}, Z_{3,t})$	0.101	20.681	4	0		20.681
$(Z_{1,t}, Z_{2,t}, Z_{3,t}, Z_{1,t+1})$	0.051	5.588	3	0.133	$K_1 = 1$	5.945
$(Z_{1,t}, Z_{2,t}, Z_{3,t}, Z_{2,t+1})$	0.032	1.516	3	0.678	$K_2 = 1$	4.477
$(Z_{1,t}, Z_{2,t}, Z_{3,t}, Z_{3,t+1})$	0.055	5.976	3	0.113	$K_3 = 1$	11.384

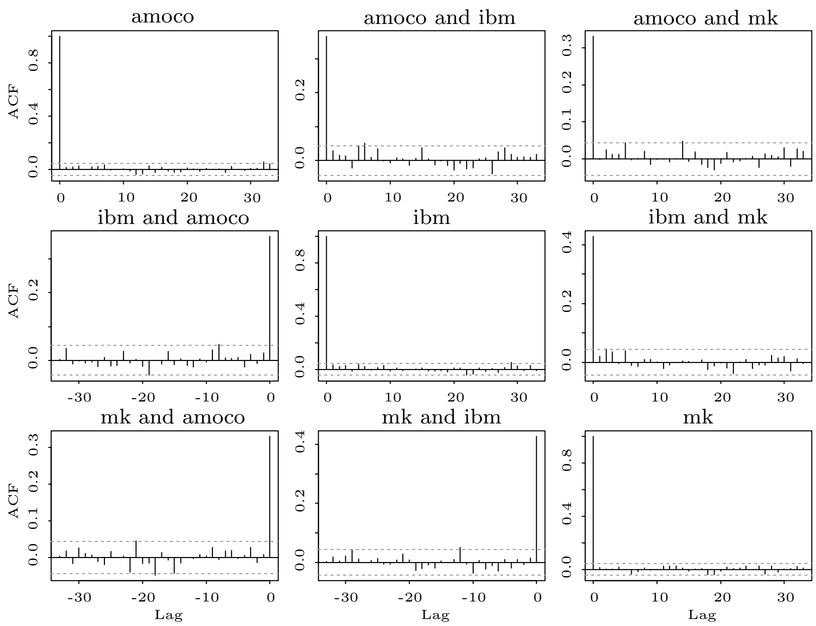

Figure 4. Auto- and cross-correlations of standardized residuals.

In summary, we entertain a vector ARMA(1,1) model with GARCH innovations for the data. For simplicity, we employ a diagonal GARCH(1,1) model for the innovations. The estimated VARMA(1,1)-GARCH(1,1) model is given below as

$$\boldsymbol{Z}_t - \begin{bmatrix} 0.005 & 1.970^{***} & 0.427 \\ 0.015 & 0.250 & 0.189^{*} \\ 0.141 & 0.894^{**} & 0.123 \end{bmatrix} \boldsymbol{Z}_{t-1}$$

$$= \begin{bmatrix} -0.060 \\ -0.009 \\ 0.093^{*} \end{bmatrix} + \boldsymbol{a}_t + \begin{bmatrix} -0.070 & 2.072^{***} & 0.445 \\ 0.057 & 0.242 & 0.233^{**} \\ 0.192^{*} & 0.896^{**} & 0.054 \end{bmatrix} \boldsymbol{a}_{t-1},$$

where the superscript *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively, and the volatility $\boldsymbol{g}_t = E(\boldsymbol{a}_t^2|\mathcal{F}_{t-1})$ follows the model

$$\boldsymbol{g}_t = \begin{bmatrix} 1.591 \\ 0.225 \\ 0.053 \end{bmatrix} + \begin{bmatrix} 0.278 & 0 & 0 \\ 0 & 0.138 & 0 \\ 0 & 0 & 0.060 \end{bmatrix} \boldsymbol{a}_{t-1}^2 + \begin{bmatrix} 0.003 & 0 & 0 \\ 0 & 0.756 & 0 \\ 0 & 0 & 0.914 \end{bmatrix} \boldsymbol{g}_{t-1},$$

where all estimates except the (1,1)th element of the coefficient matrix of $\boldsymbol{g}_{t-1}$ are significant at the 1% level. Finally, we examine the correlation matrices of the standardized residuals and their squared series of the fitted VARMA(1,1)-GARCH(1,1) model; see plots in Figures 4 and 5. The fitted model appears to be adequate in removing serial dependence in the data. The QQ-plot of the standardized residuals, however, show some departure from normality. This is common among asset return data.

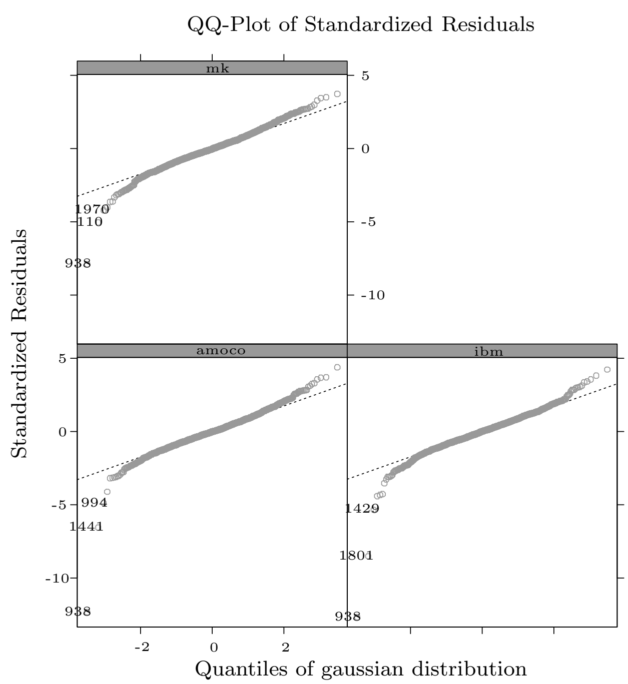

Figure 5. QQ plots of standardized residuals.

Acknowledgement

We would like to thank two referees and an associate editor for their careful reading and helpful comments that greatly improved the presentation of this paper. We also thank Professor W. B. Wu for helpful discussions. This research is supported in part by the Graduate School of Business, University of Chicago and National Science Foundation.

References

Akaike, H. (1976). Canonical Correlation Analysis of Time Series and The Use of An Information Criterion. In *Systems identification: Advances and Case Studies* (Edited by R. K. Methra and D. G. Lainiotis), 27-96. Academic Press, New York.

Berlinet, A. and Francq, C. (1997). On Bartlett's formula for non-linear processes. *J. Time Ser. Anal.* **18**, 535-552.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307-327.

Bollerslev, T., Engle, R. F and Nelson, D. B. (1994). ARCH models. In *Handbook of Econometrics* IV, pp 2959-3038. Elsevces Science B.V.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Box, G. E. P. and Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika* **64**, 355-365.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series Theory and Methods*. 2nd edition. Springer-Verlag, New York.

Cooper, D. M. and Wood, E. F. (1982). Identifying multivariate time series models. *J. Time Ser. Anal.* **3**, 153-164.

Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. *Adv. Appl. Probab.* **8**, 339-364.

Ding, Z. (1994). Time series analysis of speculative returns. Ph.D. dissertation, Dept. of Economics, University of California, San Diego.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987-1008.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error-correction: Representation, estimation and testing. *Econometrica* **55**, 251-276.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econom. Theory* **11**, 122-150.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. John Wiley, New York.

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika* **28**, 321-377.

Jewell, N. P. and Bloomfield, P. (1983). Canonical correlations of past and future for time series: Definitions and theory. *Ann. Statist.* **11**, 837-847.

Korensia, S. G. and Pukkila, T. M. (1987). Identification of non-zero elements in the polynomial matrices of mixed VARMA processes. *J. Roy. Statist. Soc. Ser. B* **49**, 112-126.

Quenouille, M. H. (1957). *The Analysis of Multiple Time Series*. Griffin, London.

Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York.

Robinson, P. M. (1973). Generalized canonical analysis for time series. *J. Multivariate Anal.* **3**, 141-160.

Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.* **91**, 590-600.

Tiao G. C. and Tsay, R. S. (1989). Model specification in multivariate time series (with discussion). *J. Roy. Statist. Soc. Ser. B* **51**, 157-213.

Tsay, R. S. (1989a). Identifying multivariate time series models. *J. Time Ser. Anal.* **10**, 357-371.

Tsay, R. S. (1989b). Parsimonious parametrization of vector autoregressive moving average models. *J. Bus. Econom. Statist.* **7**, 327-341.

Tsay, R. S. (1991). Two canonical forms for vector ARMA processes. *Statist. Sinica* **1**, 247-269.

Tsay, R. S. (2002). *Analysis of Financial Time Series.* John Wiley, New York.

Tsay, R. S. and Tiao, G. C. (1985). Use of canonical analysis in time series model identification. *Biometrika* **72**, 299-315.

Wu, W. B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9**, 809-831.

Mathematical Sciences Department, IBM T. J. Watson Research Center, Yorktown, NY 10598, U.S.A.

E-mail: wmin@galton.uchicago.edu

Graduate School of Business, University of Chicago, U.S.A.

E-mail: ruey.tsay@gsb.uchicago.edu

(Received November 2003; accepted April 2004)