Statistica Sinica **23** (2013), 873-899
doi:http://dx.doi.org/10.5705/ss.2010.219

MODEL IDENTIFICATION FOR TIME SERIES WITH DEPENDENT INNOVATIONS

Shuhao Chen, Wanli Min and Rong Chen

Bank of America, Google and Rutgers University

Abstract: This paper investigates the impact of dependent but uncorrelated innovations (errors) on the traditional autoregressive moving average model (ARMA) order determination schemes such as autocorrelation function (ACF), partial autocorrelation function (PACF), extended autocorrelation function (EACF), and the unit-root test. The ARMA models with iid innovations have been studied extensively and are well-posed, but their properties with dependent but uncorrelated innovations are relatively less studied. In the presence of such innovations, we show that the ACF, PACF, and EACF are significantly impacted while the unit-root test is not affected. We also propose a new order determination scheme to address those impacts for analyzing time series with uncorrelated innovations.

Key words and phrases: ACF, EACF, GARCH, PACF, order determination, time series, uncorrelated but dependent errors, unit-root test.

1. Introduction

Time series analysis is widely used in such fields as econometrics, finance, engineering, and metrology. The most commonly used time series model is the autoregressive moving average (ARMA) model. More specifically, $ARMA(p,q)$ model for a univariate time series X_t takes the form

$$\Phi(B)X_t = \Psi(B)\varepsilon_t, \tag{1.1}$$

with $\Phi(B) = \phi(B)U(B)$ where $\Phi(B) = 1-\Phi_1 B - \cdots - \Phi_p B^p$, $U(B) = 1-U_1B-\cdots-U_dB^d$, $\phi(B) = 1-\phi_1 B-\cdots-\phi_{p-d}B^{p-d}$ and $\Psi(B) = 1+\psi_1 B+\cdots+\psi_q B^q$ are polynomials in the back-shift operator B, defined as $BX_t = X_{t-1}$ and $B\varepsilon_t = \varepsilon_{t-1}$. We require that all zeros of $U(B)$ are on the unit circle and those of $\phi(B)$ and $\Psi(B)$ are outside the unit circle. ε_t is the innovation term. Note that when $U(B) = (1-B)^d$, model (1.1) is the autoregressive moving integrated moving average $ARIMA(p-d,d,q)$ model. If we further assume $\Psi(B) = 1$, model (1.1) is the $ARI(p-d,d)$ model.

The first step and also one of the key steps in building a time series model is order determination. In the literature, order determination schemes for time series models with identically independent distributed (*i.i.d.*) innovations$\{\varepsilon_t\}$ are well studied. Box and Jenkins (1976) introduced the autocorrelation function (ACF) and partial autocorrelation function (PACF). The Akaike Information Criterion (AIC) of Akaike (1974) and the Bayesian Information Criteria (BIC) of Schwarz (1978) are two goodness of fit measures of an estimated model that facilitate model selection. Tsay and Tiao (1984) proposed the extended autocorrelation function (EACF) for order determination of the $ARMA(p, q)$ model. On the other hand, Dickey and Fuller (1979) studied the unit-root behavior and gave the asymptotic distribution of a unit root test statistic. Standard order determination procedure combines those two techniques: taking the unit root test to decide whether to make difference(s)(*i.e.* set $Y_t = X_t - X_{t-1}$), and then using a ACF/PACF/EACF procedure on differenced series Y_t to get AR and MA orders p and q respectively. Other models and order determination schemes include the R and S array approach of Gray, Kelly, and McIntire (1978); the Corner method of Beguin, Gourieroux and Monfort (1980), and the Smallest Canonical Correlation (SCAN) of Tsay and Tiao (1984).

Most of the existing order determination methods assume that the innovation sequence $\{\varepsilon_t\}$ is *i.i.d.* and/or has constant conditional variance, excluding such interesting processes as the generalized autoregressive conditional heteroscedastic (GARCH) models of Bollerslev (1986) and the stochastic volatility (SV) model of Melino and Turnbull (1990), both often used in financial time series modeling. In these models, the innovation series is not autocorrelated but auto-dependent. Thus it becomes interesting and important to revisit the classical order determination schemes in the presence of auto-dependent but uncorrelated innovations.

Specifically, in this paper we study order determination schemes for MA, AR, ARMA and ARI models in the presence of uncorrelated but dependent innovations. Autocorrelation function (ACF) is a simple order determination procedure for MA process and for AR process partial autocorrelation function (PACF) becomes effective. For ARMA process, the extended autocorrelation function (EACF) has shown to be very useful in identifying the AR and MA orders, and it also works for the differenced sequence of an ARIMA process. We investigate how those schemes are impacted by the new type of innovations considered here.

There have been several studies in this area. Min (2004) investigated the effect of dependent innovations on pure AR and MA time series sequences; Yang and Zhang (2008) discussed the unit-root test with GARCH(1,1) innovations, one of the most used dependent but uncorrelated innovations. Our study is under a more unified framework that allows for a wide class of uncorrelated but dependent innovations.

The rest of the paper is organized as follows. In Section 2, we present a detailed study of how dependent but uncorrelated innovations affect the properties of the key statistics in classical time series model identification procedures.

Based on these findings, we propose suitable methods and establish their theoretical properties. Simulation studies and applications to examples are illustrated in Section 3. Proofs are given in the Appendix.

For simplicity, we employ the following notations and terms: $\xrightarrow{dist.}$ means convergence in distribution; $\xrightarrow{P}$ means convergence in probability; $\xrightarrow{a.s.}$ means almost sure convergence; $\triangleq$ means denotation; the term "errors" and "innovations" are used interchangeably. For a set of random variables $\{X_n\}$ and a corresponding set of function $\{f_n\}$, the notation $X_n = \mathcal{O}_p(f_n)$ means that X_n/f_n is bounded in the limit in probability.

2. Method

In this paper we consider the following structure of dependent but uncorrelated innovations. Let $\{e_t\}$ be *i.i.d.* random variables and F be a measurable function such that the innovation $\varepsilon_t = F(e_t, e_{t-1}, \ldots)$ is a well-defined random variable. We work with the following.

Condition C.1. $\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma^2, \forall\, t \in \mathbb{Z}$, where $\mathcal{F}_t \triangleq \sigma(e_t, e_{t-1}, \ldots)$ is the σ-field generated by the sequence $\{e_t\}$, representing the information available up to time t.

The error series $\{\varepsilon_t\}$ is uncorrelated since $\mathbb{E}(\varepsilon_t\varepsilon_{t-1}) = \mathbb{E}(\mathbb{E}(\varepsilon_t\varepsilon_{t-1}|\mathcal{F}_{t-1})) = \mathbb{E}(\varepsilon_{t-1}\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1})) = 0$. The second part of C.1 is weaker than the traditional condition $\mathbb{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$, which can be seen from the iterative expectation $\mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\mathbb{E}(\varepsilon_t^2|\mathcal{F}_{t-1}))$. The stronger condition $\mathbb{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ implies $\mathbb{C}ov(\varepsilon_t^2, \varepsilon_{t-1}^2) = 0$, while C.1 allows nonzero autocorrelations of $\{\varepsilon_t^2\}$.

Another important characterization of the uncorrelated and dependent innovation $\{\varepsilon_t\}$ is related to projections defined in Wu and Min (2005): $\mathcal{P}_t\eta \triangleq \mathbb{E}(\eta|\mathcal{F}_t) - \mathbb{E}(\eta|\mathcal{F}_{t-1})$. For instance, $\mathcal{P}_1\varepsilon_i\varepsilon_j \triangleq \mathbb{E}(\varepsilon_i\varepsilon_j|\mathcal{F}_1) - \mathbb{E}(\varepsilon_i\varepsilon_j|\mathcal{F}_0)$. Let $\|\cdot\|$ and $\|\cdot\|_p$ be the L_2 and L_p norm, respectively. We employ the following.

Condition C.2. $E(\varepsilon_n^4) < \infty$ and $\sum_{t=1}^{\infty} \|\mathcal{P}_1\varepsilon_t\|_4 < \infty$, $\sum_{i,j=0}^{\infty} \|\mathcal{P}_1\varepsilon_i\varepsilon_j\| < \infty$.

Remark 1. The intuition for C.2 is that the projection of the future $\varepsilon_i\varepsilon_j$ to the space $\mu_1 \ominus \mu_0 = \{Z : \sum_{i,j=0}^{\infty} \|\mathcal{P}_1 Z\| < \infty, Z \text{ is } \mathcal{F}_1 \text{ measurable and } \mathbb{E}(Z|\mathcal{F}_0) = 0\}$ has a small magnitude; namely, the future depends weakly on the current states or the current states depends weakly on the previous states.

This setup includes such interesting stochastic processes as the $ARCH(p)$ model, the $GARCH(p,q)$ model, and the Stochastic Volatility Model. For example, Min (2004) showed that the $GARCH(p,q)$ model satisfies C.2.

2.1. Moving average model

One of the fundamental building blocks of time series models is the moving average (MA) model. Specifically, a moving average process X_t with order q, $MA(q)$, has

$$X_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots + \psi_q \varepsilon_{t-q}, \tag{2.1}$$

a special case of the ARMA model (1.1) with $\Phi(B) = 1$. The autocovariance function and autocorrelation function (ACF) are

$$\gamma(h) = \mathbb{C}ov(x_t, x_{t+h}) \quad \text{and} \quad \rho(h) = \mathbb{C}or(x_t, x_{t+h}), \tag{2.2}$$

respectively. The ACF has the unique feature that for an $MA(q)$ process, the ACF cuts off at lag q. Since the ε_t are uncorrelated, we have $\gamma(h) = \sigma^2 \sum_{i=0}^{q} \psi_i \psi_{i+h}$, where $\psi_j = 0$ for $j > q$. Hence $\gamma(q+i) = \rho(q+i) = 0, \forall i > 0$. The ACF can be used effectively to determine the order q for an $MA(q)$ model.

Let $x_1, \ldots, x_n$ be a sample of the $MA(q)$ process as (2.1). We denote the sample autocovariance function and sample autocorrelation function, respectively, as

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} x_t x_{t+h}, \quad \text{and} \quad \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \tag{2.3}$$

For a linear process with finite fourth moment and innovation $\varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2)$, it is known (e.g., Brockwell and Davis (1986)) that $\sqrt{n}(\hat{\rho}(h) - \rho(h)) \xrightarrow{dist.} N(0, W)$ where W is given by the Bartlett formula

$$W = \sum_{k=-\infty}^{\infty} \{\rho^2(k) + \rho(h-k)\rho(h+k) + 2\rho^2(h)\rho^2(k) - 4\rho(h)\rho(k)\rho(k-h)\}. \tag{2.4}$$

If we further assume a $MA(q)$ model, the Bartlett formula implies that

$$W^* = 1 + 2\rho^2(1) + \cdots + 2\rho^2(q), \qquad \text{if } h > q.$$

However, when the innovations ε_t are auto-dependent, the validity of using $n^{-1}W^*$ to estimate the variance of sample ACF needs to be reexamined. In the following, we establish the asymptotic properties of $\hat{\rho}(h)$ under C.1 and C.2. Specifically, we show that the cut-off property (Lemma 1) and the asymptotic joint normality of $\hat{\gamma}(h)$ still hold, but the asymptotic variance is different (Theorem 1).

Lemma 1. *Let $\{X_t\}$ be an $MA(q)$ process with innovation $\{\varepsilon_t\}$ satisfying* C.1. *For the autocorrelation function $\rho(h)$ defined at* (2.2), $\rho(q+1) = \rho(q+2) = \cdots = 0$.

The proof is trivial since the cut-off property is only related to correlation, not dependence.

Theorem 1. *Let* $\{X_t\}$ *be an* $MA(q)$ *process with innovation* $\{\varepsilon_t\}$ *satisfying* C.1 *and* C.2. *For* $\rho(h)$ *defined at* (2.2) *and* $\hat{\rho}(h)$ *defined at* (2.3), $\forall h \geq 1$,

$$\sqrt{n}(\hat{\rho}(1) - \rho(1), \ldots, \hat{\rho}(h) - \rho(h))' \xrightarrow{dist.} N\Big(\underset{h\times 1}{\mathbf{0}}, \underset{h\times h}{\mathbf{\Sigma}}\Big),$$

where $\mathbf{\Sigma}$ *is given in the Appendix. In particular, for* $h > q$,

$$\mathbb{V}ar(\hat{\rho}(h)) \approx \frac{1}{n\gamma^2(0)}\big(\sigma(0,h) + 2\sigma(1,h) + \cdots + 2\sigma(q,h)\big), \tag{2.5}$$

where $\sigma(d,h) \triangleq E(x_0 x_h x_d x_{d+h})$. *If it is further assumed that the ACF of* $\{\varepsilon_t^2\}$ *is nonnegative and* $\mathbb{E}(\varepsilon_t^2 \varepsilon_{t-i}\varepsilon_{t-j}) = 0$, $\forall i \neq j$, *then for any* $h > q$,

$$\sqrt{n}\hat{\rho}(h) \xrightarrow{dist.} N(0, \delta_h^2), \quad \delta_h^2 \geq (1 + 2\sum_{k=1}^{q} \rho^2(k)), \tag{2.6}$$

where inequality is strict if the ACFs are positive.

The proof of the theorem is in the Appendix.

Remark 2. The theorem applies to the finite 4^{th} moment $GARCH(p,q)$ process and stochastic volatility models (see (3.1)) as they satisfy the conditions of the theorem and the additional positive ACF conditions.

Based on the theorem, one can estimate the standard error of $\hat{\rho}(h)$ by (2.5), with $\sigma(d,h)$ replaced by its estimate. However, the order of the MA process q is unknown. We use the estimator

$$\widehat{\mathbb{V}ar(\hat{\rho}(h))} = \frac{1}{n\hat{\gamma}^2(0)}\big(\hat{\sigma}(0,h) + 2\hat{\sigma}(1,h) + \cdots + 2\hat{\sigma}(h-1,h)\big). \tag{2.7}$$

This is due to the fact that $E(x_t x_{t+h} x_s x_{s+h}) = 0$ for $|s-t| > q$ for a $MA(q)$ process, hence we have $\sigma(q+1,h) = \sigma(q+2,h) = \cdots = \sigma(h-1,h) = 0$ for $h > q$. Thus including $\hat{\sigma}(q+1,h) = \hat{\sigma}(q+2,h) = \cdots = \hat{\sigma}(h-1,h)$ does not influence the estimator of the asymptotic variance.

In Section 3.1.1, we perform a comparison study between the variance calculation above and the standard variance formula $(1 + 2\hat{\rho}^2(1) + \cdots + 2\hat{\rho}^2(h-1))/n$ for *i.i.d.* innovations. Because of the differences in the variance, order identifications via ACF needs to be adjusted. A demonstration of this is given in Section 3.1.1.

It is to be noted that although our estimator is built for dependent but uncorrelated innovations, it applies to *i.i.d.* innovations automatically.

2.2. Autoregressive model

An autoregressive process X_t with order p, $AR(p)$, follows

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t.$$

It is a common practice to test the significance of the sample partial autocorrelation function (PACF) for identifying the order of $AR(p)$ series. The PACF $\phi_{h,h}$ is defined as

$$\phi_{0,0} = 1 \text{ and } \phi_{h,h} = \text{ the last component of } \boldsymbol{\phi}_h = \Gamma_h^{-1}\boldsymbol{\gamma}_h,\ \forall h \geq 1, \tag{2.8}$$

where $\Gamma_h = [\gamma(i-j)]_{i,j=1}^h$ and $\boldsymbol{\gamma}_h = (\gamma(1), \gamma(2), \ldots, \gamma(h))'$. It is estimated by replacing $\gamma(i)$ with its estimate $\hat{\gamma}(i)$ in the above formula.

It can be shown (e.g., Brockwell and Davis (1986)) that the partial autocorrelation $\phi_{h,h}$ at lag h may be regarded as the correlation between X_t and X_{t-h}, adjusted for the intervening observations $X_{t-1}, \ldots, X_{t-h+1}$. Since an $AR(p)$ model imposes a linear relationship between X_t and $X_{t-1}, \ldots, X_{t-p}$, it is easily seen that when $h > p$, X_t and X_{t-h} are conditionally uncorrelated, given $X_{t-1}, \ldots, X_{t-h+1}$, hence the cut-off property of PACF of AR processes.

If one further assumes Gaussian white noise, $\varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$, it is well known that $\forall h > p$, $s \triangleq h - p \geq 1$,

$$\sqrt{n}(\hat{\phi}_{p+1,p+1}, \ldots, \hat{\phi}_{p+s,p+s})' \xrightarrow{dist.} N\Big(\underset{s\times 1}{\mathbf{0}}, \underset{s\times s}{\mathbf{I}}\Big), \tag{2.9}$$

where I is the $s-$dimensional identity matrix. However, similar to the $MA(q)$ case, when $\{\varepsilon_t\}$ are dependent, the variance matrix of the asymptotic distribution of PACF is different from the *i.i.d.* case.

Lemma 2. *Let $\{X_t\}$ be an $AR(p)$ process with innovation $\{\varepsilon_t\}$ satisfying C.1. For the PACF $\phi_{h,h}$ (2.8), we have $\phi_{p+1,p+1} = \phi_{p+2,p+2} = \cdots = 0$.*

In analogy to (2.9), the next theorem establishes the asymptotic properties of sample PACF $\hat{\phi}_{h,h}$, $h > p$ of an $AR(p)$ process with dependent innovations.

Theorem 2. *Let $\hat{\phi}_{h,h}$ be the lag h sample PACF of a stationary and invertible $AR(p)$ time series $\{X_t\}$, where the innovations $\{\varepsilon_t\}$ satisfy* C.1 *and* C.2. *Then $\forall\, h > p$ $(s = h - p \geq 1)$,*

$$\sqrt{n}(\hat{\phi}_{p+1,p+1}, \ldots, \hat{\phi}_{p+s,p+s})' \xrightarrow{dist.} N\Big(\underset{s\times 1}{\mathbf{0}}, \underset{s\times s}{\boldsymbol{\Xi}}\Big).$$

where $\boldsymbol{\Xi}$ is given in Appendix. If we further assume ACF of $\{\varepsilon_t^2\}$ is nonnegative and $\mathbb{E}(\varepsilon_t^2\varepsilon_{t-i}\varepsilon_{t-j}) = 0$, $\forall i \neq j$, then.

$$\boldsymbol{\Xi}_{(\ell,\ell)} \geq 1, \quad \forall\, \ell \in (1, \ldots, s). \tag{2.10}$$

strict inequality holds in (2.10) *if all ACFs are positive.*

Remark 3. Dependent but uncorrelated innovations, such as GARCH innovations, render a larger variance of sample PACF compared with an asymptotic variance of $1/n$ in the case of *i.i.d.* innovations. So if we still use the traditional $(1-\alpha)\%$ confidence interval to test for lag-h PACF with $h > p$, the type I error is larger than $\alpha\%$. This over- rejection leads to a specification of $AR(h)$ model with h greater than the true order p.

We can obtain the variance estimator of the sample lag-h PACF $\mathbb{V}ar(\hat{\phi}_{h,h})$ in the following manner. First we obtain the estimates $\hat{\phi}_1, \ldots, \hat{\phi}_h$ and the corresponding residuals

$$\hat{\varepsilon}_t = x_t - \hat{\phi}_1 x_{t-1} - \cdots - \hat{\phi}_h x_{t-h}.$$

Secondly, from the sample autocovariance matrix $\hat{\Gamma}_h$ and the sample version of $\Omega_h \triangleq \mathbb{C}ov(\varepsilon_1 x_0, \varepsilon_1 x_{-1}, \ldots, \varepsilon_1 x_{1-h})'$, we obtain an estimate of $\Gamma_h^{-1}\Omega_h\Gamma_h^{-1}$ whose last diagonal entry over n is the variance estimator. See the proof of Theorem 2 in Appendix for more details.

In Section 3.1.2, we perform a comparison study between variance calculation above and the standard variance $1/n$ in the i.i.d. case. It demonstrates that the order identification scheme for $AR(p)$ process using PACF needs to be adjusted to the updated variance. The illustration of order identifications via PACF corresponding to different innovations is seen in Section 3.1.2.

2.3. Autoregressive and moving average model

The ARMA model in (1.1) is a hybrid of autoregressive and moving average models. Tsay and Tiao (1984) proposed the extended autocorrelation function (EACF) technique to identify the orders of a stationary or non-stationary ARMA process based on iterated least square estimates of the autoregressive parameters. This is based on the fact that, if X_t follows an $ARMA(p, q)$ model, the "filtered process" $Y_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}$ follows an $MA(q)$ process $Y_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \cdots + \psi_q\varepsilon_{t-q}$.

Under the dependent but uncorrelated innovations, in analogy to the procedure described in Tsay and Tiao (1984), an order determination scheme is proposed as follows.

1. For each candidate AR order s, obtain a consistent estimate of the AR coefficients, $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_s$.
2. Obtain $Y_t = X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2} - \cdots - \hat{\phi}_s X_{t-s}$. If $s = p$, then Y_t should approximately follow an $MA(q)$ sequence. On the other hand, if we over-fit the AR term by m (i.e. $s = p + m$), then Y_t should be an $MA(q+m)$ process.
3. Use the order determination procedures for moving average sequences described above on $\{Y_t\}$ to identify the MA order, build an EACF table as

proposed in Tsay and Tiao (1984), and finalize the AR and MA order selections.

Since the major difference of this scheme with that of Tsay and Tiao (1984) is dealing with the significance of ACF elaborated in Section 2.1, we skip the discussion here. The impact on EACF process inference of different innovation terms is illustrated in Section 3.1.3.

2.4. Autoregressive integrated model

The autoregressive integrated (ARI) model can be regarded as an extension of the AR model. The Augmented Dickey Fuller (ADF) test (Dickey and Fuller (1979), Said and Dickey (1984)) is important in detecting whether the process is unit-root nonstationary. In this section we study ADF under uncorrelated but dependent innovations.

Consider the ARI model,

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \varepsilon_t. \tag{2.11}$$

Following the formulation of ADF, we write (2.11) as

$$X_t = \rho X_{t-1} + \sum_{k=1}^{p-1} a_k \Delta X_{t-k} + \varepsilon_t, \tag{2.12}$$

where $\Delta X_{t-k} \triangleq X_{t-k} - X_{t-k-1}$, $\rho = \sum_{i=1}^{p} \Phi_i$, and $a_k = -\sum_{j=k+1}^{p} \Phi_j$. The unit-root problem for this model focuses on testing

$$H_0 : \rho = 1 \text{ against } H_1 : \rho_{min} < \rho < 1.$$

where ρ_{min} is the smallest of $\sum_{i=1}^{p} \Phi_i$ such that model (12) is stationary.

The ADF test statistic is $\widehat{T}_n \triangleq n(\widehat{\rho}_n - 1)$, where $\widehat{\rho}_n$ is the least square estimator of ρ when X_t is regressed on $X_{t-1}, \Delta X_{t-1}, \ldots, \Delta X_{t-p+1}$ using (2.12),

$$\widehat{\rho}_n = \frac{\sum_{t=p+1}^{n} X_{t-1}(X_t - \sum_{k=1}^{p-1} \hat{a}_k \Delta X_{t-k})}{\sum_{t=p+1}^{n} X_{t-1}^2}.$$

Dickey and Fuller (1979) proved that, under the null hypothesis,

$$\widehat{T}_n = n(\widehat{\rho}_n - 1) \overset{dist.}{\to} \frac{1}{2}(1 - \sum_{k=1}^{p-1} a_k)(\Gamma^{-1}(T^2 - 1)),$$

where $(\Gamma, T) = (\sum_{i=1}^{\infty} \gamma_i^2 Z_i^2, \sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i)$, $\gamma_i^2 = 4[(2i-1)\pi]^{-2}$, and $\{Z_i\}_{i=1}^{\infty}$ is a sequence of *i.i.d.* $N(0,1)$ random variables. Chan and Wei (1988) showed that the limiting distribution could also be further represented in terms of function of Brownian Motion $W(t)$,

$$\widehat{T}_n = n(\widehat{\rho}_n - 1) \xrightarrow{dist.} \frac{1}{2}(1 - \sum_{k=1}^{p-1} a_k) \frac{(W^2(1) - 1)}{\int_0^1 W^2(t)dt}.$$

When the innovation $\{\varepsilon_t\}$ is not a sequence of *i.i.d.* random variables but dependent and uncorrelated random variables satisfying C.1 and C.2, we show in the following that the same limiting distribution of the test statistics $\widehat{T}_n$ applies.

Before we state the results, we need the following definition from Wu and Min (2005).

Definition 1. Let $\{\omega_t\}_{t\in\mathcal{Z}}$ be independent and identically distributed random elements. The process $Y_n = g(\omega_n, \omega_{n-1}, \ldots)$, where g is a measurable function, is said to be $\mathcal{L}^p (p \geq 0)$ weakly dependent with order $r(r \geq 0)$ if $\mathbb{E}(|Y_n|^p) < \infty$ and

$$\sum_{n=1}^{\infty} n^r \|\mathcal{P}_1 Y_n\|_p < \infty. \tag{2.13}$$

If (2.13) holds with $r = 0$, then Y_n is said to be $\mathcal{L}^p$ weakly dependent.

Condition C.3. The $\{\varepsilon_t\}$ process is $\mathcal{L}^\alpha$ weakly dependent with order 1, for some $\alpha > 2$.

This innovation assumption includes a large class of nonlinear processes, and substantially relaxes the *i.i.d.* or martingale difference assumption. In particular, it is satisfied if the innovations are GARCH, random coefficient AR, bilinear AR, etc. For more details, see Wu and Min (2005).

The next theorem indicates that the limiting distribution of the test statistic $\widehat{T_n}$ does not change if we use dependent but uncorrelated errors.

Theorem 3. *If* $\rho = 1$*, for* $\{X_t\}$ *following* (2.11)*, and errors* $\{\varepsilon_t\}$ *satisfying* C.1, C.2.*, and* C.3*, we have*

$$\widehat{T}_n \xrightarrow{dist.} (1 - \sum_{k=1}^{p-1} a_k) \frac{\int_0^1 W(t)dW(t)}{\int_0^1 W^2(t)dt}.$$

Remark 4. According to Theorem 10.1.2 of Fuller (1995), $\hat{a}_k$ is a consistent estimator of a_k and we can use $\hat{a}_k$ in estimating the limiting distribution, where $(\hat{\rho}, \hat{a}_1, \ldots, \hat{a}_k)'$ is the regression coefficients obtained by regressing X_t on X_{t-1}, $\Delta X_{t-1}, \cdots, \Delta X_{t-p+1}$.

Remark 5. Theorem 3 cannot be easily extended to ARIMA models since (2.11) is no longer a simple regression model when the MA part is involved. Further study would be useful here.

3. Empirical Study

In this section, we use simulation and real applications to demonstrate the results and methods developed in the previous section. Simulation results for AR/MA/ ARMA/ARI are provided first to validate the theorems. Only part of the results are presented here. Results with different parameters of the GARCH model and results with other innovation models such as the stochastic volatility model would be given upon request. For applications, we go through the entire order determination and estimation process. We denote the EACF proposed by Tsay and Tiao (1984) as *original* EACF and the adjusted EACF for time series associated with uncorrelated and dependent innovations as *modified* EACF.

3.1. Simulation

Here we use $GARCH(1,1)$ innovations as the dependent but uncorrelated errors since GARCH is among the most common and useful models for this type. Specifically, we consider

$$\begin{cases} X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \cdots + \psi_q \varepsilon_{t-q}, \\ \varepsilon_t = \sqrt{g_t}\omega_t, \quad g_t = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta g_{t-1}, \end{cases}$$

where $\{\omega_t\}$ are *i.i.d.* $N(0,1)$.

3.1.1. Moving average model

We consider two formulae for estimating $\mathbb{V}ar(\hat\rho(h))$:

- $V(h) = n^{-1}(1 + 2\hat\rho^2(1) + \cdots + 2\hat\rho^2(h-1))$;
- $V^*(h) = (n\hat\gamma^2(0))^{-1}(\hat\sigma(0,h) + 2\hat\sigma(1,h) + \cdots + 2\hat\sigma(h-1,h))$.

Correspondingly, we denote the statistics $\hat\rho^2(h)/V(h)$ and $\hat\rho^2(h)/V^*(h)$ as $T(h)$ and $T^*(h)$ respectively.

First, we use an $MA(1)$ model for simulation.

$$\begin{cases} X_t = \varepsilon_t + \psi_1 \varepsilon_{t-1}, \\ \varepsilon_t = \sqrt{g_t}\omega_t, \quad g_t = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta g_{t-1}, \end{cases}$$

with $\psi_1 = -0.4$ and GARCH effect setting $\alpha = 0.2$ and $\beta = 0.7$.

We simulated 2,000 replications of time series with length 1,000 each. In Table 1, we report the empirical percentiles of $T(h)$ and $T^*(h)$ along with that of χ_1^2, for $h = 2$ and 3. Since the series were simulated from $MA(1)$, the true value of ρ should be zero and asymptotically the statistics should follow χ_1^2 if the variance is estimated correctly. The last column shows the percentage of rejections if $\chi_{.95}^2$ is used as the critical value for testing non-zero ACF coefficients.

Table 1. Empirical percentile of $T(h)$ and $T^*(h)$ for a simulated MA(1) series $X_t = \varepsilon_t - 0.4\varepsilon_{t-1}$ where innovation $\{\varepsilon_t\}$ is GARCH(1,1) with $\alpha = 0.2$, $\beta = 0.7$.

	Mean	S.D.	50%	75%	90%	95%	99%	p
$T(2)$	1.852	2.820	0.804	2.340	4.828	7.465	13.012	14.30%
$T(3)$	1.821	2.938	0.786	2.309	4.917	7.197	12.231	13.70%
χ_1^2	1.000	1.414	0.454	1.323	2.705	3.841	6.634	-
$T^*(2)$	1.003	1.402	0.478	1.334	2.656	3.731	6.553	4.55%
$T^*(3)$	1.007	1.332	0.484	1.330	2.806	3.776	6.084	4.45%

Table 2. Empirical percentile of $T(h)$ and $T^*(h)$ for a simulated MA(3) series $X_t = \varepsilon_t + 0.8\varepsilon_{t-1} - 0.8\varepsilon_{t-2} + 0.8\varepsilon_{t-3}$ where innovation ε_t is GARCH(1,1) with $\alpha = 0.5$, $\beta = 0.2$.

	Mean	S.D.	50%	75%	90%	95%	99%	p
$T(2)$	5.498	7.670	3.001	7.583	13.297	18.685	30.439	43.25%
$T(3)$	70.005	21.205	69.070	83.035	96.374	105.754	126.570	100.00%
$T(4)$	1.840	3.066	0.745	2.175	4.836	7.609	14.818	12.55%
$T(5)$	1.566	2.745	0.666	1.937	3.967	5.774	11.916	10.50%
χ_1^2	1.000	1.414	0.454	1.323	2.705	3.841	6.634	-
$T^*(2)$	2.286	2.549	1.426	3.325	5.992	7.739	10.941	21.80%
$T^*(3)$	26.811	10.535	26.912	33.862	40.325	44.156	51.309	98.75%
$T^*(4)$	1.035	1.448	0.505	1.410	2.682	3.670	6.479	3.85%
$T^*(5)$	1.053	1.416	0.520	1.473	2.757	3.800	6.80	4.35%

It can be seen from Table 1 that $T^*(h)$ closely follows the chi-square distribution while $T(h)$ is much larger. Since $\hat{\rho}(h)$ is the same for both $T(h)$ and $T^*(h)$, the variance estimate $V^*(h)$ is more accurate. The level of testing using $T^*(h)$ is also more accurate.

In order to compare $T(h)$ and $T^*(h)$ when $h \le q$, we use an $MA(3)$ model for a second simulation:

$$\begin{cases} X_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \psi_3\varepsilon_{t-3}, \\ \varepsilon_t = \sqrt{g_t}\omega_t, \quad g_t = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta g_{t-1}, \end{cases}$$

with $\psi_1 = 0.8, \psi_2 = -0.8, \psi_3 = 0.8$, and the two GARCH effect settings $\alpha = 0.1, \beta = 0.8$ and $\alpha = 0.5, \beta = 0.2$.

Again, we simulated 2,000 replications of the time series with length 1,000 each. The empirical percentiles of $T(h)$ and $T^*(h)$ with those of χ_1^2 are compared in Table 2.

We make the following observations.

1. For $h \le q$, both $T(h)$ and $T^*(h)$ differ significantly from χ_1^2 since $\hat{\rho}^2(h)/\widehat{\mathbb{V}ar(\hat{\rho}(h))}$ is asymptotically non-central χ^2 distributed since the asymptotic mean of $\hat{\rho}(h)$ is $\rho(h) \neq 0$ when $h \le q$.

Table 3. EACF method comparison for a simulated MA(3) series with different types of innovations: $X_t = \varepsilon_t + 0.8\varepsilon_{t-1} - 0.8\varepsilon_{t-2} + 0.8\varepsilon_{t-3}$ with innovations ε_t that were: (1) iid; (2) GARCH(1,1) with $\alpha = 0.1, \beta = 0.8$; (3) GARCH(1,1) with $\alpha = 0.5, \beta = 0.2$. The time series length was 1,000 while the replication was 1,000 as well. The numbers (in %) are the percentages which the order, row, identified by the method, column, for the simulated series. Each column sums to 100%.

	Innovation (1)		Innovation (2)		Innovation (3)	
Models	original	modified	original	modified	original	modified
MA(2)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
MA(3)	85.10%	89.80%	76.50%	87.70%	69.40%	85.00%
MA(4)	6.60%	4.10%	9.30%	5.90%	13.90%	5.10%
MA(5)	2.60%	1.90%	6.10%	2.20%	6.10%	1.90%
ARMA(1,3)	0.20%	0.00%	0.30%	0.10%	0.40%	0.20%
ARMA(1,4)	2.10%	1.60%	3.00%	0.50%	3.80%	0.00%
ARMA(1,5)	0.00%	0.00%	0.00%	0.00%	0.00%	1.00%
ARMA(2,3)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ARMA(2,4)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ARMA(2,5)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ARMA(3,3)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ARMA(3,4)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Others	3.40%	2.60%	4.80%	3.60%	6.40%	6.80%

2. $T(3)$ and $T^*(3)$ are relatively greater than the others simply because $\rho(3)$ is a larger value. Theoretically, $\rho(1) = -0.164$, $\rho(2) = -0.055$, $\rho(3) = 0.274$, $\rho(4) = \rho(5) = 0$.
3. For $h > q$, the performance of $T(h)$ and $T^*(h)$ are similar to what we observed in MA(1) simulation. $T^*(h)$ follows the χ^2_1 distribution closely while $T(h)$ is much larger.
4. It is clear that, while the power of testing using $T^*(h)$ is comparable with that of $T(h)$, its size is more accurate. Since order determination is based on the change from non-zero ACF coefficients to zero ACF coefficients, $T^*(h)$ is certainly more reliable.

The simulation confirmed that the variance of $\hat{\rho}(h)$ may change when faced with dependent innovations. Hence the new order identification scheme discussed in previous section is needed. Since the MA model is a specific case of the $ARMA$ model, we incorporate the interpretation of the results of the new scheme compared to the classical one into our discussion of the $ARMA$ model and only show the result table here (Table 3). All the interpretation of Section 3.1.3 apply.

Table 4. Empirical percentile of $T(h)$ and $T^*(h)$ for a simulated AR(1) series $X_t = 0.9X_{t-1} + \varepsilon_t$ where innovation $\{\varepsilon_t\}$ is GARCH(1,1) with $\alpha = 0.5$, $\beta = 0.2$.

	Mean	S.D.	50%	75%	90%	95%	99%	p
$T(2)$	3.633	8.808	1.339	3.977	8.459	14.096	30.992	17.45%
$T(3)$	2.449	4.661	0.948	2.874	6.307	9.032	20.378	10.90%
χ^2_1	1.000	1.414	0.454	1.323	2.705	3.841	6.634	-
$T^*(2)$	1.084	1.597	0.488	1.354	2.891	4.010	8.275	2.95%
$T^*(3)$	1.023	1.455	0.456	1.327	2.776	3.848	7.275	5.75%

3.1.2. Autoregressive model

For comparing the variance estimator under an $i.i.d$ assumption and our proposed estimator, we consider two formulae for estimating the variance $\mathbb{V}ar(\hat{\phi}_{h,h})$:

- $V(h,h) = \frac{1}{n}$;
- $V^*(h,h) = n^{-1}[\hat{\Gamma}_h^{-1}\hat{\Omega}_h\hat{\Gamma}_h^{-1}]_{(h,h)}$.

Correspondingly, we denote the statistics $\hat{\phi}^2_{h,h}/V(h,h)$ and $\hat{\phi}^2_{h,h}/V^*(h,h)$ as $T(h)$ and $T^*(h)$ respectively. We used the $AR(1)$ model for the first simulation experiment in this setting:

$$\begin{cases} X_t = \phi_1 X_{t-1} + \varepsilon_t, \\ \varepsilon_t = \sqrt{g_t}\omega_t, \quad g_t = \alpha_0 + \alpha\varepsilon^2_{t-1} + \beta g_{t-1}, \end{cases}$$

with $\phi_1 = 0.9$ and GARCH effect setting $\alpha = 0.5$ and $\beta = 0.2$.

We simulated 2,000 replications of time series with length 1,000 each. The empirical percentiles of $T(h)$ and $T^*(h)$ with those of χ^2_1 are listed in the Table 4. It shows very similar results to those of Table 1 and leads to similar conclusions.

For $h \le p$, we use an $AR(3)$ model for a second simulation.

$$\begin{cases} X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \varepsilon_t, \\ \varepsilon_t = \sqrt{g_t}\omega_t, \quad g_t = \alpha_0 + \alpha\varepsilon^2_{t-1} + \beta g_{t-1}, \end{cases}$$

with $\phi_1 = 0.8, \phi_2 = -0.8, \phi_3 = 0.8$ and the two GARCH effect settings $\alpha = 0.1, \beta = 0.8$ and $\alpha = 0.5, \beta = 0.2$.

Again, we simulated 2,000 replications of time series with length 1,000 each. The empirical percentiles of $T(h)$ and $T^*(h)$ with those of χ^2_1 are compared in the Table 5. Again, we see very similar results as those in Table 2.

The confirmation that the variance of $\hat{\phi}_{h,h}$ may change when faced with dependent innovations indicates the need to use the new order identification scheme discussed in previous section. Since the AR model is a specific case of the $ARMA$ model, indicates the incorporate the interpretation of the results of

Table 5. Empirical percentile of $T(h)$ and $T^*(h)$ for a simulated AR(3) series $X_t = 0.8X_{t-1} - 0.8X_{t-2} + 0.8X_{t-3} + \varepsilon_t$ where innovation ε_t is GARCH(1,1) with $\alpha = 0.5$, $\beta = 0.2$.

	Mean	S.D.	50%	75%	90%	95%	99%	p
$T(2)$	210.077	63.594	203.664	249.110	290.703	321.813	380.473	100.00%
$T(3)$	622.774	40.188	625.208	649.088	669.552	682.707	708.027	100.00%
$T(4)$	3.326	8.237	1.238	3.897	8.217	12.303	23.530	23.30%
$T(5)$	2.112	3.953	0.813	2.597	5.296	7.835	17.106	8.30%
χ_1^2	1.000	1.414	0.454	1.323	2.705	3.841	6.634	-
$T^*(2)$	55.874	26.798	53.597	72.045	90.328	102.028	127.673	99.55%
$T^*(3)$	178.789	65.745	182.865	226.146	259.340	278.116	320.841	100.00%
$T^*(4)$	1.224	1.998	0.495	1.634	3.197	4.583	8.377	2.85%
$T^*(5)$	1.011	1.408	0.433	1.322	2.824	3.793	6.533	6.35%

Table 6. EACF method comparison for a simulated AR(3) series with different types of innovations: $X_t = 0.8X_{t-1} - 0.8X_{t-2} + 0.8X_{t-3} + \varepsilon_t$ with innovations ε_t that were (1) iid; (2) GARCH(1,1) with $\alpha = 0.1, \beta = 0.8$; (3) GARCH(1,1) with $\alpha = 0.5, \beta = 0.2$. The time series length was 1,000 while the replication is 1,000 was well. The numbers (in %) are the percentage which the order, row, identified by the method, column, for the simulated series. Each column sums to 100%.

	Innovation (1)		Innovation (2)		Innovation (3)	
Models	original	modified	original	modified	original	modified
AR(2)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ARMA(2,1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ARMA(2,2)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AR(3)	57.90%	64.80%	48.40%	69.20%	33.60%	70.60%
ARMA(3,1)	8.50%	10.40%	10.00%	8.60%	16.50%	7.40%
ARMA(3,2)	5.30%	5.00%	7.70%	4.10%	10.10%	3.90%
AR(4)	0.70%	0.60%	0.90%	0.50%	0.60%	1.00%
ARMA(4,1)	18.20%	13.70%	19.30%	13.10%	18.20%	12.60%
ARMA(4,2)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
AR(5)	1.20%	0.30%	1.30%	0.30%	1.60%	1.10%
ARMA(5,1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ARMA(5,2)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Others	8.20%	5.20%	12.40%	4.20%	19.40%	3.40%

the new scheme compared to the classical one into our discussion of the $ARMA$ model and only report the results in Table 6. All the interpretation of Section 3.1.3 can still be applied.

3.1.3. Autoregressive moving average model

Here we compare the different original and modified inferences of EACF on time series with either *i.i.d.* innovations or uncorrelated but dependent innovations. The experiment was designed as follows.

1. Simulate $ARMA(1,1)$ time series with innovations that are *i.i.d.*, GARCH innovations (with two parameter settings), and that follow a stochastic volatility model (to be specified). The length of the time series is 1,000, with 1,000 replications.
2. For each time series generated, estimate the EACF table and use the original inference procedure (under *i.i.d.* assumption) to identify the model order.
3. Of the 1,000 replications, count the frequency of each candidate models selected by the original EACF procedure.
4. Repeat step 2 and 3 using the modified EACF procedure.

The parameter for the $ARMA(1,1)$ mean equation is $(0.8, 0.5)$ for the AR and MA coefficients. We use innovations that are (i) *i.i.d.*; (ii) GARCH(1,1) with $\alpha = 0.1$, $\beta = 0.8$; (iii) GARCH(1,1) with $\alpha = 0.5$, $\beta = 0.2$ and (iv) that follow stochastic volatility with $\alpha_1 = 0.5$ where the model is

$$\varepsilon_t = \sigma_t e_t, \quad (1 - \alpha_1 B)\ln(\sigma_t^2) = v_t, \tag{3.1}$$

where e_t are *i.i.d.* $N(0,1)$, v_t are *i.i.d.* $N(0, \sigma_v^2)$, $\{e_t\}$ and $\{v_t\}$ are independent.

The reason we simulated two settings of the GARCH model is that the GARCH effects differ. Min and Tsay (2005) showed that $\mathbb{V}ar(\hat{\rho})$ is significantly impacted if $\mathbb{C}ov(\varepsilon_0^2, \varepsilon_{q-k+i}^2)/\mathbb{E}^2(\varepsilon_0^2)$ is large. For instance, if $\{\varepsilon_t\}$ is a GARCH(1,1) process, as in our simulations, $\mathbb{C}ov(\varepsilon_0^2, \varepsilon_1^2)/\sigma^4 = 2\alpha + (6\alpha^2(\alpha + \beta/3))/(1 - 2\alpha^2 - (\alpha + \beta)^2)$. Given $\alpha = 0.5$ and $\beta = 0.2$ the ratio is 86, and with $\alpha = 0.1$ and $\beta = 0.8$, the ratio is 0.33.

The simulation results are reported in Table 7, showing the percentage of a candidate model being selected based on either the original inference or the modified inference of the EACF table.

From Table 7, it is seen that the modified EACF procedure has similar performance as the original EACF for pure ARMA models, but outperforms it for ARMA+GARCH Models. We ran a set of formal tests to see it more clearly. Since the same sets of simulated time series are used for original EACF and modified EACF procedure, we created a 2×2 contingency table for each innovation type and run chi-square test to compare identification power. The results are shown in Table 8 with the table and corresponding chi-square statistic and its p-value, for each innovation type.

From Tables 7 and 8, we see that the procedures work equally well with *i.i.d.* innovations (large p-value), confirming that the two algorithms are almost equivalent for *i.i.d.* error terms. With GARCH innovation but relatively weak dependence (ii), the improvement by the modified EACF compared to original

Table 7. EACF method comparison for a simulated ARMA(1,1) series with different types of innovations: $X_t = 0.8X_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$ with innovations ε_t that are (1) iid, (2) GARCH(1,1) with $\alpha = 0.1, \beta = 0.8$, (3) GARCH(1,1) with $\alpha = 0.5, \beta = 0.2$, (4) Stochastic Volatility with $\alpha_1 = 0.5$. The time series length was 1,000 while the replication was 1,000 as well. The numbers (in %) are the percentages which the order, row, identified by the method, column, for the simulated series. Each column sums to 100%.

	Innovation (i)		Innovation (ii)		Innovation (iii)		Innovation (iv)	
Models	original	modified	original	modified	original	modified	original	modified
white noise	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
MA(1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
MA(2)	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%	0.00%	0.00%
MA(3)	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%	0.00%	0.00%
AR(1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%	0.00%	0.00%
ARMA(1,1)	81.40%	82.40%	76.90%	82.70%	66.20%	87.60%	70.10%	84.30%
ARMA(1,2)	9.90%	8.50%	11.50%	7.60%	16.30%	4.40%	10.80%	7.40%
ARMA(1,3)	4.10%	3.90%	5.30%	4.50%	7.50%	3.00%	6.10%	3.70%
AR(2)	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%
ARMA(2,1)	0.20%	0.40%	0.40%	0.10%	0.80%	0.20%	0.40%	0.00%
ARMA(2,2)	2.70%	3.10%	3.60%	3.30%	3.50%	2.10%	7.10%	3.00%
ARMA(2,3)	0.80%	0.30%	0.80%	0.40%	0.70%	0.40%	1.50%	0.50%
AR(3)	0.00%	0.00%	0.00%	0.10%	0.10%	0.20%	0.10%	0.00%
ARMA(3,1)	0.00%	0.00%	0.10%	0.00%	0.20%	0.10%	0.00%	0.00%
ARMA(2,2)	0.10%	0.20%	0.10%	0.20%	0.60%	0.20%	0.20%	0.10%
ARMA(3,3)	0.50%	0.90%	0.70%	0.50%	1.90%	0.50%	2.30%	0.50%
Others	0.30%	0.30%	0.60%	0.60%	2.20%	0.50%	1.40%	0.50%

EACF is significant but not as strong as the stronger dependent GARCH innovation case (iii). For the case of innovation following the Stochastic Volatility model, the effect is quite similar to that of GARCH, hence the modified EACF procedure works for this type of innovations as well.

3.1.4. Autoregressive integrated model

In this subsection, we study the ADF test for ARI process with dependent but uncorrelated innovations. As we saw in the previous section, the ADF test continues to work in this case and we illustrate the conclusion in a simulation.

We used an $ARIMA(1, 1, 0)$ model as the mean equation and GARCH(1,1) for the innovations. Specifically,

$$\begin{cases} \Delta X_t = \phi \Delta X_{t-1} + \varepsilon_t, \\ \quad \varepsilon_t = \sqrt{g_t}\omega_t, \quad g_t = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta g_{t-1}. \end{cases}$$

Simulation was conducted with the parameter settings$\alpha = 0.1$, $\beta = 0.8$ and $\alpha = 0.5$, $\beta = 0.2$. The results of 5% level test for 10,000 replications are reported

Table 8. EACF procedure comparison by 2 × 2 contingency table and chi square test.

Innovation (i)		original	modified		
	ARMA(1,1)	814	824	chi-square	0.337293
	non ARMA(1,1)	186	176	p-value	0.561396
Innovation (ii)		original	modified		
	ARMA(1,1)	769	827	chi-square	10.4345
	non ARMA(1,1)	231	173	p-value	0.001237
Innovation (iii)	original	modified			
	ARMA(1,1)	662	876	chi-square	128.9019
	non ARMA(1,1)	338	124	p-value	7.13E-30
Innovation (iv)	original	modified			
	ARMA(1,1)	701	843	chi-square	57.27888
	non ARMA(1,1)	299	157	p-value	3.78E-14

Table 9. Results of ADF Test: replication = 10^4; time series length = 10^4.

Model	ϕ	Innovations	Percentage of rejection
ARIMA(1,0,0)	0.95	*i.i.d.*	100.00%
ARIMA(1,1,0)	0.95	*i.i.d.*	5.09%
ARIMA(1,1,0)	0.95	$\alpha = 0.1$, $\beta = 0.8$	4.80%
ARIMA(1,1,0)	0.95	$\alpha = 0.5$, $\beta = 0.2$	5.12%
ARIMA(1,0,0)	0.8	*i.i.d.*	100.00%
ARIMA(1,1,0)	0.8	*i.i.d.*	4.92%
ARIMA(1,1,0)	0.8	$\alpha = 0.1$, $\beta = 0.8$	5.06%
ARIMA(1,1,0)	0.8	$\alpha = 0.5$, $\beta = 0.2$	4.87%
ARIMA(1,0,0)	0.5	*i.i.d.*	100.00%
ARIMA(1,1,0)	0.5	*i.i.d.*	5.03%
ARIMA(1,1,0)	0.5	$\alpha = 0.1$, $\beta = 0.8$	4.73%
ARIMA(1,1,0)	0.5	$\alpha = 0.5$, $\beta = 0.2$	4.90%

in Table 9. It also include the results for an $ARIMA(1, 0, 0)$ setting to check the power of the test.

With a very large number of replications, we confirmed that the ADF test maintains roughly the correct level under the null hypothesis, regardless the type of innovations, as our theorem indicates.

3.2. Applications

Here we analyze two data sets utilizing our modified EACF procedure.

3.2.1. General motor stock price

We analyze the log-return series of General Motor company (GM) for the period from January 2, 1996 to May 8, 2006, shown in Figure 1. Order determination schemes were applied. The EACF significance tables with the orig-

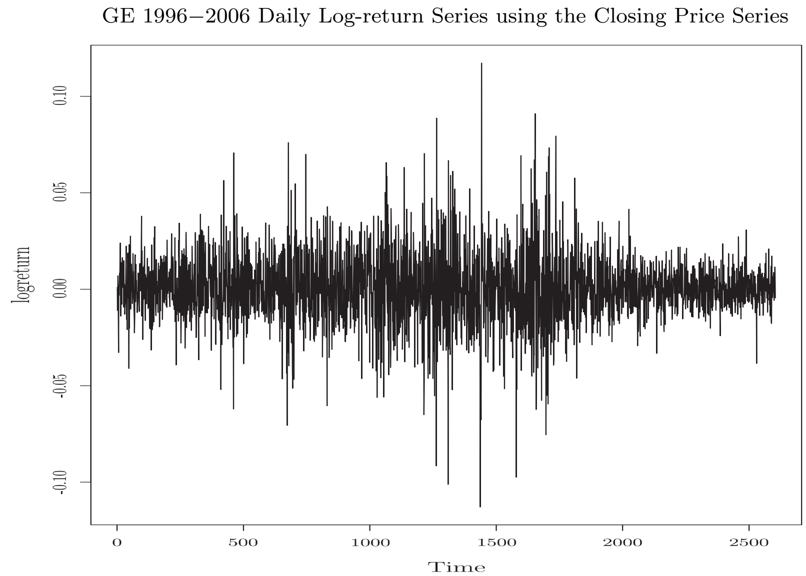

Figure 1. General Motor Stock Return Series.

inal and modified estimated variance are shown in Table 10, where 0 indicates insignificant EACF coefficient at 5% level and 2 indicates a significant one. It is seen that the original inference procedure indicates an $ARMA(2,2)$ model while the modified EACF procedure indicates a white noise model. When ARMA(2,2)+GARCH(1,1) model in the form

$$\begin{cases} X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2}, \\ \varepsilon_t = \sqrt{g_t}\omega_t, \quad g_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \beta_1 g_{t-1}, \end{cases}$$

was fitted to the data, none of the AR and MA coefficients was significant, shown in Table 11. This suggests that the modified EACF procedure provides more reliable model identification results. Although starting with a wrong model does not necessarily result in an inaccurate final model, one has to go through a more tedious and effort consuming estimation and model checking process.

3.2.2. Traffic volume in a large city

Time series analysis is widely used in traffic control applications. Large amount of traffic information including speed and volume has been collected but accurate prediction remains a challenge.

Here we analyze traffic volume data collected in a major city in China. The traffic volume (number of vehicles) passing a specific location was recorded every 3 minutes. The data set contains 8 weeks of data, with 480 observations each day. We used the first seven weeks for model identification and the last week to

Table 10. Comparison of EACF tables for GM return data.

	Original EACF Table								Modified EACF Table						
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]		[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	0	2	0	0	2	0	0	[1,]	0	0	0	0	0	0	0
[2,]	2	2	0	0	2	0	0	[2,]	2	0	0	0	0	0	0
[3,]	2	2	0	0	0	0	0	[3,]	2	2	0	0	0	0	0
[4,]	2	0	2	0	0	0	0	[4,]	2	0	2	0	0	0	0
[5,]	2	0	2	2	0	0	0	[5,]	2	0	2	2	0	0	0
[6,]	2	2	2	2	2	0	0	[6,]	2	2	2	2	2	0	0

Table 11. The fitting results of GM data set using ARMA(2,2) +GARCH(1,1) model.

	Estimate	Std. Error	t value	$Pr(>\|t\|)$	Significance
μ	2.37E-04	1.43E-04	1.658	0.0973	.
ϕ_1	5.55E-01	3.69E-01	1.506	0.1321	
ϕ_2	9.34E-02	3.34E-01	0.280	0.7794	
ψ_1	-5.57E-01	3.67E-01	-1.520	0.1284	
ψ_2	-1.42E-01	3.37E-01	-0.419	0.6749	
α_0	5.49E-07	3.14E-07	1.747	0.0806	.
α_1	3.75E-02	6.41E-03	5.847	5.02E-09	***
β_1	9.62E-01	6.45E-03	149.1	<2.00E-16	***

Table 12. Comparison of EACF Tables for Traffic Data Set.

	Original EACF Table									Modified EACF Table							
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]		[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	2	2	2	2	2	2	2	2	[1,]	2	2	2	2	2	2	2	2
[2,]	2	0	0	2	0	0	0	0	[2,]	2	0	0	0	0	0	0	0
[3,]	2	2	0	2	0	0	0	0	[3,]	2	2	0	0	0	0	0	0
[4,]	2	2	2	0	0	0	0	0	[4,]	2	2	2	0	0	0	0	0
[5,]	2	2	2	2	0	0	0	0	[5,]	2	2	2	2	0	0	0	0
‘[6,]	2	2	2	2	2	0	0	0	[6,]	2	2	2	2	2	0	0	0

test the prediction results. Figure 2 shows four of the eight weeks of the time series.

Clearly the time series possesses strong daily and weekly seasonality. To remove the seasonality and to stabilize the variance, for each observation we subtracted from it the average of the volumes observed at the same time in the seven weeks, then divided it by the standard deviation of these seven volumes. Again, we used the original and modified EACF procedures on the resulting series. Table 12 shows the results.

According to the EACF table, the modified EACF procedure identifies an ARMA(1,1) while the original EACF procedure identifies an ARMA(1,4) model, if the significant entries are treated strictly. The estimation results for both models are shown in Table 13.

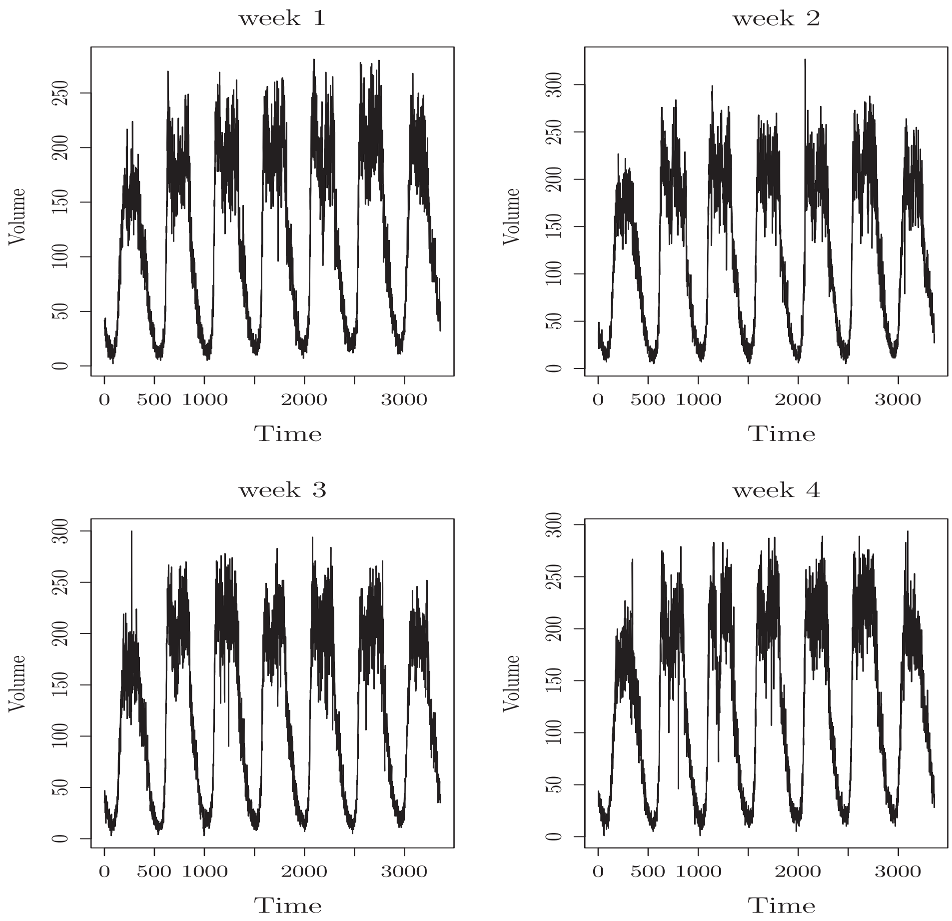

Figure 2. Big City Traffic Volume Data Set.

Table 13. Estimation results for the traffic data.

	ARMA(1,4)+GARCH(1,1)				ARMA(1,1)+GARCH(1,1)			
	Est	Std. Err	t value	p value	Est	Std. Err	t value	p value
ar1	9.90e-01	1.37e-03	722.7	$<$ 2e-16	9.85e-01	1.77e-03	557.0	$<$ 2e-16
ma1	-8.81e-01	6.98e-03	-126.3	$<$ 2e-16	-9.01e-01	4.98e-03	-181.0	$<$ 2e-16
ma2	-2.58e-03	8.90e-03	-0.290	0.772				
ma3	-1.23e-02	9.00e-03	-1.363	0.173				
ma4	-2.79e-02	6.73e-03	-4.150	3.33e-05				
ω	5.44e-02	8.78e-03	6.198	5.73e-10	5.72e-02	9.21e-03	6.209	5.35e-10
α_1	5.06e-02	5.01e-03	10.10	$<$ 2e-16	5.38e-02	5.19e-03	10.36	$<$ 2e-16
β_1	8.73e-01	1.58e-02	55.11	$<$ 2e-16	8.66e-01	1.66e-02	52.27	$<$ 2e-16

The estimation shows that the last of the three extra MA terms in the ARMA(1,4)+GARCH(1,1) model is actually significant. The two models have

Table 14. Sum of Squares of Prediction Errors: The numbers listed below are based on the summation of $n = 3,360$ prediction error squares obtained from the eighth weeks transformed series.

Sum of Squares of Prediction Error				
Model	Prediction Steps d			
	1	2	5	20
ARMA(1,4)+GARCH(1,1)	29.75342	29.34402	31.89418	32.54956
ARMA(1,1)+GARCH(1,1)	29.43388	29.32293	31.66965	32.41064

very comparable BIC values (2.485 and 2.487). To further compare the two models, we obtained the multi-step prediction error using observation of the next week. Table 14 shows the Sum of Squares of Prediction Errors, defined as $\sum_t (\hat{x}_t(d) - x_{t+d})^2$, where both the predictions and true observations are based on the transformed series. Here the d-step ahead prediction $\hat{x}_t(d)$, is a prediction of x_{t+d} made at time t with observations up to x_t, and the model parameters were estimated using the first seven weeks of data. It is seen that the simpler model identified by the modified EACF procedure performs slightly better than the more complex one.

Acknowledgement

The authors are grateful to the referees and an associate editor for many helpful comments. Rong Chen's research is supported by NSF grants DMS 0905763, DMS 0915139 and DMS 0800183.

Appendix: Proof of the Theorems

Proof of Theorem 1. Given the conditions C.1 and C.2, Theorem 3 of Wu and Min (2005) implies that, for a given h,

$$\sqrt{n}(\hat{\gamma}(0) - \gamma(0), \hat{\gamma}(1) - \gamma(1), \ldots, \hat{\gamma}(h) - \gamma(h))' \xrightarrow{dist.} N\Big(\underset{(h+1)\times 1}{\mathbf{0}}, \underset{(h+1)\times(h+1)}{\mathbf{\Pi}}\Big),$$

where $\mathbf{\Pi}$ is the covariance matrix of random vector $\sum_{t=-\infty}^{\infty} \mathcal{P}_1(x_t x_t, x_{t-1}x_t, \ldots, x_{t-h}x_t)'$.

With $H((x_0, \ldots, x_h)') = (x_1/x_0, \ldots, x_h/x_0)'$, and a Delta method argument, we have that the sample autocorrelation functions $\hat{\rho}(1), \ldots, \hat{\rho}(h)$ satisfy

$$\sqrt{n}(\hat{\rho}(1) - \rho(1), \ldots, \hat{\rho}(h) - \rho(h))' \xrightarrow{dist.} N\Big(\underset{h\times 1}{\mathbf{0}}, \underset{h\times h}{\mathbf{\Sigma}}\Big),$$

with $\mathbf{\Sigma} = D\mathbf{\Pi}D'$. Here $D = [-\boldsymbol{\rho}_h, \mathbf{I}_h]/\gamma(0)$, where $\boldsymbol{\rho}_h = (\rho_1, \ldots, \rho_h)'$ and $\mathbf{I}_h$ is the identity matrix of $h \times h$. This completes the proof for the first part of

Theorem 1. For $h > q$, by the obvious fact $\rho(h) = 0$ for $MA(q)$ process and the above expression for $\mathbf{\Sigma}$, we have that

$$\sqrt{n}\hat{\rho}(h) \xrightarrow{dist.} N(0, \delta_h^2),$$

where $\delta_h^2 = \|\xi\|^2/\gamma^2(0)$ and $\xi = \sum_{t=-\infty}^{\infty} \mathcal{P}_1 x_{t-h} x_t$. Since $\{x_t\}$ is $MA(q)$, $x_t = \sum_{i=1}^{q} \psi_i x_{t-i} + \varepsilon_t$, we have $\xi = \varepsilon_1 \sum_{i=0}^{q} \psi_i x_{1-h+i}$ if $h > q$, and therefore

$$\begin{aligned}
\|\xi\|^2 &= \mathbb{E}(\varepsilon_1 \sum_{i=0}^{q} \psi_i x_{1-h+i})^2 \\
&= \mathbb{E}[\varepsilon_1^2 (\sum_{i=0}^{\infty} \kappa_i \varepsilon_{-i})^2] \quad (\text{MA representation of } \{x_t\}) \\
&= \sum_{i,j=0}^{\infty} \kappa_i \kappa_j \mathbb{E}(\varepsilon_1^2 \varepsilon_{-i} \varepsilon_{-j}) \\
&= \sum_{i=0}^{\infty} \kappa_i^2 \mathbb{E}(\varepsilon_1^2 \varepsilon_{-i}^2) \qquad (i \neq j \text{ terms vanish by assumption}) \\
&= \sum_{i=0}^{\infty} \kappa_i^2 \mathbb{E}(\varepsilon_1^2) \mathbb{E}(\varepsilon_{-i}^2) + \sum_{i=0}^{\infty} \kappa_i^2 Cov(\varepsilon_1^2, \varepsilon_{-i}^2).
\end{aligned}$$

The first term here is

$$\begin{aligned}
\mathbb{E}(\varepsilon_1^2) \sum_{i=0}^{\infty} \kappa_i^2 \mathbb{E}(\varepsilon_{-i}^2) &= \mathbb{E}(\varepsilon_1^2) \mathbb{E}(\sum_{i=0}^{\infty} \kappa_i^2 \varepsilon_{-i}^2) = \mathbb{E}(\varepsilon_1^2) \mathbb{E}(\sum_{i=0}^{\infty} \kappa_i \varepsilon_{-i})^2 \\
&= \mathbb{E}(\varepsilon_1^2) \mathbb{E}(\sum_{i=0}^{q} \psi_i x_{1-h+i})^2 \\
&= \sigma^2 \big(\sum_{i=0}^{q} \psi_i^2 \gamma(0) + 2 \sum_{k=1}^{q} \sum_{i=0}^{q} \psi_i \psi_{i+k} \gamma(k) \big) \\
&= \gamma^2(0) + 2 \sum_{k=1}^{q} \gamma^2(k).
\end{aligned}$$

The last equality is due to the fact that $\sigma^2 \big(\sum_{i=0}^{q} \psi_i^2 \big) = \gamma(0)$ and $\sigma^2 \big(\sum_{i=0}^{q} \psi_i \psi_{i+k} \big) = \gamma(k)$. Thus we have

$$\delta_h^2 - (1 + 2 \sum_{k=1}^{q} \rho^2(k)) = \frac{\|\xi\|^2}{\gamma^2(0)} - (1 + 2 \sum_{k=1}^{q} \rho^2(k)) = \sum_{i=0}^{\infty} \kappa_i^2 Cov(\varepsilon_1^2, \frac{\varepsilon_{-i}^2}{\gamma^2(0)}).$$

Hence $\delta_h^2 \geq 1 + 2 \sum_{k=1}^{q} \rho^2(k)$ if $\sum_{i=0}^{\infty} \kappa_i^2 Cov(\varepsilon_1^2, \varepsilon_{-i}^2)/\gamma^2(0)$ is nonnegative. This is true if all ACF of the ε_t^2 series are non-negative. If all ACF are strictly positive, then $\delta_h^2 > 1+2\sum_{k=1}^q \rho^2(k)$. In addition, when $\mathbb{E}(\varepsilon_n^2|\mathcal{F}_{n-1}) = \sigma^2$, we have equality, which is the same as the *i.i.d.* case.

To estimate the variance of $\hat{\gamma}(h)$, $h > q$, note that

$$\begin{aligned}
\mathbb{V}ar(\hat{\gamma}(h)) &= \mathbb{V}ar(\frac{1}{n}\sum_{t=1}^{n-h} x_t x_{t+h}) = \frac{1}{n^2}\mathbb{V}ar(\sum_{t=1}^{n-h} x_t x_{t+h}) \\
&= \frac{1}{n^2}\Big(\sum_{t=1}^{n-h}\mathbb{V}ar(x_t x_{t+h}) + \sum_{t=1}^{n-h}\sum_{s=1,s\neq t}^{n-h}\mathbb{C}ov(x_t x_{t+h}, x_s x_{s+h})\Big) \\
&\overset{MA(q)}{=} \frac{1}{n^2}\Big((n-h)\mathbb{V}ar(x_0 x_h) + 2\sum_{t=1}^{q}(n-t)\mathbb{C}ov(x_0 x_h, x_t x_{t+h})\Big) \\
&\overset{n-h\sim n\to\infty}{\longrightarrow} \frac{1}{n}\Big(\mathbb{V}ar(x_0 x_h) + 2\sum_{t=1}^{q}\mathbb{C}ov(x_0 x_h, x_t x_{t+h})\Big) \\
&\overset{Symmetry}{=} \frac{1}{n}\sum_{|d|\le q}\mathbb{C}ov(x_0 x_h, x_d x_{d+h}) \\
&= \frac{1}{n}\sum_{|d|\le q}\Big(\mathbb{E}(x_0 x_h x_d x_{d+h}) - \mathbb{E}(x_0 x_h)\mathbb{E}(x_d x_{d+h})\Big) \\
&\overset{E(x_0 x_h)=0}{=} \frac{1}{n}\sum_{|d|\le q}\mathbb{E}(x_0 x_h x_d x_{d+h}) \triangleq \frac{1}{n}\sum_{|d|\le q}\sigma(d,h). \qquad \text{(A.1)}
\end{aligned}$$

Recalling (2.3), we have $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$. By Theorem 3 of Wu and Min (2005),

$$\begin{bmatrix}\hat{\gamma}(0)\\ \hat{\gamma}(h)\end{bmatrix} \overset{dist.}{\longrightarrow} N\Big(\begin{bmatrix}\gamma(0)\\ \gamma(h)\end{bmatrix}, n^{-1}V\Big). \qquad \text{(A.2)}$$

By the Delta method (see Casella and Berger (2002)), we have for $h > q$,

$$\begin{aligned}
\mathbb{V}ar(\hat{\rho}(h)) &= \mathbb{V}ar(\frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}) \approx \Big(-\frac{\gamma(h)}{\gamma^2(0)}, \frac{1}{\gamma(0)}\Big)[n^{-1}V]\Big(-\frac{\gamma(h)}{\gamma^2(0)}, \frac{1}{\gamma(0)}\Big)^T \\
&\overset{\gamma(h)=0}{=} \frac{1}{\gamma^2(0)}n^{-1}V_{22} \approx \frac{\mathbb{V}ar(\hat{\gamma}(h))}{\gamma^2(0)} \\
&\overset{\text{by (A.1)}}{=} \frac{1}{n\gamma^2(0)}\big(\sigma(0,h) + 2\sigma(1,h) + \cdots + 2\sigma(q,h)\big).
\end{aligned}$$

Proof of Theorem 2. We adopt the notation $\mathbf{X_k} \in \mathbb{R}^{n\times k}$ and $\varepsilon \in \mathbb{R}^{n\times 1}$ for any given $k \in (p, p+s]$,

$$\mathbf{X_k} = \begin{bmatrix} x_0 & x_{-1} & \cdots & x_{1-k} \\ x_1 & x_0 & \cdots & x_{2-k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_{n-k} \end{bmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

where $\{x_t | t = 1, \ldots, n\}$ is the observed sequence and in the definition for $\mathbf{X_k}$ we assume that $x_l = 0$ for $l \le 0$. We first argue the asymptotic joint normal distribution for $(\mathbf{X}'_{\mathbf{p+s}}\varepsilon)/\sqrt{n}$. In fact, under C.1 and C.2 we can apply Theorem 3 of Wu and Min (2005) respectively, to the new sequences $\{x_{t-j} + \varepsilon_{t+1}\}$ and $\{x_{t-j} - \varepsilon_{t+1}\}$ for any given j. Since $x_{t-j}\varepsilon_{t+1} = [(x_{t-j}+\varepsilon_{t+1})^2 - (x_{t-j}-\varepsilon_{t+1})^2]/4$, we have $\sum_{t=1}^{\infty} \|\mathcal{P}_1(x_{t-j}\varepsilon_{t+1})\| < \infty$, which leads to the asymptotic normality of $\sum_{t=1}^{n} x_{t-j}\varepsilon_{t+1}/\sqrt{n}$. Through the Cramér-Wold device, we obtain asymptotic joint normality of $\{\sum_{t=1}^{n} x_{t-j}\varepsilon_{t+1}/\sqrt{n} \,|\, j = 1, \ldots, p+s\}$, i.e. of $(\mathbf{X}'_{\mathbf{p+s}}\varepsilon)/\sqrt{n}$. In particular, the mean vector of the asymptotic joint normal distribution is a vector of zeros since $E(x_{t-j}\varepsilon_{t+1}) = 0$.

On the other hand, the lag-k PACF $\hat{\phi}_{k,k}$ is obtained from the OLS regression of $x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \cdots + \phi_k x_{t-k}$ where the true values of $\phi_{p+1}, \ldots, \phi_k$ are zero. For each $k \in (p, p+s]$, $\sqrt{n}\hat{\phi}_{k,k} = [(\mathbf{X'_k X_k}/n)^{-1}]_{(kth\,\mathrm{row})}(\mathbf{X'_k}\varepsilon)/\sqrt{n}$ is a linear transformation of the vector $\mathbf{X'_{p+s}}\varepsilon/\sqrt{n}$. Hence the vector $\sqrt{n}(\hat{\phi}_{p+1,p+1}, \cdots, \hat{\phi}_{p+s,p+s})'$ also approaches a linear transformation of $\mathbf{X'_{p+s}}\varepsilon/\sqrt{n}$. Since $(\mathbf{X}'_{\mathbf{p+s}}\varepsilon)/\sqrt{n}$ is asymptotically joint normal with zero mean, it follows

$$\sqrt{n}(\hat{\phi}_{p+1,p+1}, \ldots, \hat{\phi}_{p+s,p+s})' \xrightarrow{dist.} N\Big(\underset{s\times 1}{\mathbf{0}}, \underset{s\times s}{\mathbf{\Xi}}\Big).$$

This completes the proof for the first part of Theorem 2. To show (2.10), for simplicity of notations, we write $k = p+l$ hereafter. $\mathbf{\Xi}_{(l,l)}$ is related to the variance of asymptotic distribution of lag-k sample PACF $\hat{\phi}_{k,k}$. We first notice by Theorem 3 of Wu and Min (2005), $(\mathbf{X}'_{\mathbf{k}}\varepsilon)/\sqrt{n}$ has an asymptotic normal distribution with covariance matrix $\Omega_k = \mathbb{E}(\xi\xi')$ where $\xi = \sum_{t=-\infty}^{\infty}(\mathcal{P}_1(\varepsilon_t x_{t-1}), \ldots, \mathcal{P}_1(\varepsilon_t x_{t-k}))' = (\varepsilon_1 x_0, \varepsilon_1 x_{-1}, \ldots, \varepsilon_1 x_{1-k})'$. Since $\sqrt{n}\hat{\phi}_{k,k} = [(\mathbf{X'_k X_k}/n)^{-1}]_{(kth\,\mathrm{row})}\,(\mathbf{X'_k}\varepsilon)/\sqrt{n}$, we have $\mathbf{\Xi}_{(l,l)} = [\Gamma_k^{-1}\Omega_k\Gamma_k^{-1}]_{(k,k)}$, the (k,k) entry of matrix $[\Gamma_k^{-1}\Omega_k\Gamma_k^{-1}]$ where $\Gamma_k := [\gamma(i-j)]_{1\le i,j\le k}$. We will show that $\Omega_k - \sigma^2\Gamma_k$ is a positive semidefinite matrix. To this end, consider that for any non-random vector $C = (c_0, c_{-1}, \ldots, c_{1-k})' \in \mathbb{R}^k$ with a positive L_2 norm,

$$\begin{aligned} C'\Omega_k C &= C'[\mathbb{C}ov(\varepsilon_1 x_0, \varepsilon_1 x_{-1}, \ldots, \varepsilon_1 x_{1-k})']C \\ &= \mathbb{E}[\varepsilon_1^2(c_0 x_0 + c_{-1}x_{-1} + \cdots + c_{1-k}x_{1-k})^2] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[\varepsilon_1^2(\sum_{i=0}^{\infty}\kappa_i\varepsilon_{-i})^2] \quad \text{(MA representation of } \{x_t\}\text{)} \\
&= \sum_{i,j=0}^{\infty}\kappa_i\kappa_j\mathbb{E}(\varepsilon_1^2\varepsilon_{-i}\varepsilon_{-j}) \\
&= \sum_{i=0}^{\infty}\kappa_i^2\mathbb{E}(\varepsilon_1^2\varepsilon_{-i}^2) \quad (i \neq j \text{ terms vanish by assumption}) \\
&\geq \sum_{i=0}^{\infty}\kappa_i^2\mathbb{E}(\varepsilon_1^2)\mathbb{E}(\varepsilon_{-i}^2) \quad \text{(ACF of } \{\varepsilon_t^2\} \text{ are nonnegative by assumption)} \\
&= \sigma^2\sum_{i=0}^{\infty}\kappa_i^2\mathbb{E}(\varepsilon_{-i}^2) = \sigma^2\mathbb{E}[(c_0x_0 + c_{-1}x_{-1} + \cdots + c_{1-k}x_{1-k})^2] \\
&= \sigma^2 C'\Gamma_k C > 0.
\end{aligned}
$$

So $\Gamma_k^{-1}\Omega_k\Gamma_k^{-1} \geq \Gamma_k^{-1}\sigma^2\Gamma_k\Gamma_k^{-1} = \sigma^2\Gamma_k^{-1}$. It is known that for an $AR(p)$ process the (k,k) entry of Γ_k^{-1} is $1/\sigma^2$ for $k > p$. The result follows in view of $[\Gamma_k^{-1}\Omega_k\Gamma_k^{-1}]_{(k,k)} \geq \sigma^2[\Gamma_k^{-1}]_{(k,k)} = 1$. If all ACFs are positive, then the inequality in above derivations holds strictly since $\Omega_k > \sigma^2\Gamma_k$.

Let $S_n \triangleq \sum_{t=1}^n \varepsilon_t$, then $\|S_n^2\|^2 = \mathbb{E}(\sum_{t=1}^n \varepsilon_t)^2 = \mathbb{E}(\sum_{t=1}^n \varepsilon_t^2)$. Take $V_n^2 \triangleq \widehat{\|S_n^2\|^2} = \sum_{t=1}^n \varepsilon_t^2$. For $k = 0, \ldots, n$, set

$$
W_n(t) = \begin{cases} \frac{S_k}{V_n}, & t = \frac{k}{n}\,; \\ \frac{S_k}{V_n} + (nt-k)\frac{\varepsilon_{k+1}}{V_n}, & \frac{k}{n} \leq t \leq \frac{k+1}{n}\,. \end{cases}
$$

Lemma A.1. Suppose $\{X_t\}$ follows (2.11), with θ_i the coefficient at lag i of the MA infinity representation of $\{X_t\}$, $\Theta_n = \sum_{i=0}^n \theta_i$, $B_n^2 = \sum_{i=0}^{n-1}\Theta_i^2$. If the error series $\{\varepsilon_t\}$ satisfies C.3, $\sum_{i=1}^{\infty}(\Theta_{n+i} - \Theta_i)^2 = o(B_n^2)$, and $B_n \to \infty$, then $W_n \xrightarrow{dist.} W$.

Remark 6. The condition $\sum_{i=1}^{\infty}(\Theta_{n+i} - \Theta_i)^2 = o(B_n^2)$ is related to the mean equation's coefficients and is automatically satisfied for stationary and invertible time series.

Proof of Lemma 3. See Theorem 1 of Wu and Min (2005).

Proof of Theorem 3. Recall that $S_n = \sum_{t=1}^n \varepsilon_t$ and $V_n^2 = \widehat{\|S_n^2\|^2} = \sum_{t=1}^n \varepsilon_t^2$. Also recall that $\hat{\rho}_n$, the least square estimator of ρ, has the form

$$
\widehat{\rho}_n = \frac{\sum_{t=p+1}^n X_{t-1}(X_t - \sum_{k=1}^{p-1}\hat{a}_k\Delta X_{t-k})}{\sum_{t=p+1}^n X_{t-1}^2}.
$$

Under the null hypothesis that $\rho = 1$, since no MA part is involved in X_t, from the multiple regression property $\hat{a}_k \to a_k$, we immediately have

$$\begin{aligned}\widehat{T}_n = n(\widehat{\rho}_n - 1) &= n\frac{\sum_{t=p+1}^n X_{t-1}(X_t - X_{t-1} - \sum_{k=1}^{p-1}\hat{a}_k \Delta X_{t-k})}{\sum_{t=p+1}^n X_{t-1}^2}\\ &= \frac{\frac{1}{n}\sum_{t=p+1}^n X_{t-1}\varepsilon_t}{\frac{1}{n^2}\sum_{t=p+1}^n X_{t-1}^2} + o_p(1).\end{aligned}$$

Take $c = (1 - \sum_{k=1}^{p-1} a_k)^{-1}$, and use the same polynomial decomposition of Theorem 10.1.2 of Fuller (1995). Although the conditions on innovations are different, they do not impact the argument for this specific part. Using the same argument tat still holds for the our assumptions, we get

$$\frac{1}{n^2}\Big[\sum_{t=p+1}^n X_{t-1}^2 - c^2 \sum_{t=p+1}^n S_{t-1}^2\Big] = o_p(1);$$

$$\frac{1}{n}\Big[\sum_{t=p+1}^n X_{t-1}\varepsilon_t - c\sum_{t=p+1}^n S_{t-1}\varepsilon_t\Big] = o_p(1).$$

Hence,

$$\begin{aligned}\hat{T}_n &= = \frac{\frac{1}{n}\sum_{t=p+1}^n X_{t-1}\varepsilon_t}{\frac{1}{n^2}\sum_{t=p+1}^n X_{t-1}^2} + o_p(1) = \frac{nc\sum_{t=p+1}^n S_{t-1}\varepsilon_t}{c^2\sum_{t=p+1}^n S_{t-1}^2} + o_p(1)\\ &= \frac{n}{c}\frac{\frac{1}{2}(S_n^2 - V_n^2)}{\sum_{t=p+1}^n S_{t-1}^2} + o_p(1) = \frac{1}{c}\frac{\frac{1}{2}((S_n/V_n)^2 - 1)}{\frac{1}{n}\sum_{t=p+1}^n (S_{t-1}/V_n)^2} + o_p(1)\\ &\xrightarrow{P} (1 - \sum_{k=1}^{p-1} a_k)\frac{\frac{1}{2}(W_n^2(1) - 1)}{\frac{1}{n}\sum_{t=p+1}^n W_n^2(\frac{t-1}{n})}\\ &\xrightarrow{dist.} (1 - \sum_{k=1}^{p-1} a_k)\frac{\int_0^1 W(t)dW(t)}{\int_0^1 W^2(t)dt} \quad \text{(by Lemma A.1).}\end{aligned}$$

References

Akaike, H. (1974). A new look at the statistical model identi cation, *IEEE Trans. Automat. Control* **19**, 716-723.

Beguin, J. M. , Gourieroux, C. and Monfort, A. (1980). Identi cation of a mixed autoregressive moving average process: The corner method. In *Time Series.* (Edied by O. D. Anderson), Amsterdam: North-Holland, 423-436.

Brockwell, P. J. and Davis, R. A. (1986). *Time Series: Theory and Methods.* Springer-Verlag.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *J. Econom.* 31, 307-327.

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

Casella, G. and Berger, R. L. (2002). *Statistical Inference.* 2nd edition. Duxbury Press.

Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367-401.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *J. Amer. Statist. Assoc.* **74**.

Fuller, W. A. (1995). *Introduction to Statistical Time Series.* 2nd edition. Wiley, New York.

Gray, H. L., Kelly, G. D. and McIntire, D. D. (1978), A new approach to ARMA modeling, *Comm. Statist.*, **B7**, 1-77.

Min, W. L. (2004). Effect of dependent innovations on time series inference. Ph.D Thesis.

Min, W. L. and Tsay, R. S. (2005). On Canonical Analysis of Multivariate Time Series. *Statist. Sinica* **15**, 303-323.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *J. Econom.* **45**, 239-265.

Said, E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika* **71**, 599-607.

Schwarz, G. E. (1978). Estimating the dimension of a model, *Ann. Statist.* **6**, 461-464.

Tsay, R. S. and Tiao, G. C. (1984). Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. *J. Amer. Statist. Assoc.* **79**.

Wu, W. B. and Min, W. L. (2005). On linear processes with dependent innovations. *Stochastic Process. Appl.* **115**, 939-958.

Yang, X. R. and Zhang, L. X. (2008). A note on self-normalized Dickey-Fuller test for unit root in autoregressive time series with GARCH errors. *Appl. Math. J. Chinese Univ.* **23**, 197-201.

Bank of America Merrill Lynch, New York City, NY 10036, USA.

E-mail: shuhao.chen@baml.com

Google, Mountain View, CA 94043, USA.

E-mail: wmin@google.com

Department of Statistics, Rutgers University, 475 Hill Center, Busch Campus, 110 Frelinghuysen Road, Piscataway, NJ 08854, USA.

E-mail: rongchen@stat.rutgers.edu

(Received September 2010; accepted August 2012)