ON ORDER IDENTIFICATION OF TIME SERIES MODELS AND ITS APPLICATIONS

BY SHUHAO CHEN

A dissertation submitted to the Graduate School—New Brunswick Rutgers, The State University of New Jersey in partial fulfillment of the requirements for the degree of Doctor of Philosophy Graduate Program in Statistics and Biostatistics

Written under the direction of Rong Chen and approved by

_________________________________

_________________________________

_________________________________

_________________________________

New Brunswick, New Jersey

May, 2011
My thesis focuses on the order identification schemes of the widely-used time series model - Autoregressive Integrated Moving-Average (ARIMA) model and the applications of the order determination methods.

The first part investigates the impact of dependent but uncorrelated innovations (errors) on the traditional autoregressive integrated moving average (ARIMA) model order determination schemes such as autocorrelation function (ACF), partial autocorrelation function (PACF), extended autocorrelation function (EACF) and unit-root test. We also propose a new order determination scheme to address those impacts and can be used to time series sequences with uncorrelated innovations.

In the second part, a unified approach for the tentative specification of both the seasonal and nonseasonal orders of general multiplicative seasonal model is proposed. This new approach has the advantages of determining the seasonal and nonseasonal orders simultaneously and automatically.
In the third part, a hierarchical model approach is presented for predicting the end-of-day stock trading volume (total daily volume). It effectively combines two sources of information: the trading volume already accumulated from the beginning of the trading day to the time of prediction, and the historical daily trading volume dynamics.
I am deeply grateful to my thesis advisor Dr. Rong Chen for his helpful guidance and invaluable brainstorm. During my second year at Rutgers, I took his course of Applied Time Series and his intriguing teaching added lots of fun to my study and encouraged me to go further in time series research. Through researching with him for the following two years, I learned not only the knowledge and analytical skills but also the strong will and indomitable spirit to solve difficult problems independently and effectively, which I think would help a lot not only in my future career but also life.

I would also like to express my thanks to Dr. Wanli Min, IBM Singapore, for his extensive support, constant encouragement and kind presence in my thesis committee. Not only does he give me hands-on guidance on the topics of part of this thesis, but he also teaches me how to learn effectively and think deeply. His devotion to the application of mathematical and statistical methods will always guide and inspire me.

I also want to express my sincere appreciation to Dr. Frank Lin, ConvergEx Group, for the internship opportunity he offered to me of applying statistics into practical problem.

Without altruistic advice, help and encouragement from above people, my thesis can never be done.

Moreover, I would like to thank my graduate director Dr. John Kolassa, the
other faculty and staff at Department of Statistics and Biostatistics of Rutgers University for supporting me in various ways and at various occasions during my study.

Last but not least, I want to thank the other members of my dissertation committee, Professor Minge Xie, Professor Lee Dicker for their helpful comments on the manuscript.

Finally, I would like to show my great gratitude to my parents, wife and friends for their continuous encouragement and supports.
Dedication

To my parents Zhimin and Guixiang; To my wife Yingqiu.
Table of Contents

Abstract ......................................................... ii
Acknowledgements .............................................. iv
Dedication ......................................................... vi

List of Tables .................................................. x
List of Figures .................................................. xiii

1. Preliminary .................................................... 1
   1.1. Autoregressive Integrated Moving Average Model ........ 1
   1.2. Order Identification Tools for ARIMA Model ............... 4
      1.2.1. Autocorrelation Function ............................. 5
      1.2.2. Partial Autocorrelation Function ....................... 6
      1.2.3. Extended Autocorrelation Function ..................... 7
      1.2.4. Unit Root Test ....................................... 11
   1.3. Generalized Autoregressive Conditional Heteroscedasticity Model 13
   1.4. Seasonal Autoregressive Integrated Moving Average Model .... 15
   1.5. Topics in this Thesis .................................... 17

2. Model Identification for Time Series with Dependent Innovations 19
   2.1. Introduction .............................................. 19
   2.2. Methodology ............................................. 21
      2.2.1. Assumptions on the Innovations ....................... 21
2.2.2. Moving Average Model ............................................ 22
2.2.3. Autoregressive Model ............................................. 26
2.2.4. Autoregressive and Moving Average Model ................. 29
2.2.5. Autoregressive Integrated Moving Average Model .......... 30
2.3. Application .............................................................. 33
  2.3.1. Simulation .......................................................... 33
  Moving Average Model ............................................... 34
  Autoregressive Model ............................................... 35
  Autoregressive Moving Average Model ............................ 35
  Autoregressive Integrated Moving Average Model ............... 42
2.3.2. Real Data ............................................................ 44
  General Motor Stock Price Data Set ............................... 44
  Big City Traffic Volume Data Set .................................... 46
2.4. Proof ................................................................. 48

3. Automatic Order Identification for Seasonal ARIMA Series . 56
  3.1. Introduction .......................................................... 56
  3.2. Methodology ........................................................ 58
  3.2.1. Separating Seasonal Impact and Getting Nonseasonal Orders 60
    Assume SAR as the seasonal model ............................. 61
    Assume SMA as the seasonal model ............................. 63
    Assume SARMA as the seasonal model ......................... 65
  3.2.2. Automatic Order Identification Strategy ..................... 66
    Seasonal Part Comparison ......................................... 67
    Non-Seasonal Part Comparison .................................... 67
    Balance between Seasonal Comparison and Non-seasonal Comparison .................. 68
4. End-of-Day Stock Trading Volume Prediction with a Two-Component Hierarchical Model

4.1. Introduction ......................................................... 79
4.2. Two Component Hierarchical Model ............................... 81
4.3. Prediction based on the two-component hierarchical models .... 84
4.4. Empirical Study ....................................................... 86

References ................................................................. 96

Vita .............................................................. 102
List of Tables

1.1. Theoretical EACF table for an ARMA(1,2) series. .......................... 8

2.1. Empirical percentile of $\hat{\rho}^2(h)/\text{Var}(\hat{\rho}(h))$. .................. 35

2.2. EACF results for a simulated MA(3) series with different type of innovations: $X_t = \varepsilon_t + 0.8\varepsilon_{t-1} - 0.8\varepsilon_{t-2} + 0.8\varepsilon_{t-3}$ where innovation $\varepsilon_t$’s types are: (1) iid; (2) GARCH(1,1) with $\alpha = 0.1, \beta = 0.8$; (3) GARCH(1,1) with $\alpha = 0.5, \beta = 0.2$. The first half are results from original EACF procedure using Bartlett formula while the second half are results from modified EACF procedure. .................. 36

2.3. EACF results for a simulated AR(3) series with different type of innovations: $X_t = 0.8X_{t-1} - 0.8X_{t-2} + 0.8X_{t-3} + \varepsilon_t$ where innovation $\varepsilon_t$’s types are: (1) iid; (2) GARCH(1,1) with $\alpha = 0.1, \beta = 0.8$; (3) GARCH(1,1) with $\alpha = 0.5, \beta = 0.2$. The first half are results from original EACF procedure using Bartlett formula while the second half are results from modified EACF procedure. .................. 37

2.4. Original EACF results for a simulated ARMA(1,1) series with different type of innovations: $X_t = 0.8X_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$ where innovation $\varepsilon_t$’s types 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 $\alpha = 0.5$. .......................................................... 39
2.5. Modified EACF results for a simulated ARMA(1,1) series with different type of innovations: $X_t = 0.8X_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$ where innovation $\varepsilon_t$'s types 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 $\alpha = 0.5$.

2.6. Results of Dickey Fuller test (1); replication = $10^5$.

2.7. Results of Dickey Fuller test (2) with different set of parameters; replication=$10^5$.

2.8. Comparison of EACF tables for GM data set.

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

2.10. Comparison of EACF tables for traffic data set.

2.11. Sum of prediction error squares: the numbers listed below are based on the summation of $n = 3360$ error squares gained from the last week’s transformed series.

3.1. Comparison of the recursive and least square estimators for SAR(1) model $(1 - 0.8B^4)(1 - 0.8B)X_t = (1 + 0.8B)\varepsilon_t$.

3.2. Comparison of the recursive and least square estimators for SAR(1) model $(1 - 0.8B^7)(1 - 0.8B + 0.8B^2)X_t = (1 + 0.8B - 0.8B^2)\varepsilon_t$.

3.3. ASNOIS Output for SAR(1):$(1 - 0.9B^7)X_t = \varepsilon_t$.

3.4. Accuracy Count for SAR(1):$(1 - 0.9B^7)X_t = \varepsilon_t$.

3.5. ASNOIS Output for SMA(1)+ARMA(1,1): $(1 - 0.9B)X_t = (1 - 0.3B^7)(1 + 0.3B)\varepsilon_t$.

3.6. Accuracy Count for SMA(1)+ARMA(1,1): $(1 - 0.9B)X_t = (1 - 0.3B^7)(1 + 0.3B)\varepsilon_t$.

3.7. ASNOIS Output for SARMA(1,1)+ARMA(1,2): $(1 - 0.5B^7)(1 - 0.9B)X_t = (1 + 0.5B^7)(1 - 0.2B + 0.3B^2)\varepsilon_t$.
3.8. Accuracy Count for SARMA(1,1)+ARMA(1,2): 
\[ (1 - 0.5B^7)(1 - 0.9B)X_t = (1 + 0.5B^7)(1 - 0.2B + 0.3B^2)\varepsilon_t. \]

4.1. Maximum likelihood estimates (standard errors) of the gamma.  87
4.2. Fitting results for candidate ARMA models.  89
4.3. Variance Comparison I.  91
4.4. Variance Comparison II.  91
4.5. Prediction results from various methods. (in millions of shares)  92
4.6. Root mean square error from various methods.  93
4.7. Prediction comparison on Dow Jones Industry Average components based on Prediction Mean Square Error. (in millions of shares)  94
## List of Figures

2.1. General motor stock price series ............................................. 45
2.2. Metropolitan traffic volume data set ........................................ 47
4.1. Apple Inc. volume series: the volumes to the left of vertical dash line are used in fitting while the right hand side volumes are used for prediction and testing purpose ............................................. 87
4.2. Intraday volume distribution of different stocks: x-axis corresponds to the 13 trading periods and y-axis represents the volume percentage $\theta$. Volumes would accumulate to 100% ............................................. 88
4.3. ACF and PACF for Apple Inc.'s daily volume series ................. 89
4.4. Comparison of different prediction methods .................................. 92
Chapter 1
Preliminary

1.1 Autoregressive Integrated Moving Average Model

In statistics, econometrics, and particularly time series analysis, an autoregressive integrated moving-average (ARIMA) model is, in theory, the most general class of models for forecasting a time series. These models are fitted to autocorrelated time series data either to better understand the data (fitting) or to predict future points in the series (forecasting).

An autoregressive integrated moving average (ARIMA) model is usually applied in cases where data show evidence of non-stationarity and an initial differencing step (corresponding to the ”integrated” part of the model) can be applied to remove the non-stationarity. Afterwards, it simplifies to autoregressive moving average (ARMA) model, which itself consists of two even simpler and more fundamental models: an autoregressive (AR) part and a moving-average (MA) part.

It should be noted that the methodology of ARIMA estimation and model selection is a classical topic covered in most textbooks on time series analysis (e.g. Brockwell and David(1986)[11]; Brockwell and David(2002)[12]; Hamilton(1994)[38]; Tsay(2005)[77]). These models are also called Box and Jenkins(1984)[8] models on the basis of these authors’ pioneering work regarding time-series forecasting techniques. Some other contributions are numerated as follows. The general
transfer function model employed by the ARIMA procedure was discussed by Box and Tiao(1975)[10]. When an ARIMA model includes other time series as input variables, the model is sometimes referred to as an ARIMAX model. Pankratz(1991)[62] refers to the ARIMAX model as dynamic regression.

It is not this thesis’s intension to duplicate here the description of already well documented methodologies, but rather to list precise definitions of terms and concepts to be referenced in this thesis.

The notation $AR(p)$ refers to the autoregressive model of order $p$. The $AR(p)$ model is defined as

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t,$$  \hspace{1cm} (1.1)

where $\phi_1, \cdots, \phi_p$ are the parameters of the model, $c$ is a constant and $\varepsilon_t$ is the error. The constant term is omitted by many authors for simplicity. The value of $p$ is called the order of the AR model.

The notation $MA(q)$ refers to the moving average model of order $q$. The $AR(p)$ model is defined as

$$X_t = c + \varepsilon_t + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i},$$  \hspace{1cm} (1.2)

where $c$ is the mean of the series, $\psi_1, \cdots, \psi_q$ are the parameters of the model and $\varepsilon_t$ is the error. The value of $q$ is called the order of the MA model.

The notation $ARMA(p,q)$ refers to the model with $p$ autoregressive terms and $q$ moving-average terms. The $ARMA(p,q)$ containing both the $AR(p)$ and the $MA(q)$ models is defined as,

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i}.$$  \hspace{1cm} (1.3)
There is an alternative way to represent the above models using back-shift operators. More specifically, $ARMA(p, q)$ model for a univariate time series $X_t$ takes the form of

$$\Phi(B)X_t = \Psi(B)\varepsilon_t, \quad (1.4)$$

where $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p$ 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 $B\varepsilon_t = \varepsilon_{t-1}$.

Assume now that the polynomial $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p = (1 - \sum_{i=1}^{p} \Phi_i B^i)$ has a unitary root of multiplicity $d$, then it can be rewritten as:

$$(1 - \sum_{i=1}^{p} \Phi_i B^i) = (1 - \sum_{i=1}^{p-d} \phi_i B^i)(1 - B)^d.$$

An $ARIMA(p, d, q)$ process expresses this polynomial factorization property, and is given by

$$(1 - \sum_{i=1}^{p} \phi_i B^i)(1 - B)^d X_t = (1 + \sum_{i=1}^{q} \psi_i B^i)\varepsilon_t \quad (1.5)$$

Thus, it can be thought as a particular case of an $ARMA(p + d, q)$ process having the auto-regressive polynomial with some roots in the unity.

The ARIMA model is generally referred to as an $ARIMA(p, d, q)$ model where $p$, $d$, and $q$ are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. For example, an $AR(1)$ model is $ARIMA(1, 0, 0)$, and a $MA(1)$ model is $ARIMA(0, 0, 1)$.

The error terms $\{\varepsilon_t\}$s are generally assumed to be independent identically-distributed ($iid$) random variables sampled from a normal distribution with zero mean: $\varepsilon_t \sim N(0, \sigma^2)$ where $\sigma^2$ is the variance. These assumptions may be weakened, but doing so will change the properties of the model. In particular, a change to
the iid assumption makes a rather fundamental difference as would be elaborated in the first topic of this thesis.

1.2 Order Identification Tools for ARIMA Model

The first step, also one of the key steps in building a time series model using ARIMA setting is the order determination step, i.e., identifying $p$, $d$ and $q$ in 1.5. In the literature, order determination schemes for time series models with identically independent distributed (iid) innovations are well studied. Box and Jenkins(1976)[8] introduced the autocorrelation function (ACF) and partial autocorrelation function (PACF). The Akaike Information Criterion(AIC) by Akaike(1974)[1] and Bayesian Information Criteria (BIC) by Schwarz(1978)[69] are two goodness of fit measures of an estimated model to facilitate the model selection, which further extended into AICC, a bias correlated version of AIC by Hurvich and Tsai (1989)[45] and Hannan Quinn Information Criteria (HQIC) by Hannan and Quinn(1979)[36]. Tsay and Tiao(1984)[74] proposed the extended autocorrelation function (EACF) for order determination of $ARMA(p,q)$ model, and its properties were discussed in [74] and [73]. On the other hand, Dickey and Fuller(1979)[22] 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 the necessity of making difference(s)(for example, set $Y_t = X_t - X_{t-1}$) and then using ACF/PACF/EACF procedure on differenced series $Y_t$ to get AR and MA orders $p$ and $q$ respectively. Other order determination schemes include R and S array approach by Gray, Kelley and McIntire(1978)[34]; Corner method by Beguin, Gourieroux and Monfort(1980)[3] and Smallest Canonical Correlation(SCAN) by Tsay and Tiao(1985)[75] and discussed in Box, Jenkins, and Reinsel(1994)[9].
Choi(1992)[17] provided comprehensive reviews and descriptions of pattern identification methods and algorithms.

The definitions and properties of those order determination schemes to be discussed in this thesis are listed as follows.

1.2.1 Autocorrelation Function

The auto-covariance function and autocorrelation function (ACF) are defined as

\[ \gamma(h) = \text{Cov}(x_t, x_{t+h}) \text{ and } \rho(h) = \text{Cor}(x_t, x_{t+h}) \] respectively. (1.6)

Let \( x_1, \ldots, x_n \) be a time series. We denote the sample auto-covariance function and sample autocorrelation function respectively as

\[ \hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} x_t x_{t+h} \text{ and } \hat{\rho}(h) = \hat{\gamma}(h) / \hat{\gamma}(0). \] (1.7)

Autocorrelation function is a commonly used tool for model identification. Specifically, autocorrelations are useful in identifying the order of a moving average sequences. The autocorrelation of a \( MA(p) \) process is zero at lag \( p + 1 \) and greater. If a \( MA \) model might be appropriate, then the sample autocorrelation plot is examined to help identify the order. One looks for the point on the plot where the autocorrelations for all higher lags are essentially zero. Placing on the plot an indication of the sampling uncertainty of the sample ACF is helpful for this purpose: this is usually constructed on the basis that the true value of the ACF, at any given positive lag, is zero.
1.2.2 Partial Autocorrelation Function

The partial autocorrelation function (PACF) is defined by

\[ \alpha(0) = 1 \text{ and } \alpha(h) = \phi_{hh}, h \geq 1, \tag{1.8} \]

where \( \phi_{hh} \) is the last component of \( \phi_h = \Gamma_h^{-1} \gamma_h \) where \( \Gamma_h = [\gamma(i-j)]_{i,j=1}^h \) and \( \gamma_h = [\gamma(1), \gamma(2), \ldots, \gamma(h)]' \).

For any set of observations \( x_1, \ldots, x_n \) with \( x_i \neq x_j \) for some \( i \) and \( j \), the sample PACF \( \hat{\alpha}(h) \) is given by

\[ \hat{\alpha}(0) = 1 \text{ and } \hat{\alpha}(h) = \hat{\phi}_{hh}, \quad \forall h \geq 1, \tag{1.9} \]

where \( \hat{\phi}_{hh} \) is the last component of \( \hat{\phi}_h = \hat{\Gamma}_h^{-1} \hat{\gamma}_h \).

Partial autocorrelation function is another commonly used tool for model identification. Specifically, partial autocorrelations are useful in identifying the order of an autoregressive series. The partial autocorrelation of an AR(\( p \)) process is zero at lag \( p + 1 \) and greater. If the sample autocorrelation plot indicates that an AR model may be appropriate, then the sample partial autocorrelation plot is examined to help identify the order. One looks for the point on the plot where the partial autocorrelations for all higher lags are essentially zero. Placing on the plot an indication of the sampling uncertainty of the sample PACF is helpful for this purpose: this is usually constructed on the basis that the true value of the PACF, at any given positive lag, is zero.
1.2.3 Extended Autocorrelation Function

Although autocorrelation function and partial autocorrelation function work perfectly for pure moving average series and autoregressive series respectively, however, they do not show strong identification capability for \( ARMA(p,q) \) series. Tsay and Tiao(1984)[74] proposed the extended autocorrelation function (EACF) technique which can tentatively identify the orders of a stationary or non-stationary ARMA process based on iterated least squares estimates of the autoregressive parameters. Basically, an iterative regression procedure is given to produce consistent estimates of the autoregressive parameters and based on these consistent estimates we utilize the order determination scheme such as ACF for the series subtracting the autoregressive terms.

More specifically, the order determination scheme is arranged in the following manner,

1. The complexity of the EACF algorithm and nowadays’ cheap computational cost make it possible for us to enumerate all possible orders of the autoregressive term.

2. For each candidate \( AR \) order \( p \), we first get the consistent estimates of \( \hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p \).

3. Denote \( Y_t = X_t - \hat{\phi}_1X_{t-1} - \hat{\phi}_2X_{t-2} - \cdots - \hat{\phi}_pX_{t-p} \). For each candidate autoregressive order \( p \), given the consistent estimates \( \hat{\phi}_1, \cdots, \hat{\phi}_p \), we can get \( y_t \) from \( x_t, \cdots, x_{t-p} \) and what remains should be a \( MA(q) \) model if we get the \( AR \) order \( p \) correctly.

4. By calculating the autocorrelation function of series \( y_t \), we can have the significant test results on every moving average lag \( q \).
5. Marking significant levels on different moving average lags $q$ for all candidate $AR$ orders $p$, we plot the EACF table and choose ARMA orders by identifying the upper right zero triangle. See following Table 1.1 for details.

The argument of the last step relies on the following facts,

- If we under-estimate the $AR$ order, the transferred series should not have a clean $MA$ pattern and thus lead to whole significant row for the AR term;

- If we get the $AR$ order correctly, the autocorrelation function would provide the correct $MA$ order $q$.

- If, on the other hand, we over-fit the $AR$ order by lag $n$, the transformed sequence becomes $MA(q+n)$ series correspondingly.

Table 1.1 depicts the theoretical pattern associated with an $ARMA(1,2)$ series. For each pair of AR and MA orders, we test the significance of $t$-statistics of the autocorrelation function and mark insignificant points as 0, boundary significant points as 1 and significant points as 2. The insignificant points (marked as 0) compose the upper right zero triangle and the starting point of the triangle’s coordinate is $(2,3)$ ($(p,q)$ in general), and thus we identify the model as $ARMA(1,2)$ ($ARMA(p-1,q-1)$ in general).

Table 1.1: Theoretical EACF table for an ARMA(1,2) series.

<table>
<thead>
<tr>
<th>MA</th>
<th>AR</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

2 = significant
1 = boundary significant
0 = insignificant
Since the most important usage of the EACF table is to determine the tentative ARMA orders, it is of great interest to have an effective automatic algorithm to figure out the most suitable ARMA orders from a given EACF table. Following is two alternative algorithms which will be intensive used in the following chapters.

**Algorithm 1.** Identify autoregressive (AR) and moving-average (MA) orders from a given EACF table.

Denote $EACF(i,j)$ as the significance level located at the $i^{th}$ row and $j^{th}$ column of the EACF table, $1 \leq i \leq n$, $1 \leq j \leq m$.

**Step 1:** $ARMA(i−1,j−1)$ can be candidate model if

1. $EACF(i,j) = 0$ and
2. $Real(i,j) \leq Theory(i,j)$ where

$$Real(i,j) \triangleq \sum_{s=0}^{n-i} \sum_{t=s}^{m-j} (s + i + t + j)^{-0.5} \times EACF(s + i, t + j) \quad (1.10)$$

$$Theory(i,j) \triangleq \sum_{s=0}^{n-i} \sum_{t=s}^{m-j} (s + i + t + j)^{-0.5} \times 1 \quad (1.11)$$

**Remark.** We want the upper-right triangle leading by $EACF(i,j)$ to beat at least the triangle of 1 to qualify $ARMA(i−1,j−1)$ as a candidate model.

**Step 2:** Best $ARMA$ models among those qualified candidates can be found by

Set $k = i + j$ and enumerate $k$ from 2 to $m + n$.

**If there is only one qualified candidate:** This is the final model;

**If more than one qualified candidates:** The one with lower $\frac{Real}{Theory}$ ratio wins;
If no qualified candidate: $k = k + 1$. □

The EACF table is composed of three levels of significance as a (step) function of the underlying $t$-statistic of the autocorrelation function. Actually, we lose quite a lot information in determining the ARMA orders by the significance levels instead of the $t$-statistics. Thus it should be better to determine the ARMA orders by the $t$-statistics, or $p$-value corresponding to the $t$-statistics. This leads to the following alternative algorithm.

Algorithm 2. Identify autoregressive (AR) and moving-average (MA) orders from a given EACF table.

Denote $pEACF(i, j)$ as the $p$-value for the autocorrelation function located at the $i^{th}$ row and $j^{th}$ column of the EACF table, $1 \leq i \leq n, 1 \leq j \leq m$.

Step 1: $ARMA(i - 1, j - 1)$ can be candidate model if

1. $pEACF(i, j) \geq \alpha^1$ and
2. $pReal(i, j) \leq pTheory(i, j)$ where

$$pReal(i, j) \triangleq \sum_{s=0}^{n-i} \sum_{t=0}^{m-j} \sum_{s=0}^{m-j} (s + i + t + j)^{-0.5} \times pEACF(s + i, t + j)$$  \hspace{1cm} (1.12) $$pTheory(i, j) \triangleq \sum_{s=0}^{n-i} \sum_{t=0}^{m-j} (s + i + t + j)^{-0.5} \times \psi \times \alpha$$  \hspace{1cm} (1.13)

Remark. We want the upper-right triangle leading by $pEACF(i, j)$ to beat at least the triangle of $\psi \times \alpha$ to qualify $ARMA(i - 1, j - 1)$ as a candidate model where $\psi$ is a tuning parameter.

Step 2: Best $ARMA$ models among those qualified candidates can be found by

$^1\alpha = 0.05$ for example.
Set $k = i + j$ and enumerate $k$ from 2 to $m + n$.

**If there is only one qualified candidate:** This is the final model;

**If more than one qualified candidates:** The one with lower $\frac{p_{Real}}{p_{Theory}}$ ratio wins;

**If no qualified candidate:** $k = k + 1$. □

### 1.2.4 Unit Root Test

In statistics, a unit root test examines whether a time series is non-stationary using an autoregressive model. A well-known test that is valid in large samples is the Dickey Fuller test. When a time series has a unit root, the series is non-stationary and the ordinary least squares (OLS) estimator is not normally distributed. Dickey(1976)[21], Dickey and Fuller(1979)[22] studied the limiting distribution of the OLS estimator of autoregressive models for time series with a simple unit root.

A time series can be non-stationary because of a deterministic trend or a stochastic trend. Dicker Fuller test is intended to detect the latter, however, it would give misleading inference if a deterministic trend is present but is not allowed for. The Augmented Dickey Fuller (ADF) test, which adds lagged dependent variables to the test equation, is often used instead. It is an augmented version of the Dickey Fuller test for a larger and more complicated set of time series models. The augmented Dicker Fuller statistic, used in the test, is a negative number: the more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

Further research on this topic follows: Dickey, Hasza and Fuller(1984)[23] obtained the limiting distribution for time series with seasonal unit root; Said
and Dickey(1984)[68] extended the test to autoregressive and moving average models. There are also alternative unit root tests such as the Phillips-Perron Test by Phillips and Perron(1988)[65] or ADF-GLS procedure developed by Elliot, Rothenberg and Stock(1996)[25]. Philips-Perron test is used in time series analysis to test the null hypothesis that a time series is integrated of order 1, instead of 0 as the null hypothesis in Dickey Fuller test. Various types of unit root testing have already been detailed discussed in many textbooks, for example, Hamilton(1984)[38], Fuller(1995)[29].

For simplicity, consider a special case of ARMA model (1.4) when we set $\Psi(B) = 1$,

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \varepsilon_t. \quad (1.14)$$

A classical representation of (1.14) used for testing purpose can be written in the following manner

$$X_t = \rho X_{t-1} + \sum_{k=1}^{p-1} a_k \Delta X_{t-k} + \varepsilon_t, \quad (1.15)$$

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

$$H_0 : \rho = 1 \text{ against } H_1 : |\rho| < 1, \quad (1.16)$$

and a basic test statistic is $\widehat{T}_n \overset{\Delta}{=} n(\hat{\rho} - 1)$, where $\hat{\rho}$ is the least-square estimator (LSE) of $\rho$. Dickey and Fuller(1979)[22] proved that under the null hypothesis

$$\widehat{T}_n = n(\hat{\rho} - 1) \overset{\text{dist}}{\to} \frac{1}{2} \left( 1 - \sum_{k=1}^{p-1} a_k \right) \Gamma^{-1}(T^2 - 1), \quad (1.17)$$
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 iid \(N(0, 1)\) random variables.

1.3 Generalized Autoregressive Conditional Heteroscedasticity Model

Volatility is an important factor in many aspects, for example, option trading\(^2\): it is among the most important parameters of the well known Black-Scholes option pricing equation, see Black and Scholes(1973)[6] and Merton(1973)[55] for details. Volatility is also important in risk management: it provides a simple approach to calculate value at risk of a financial position and modeling the volatility of a time series can improve the efficiency in parameter estimation and the accuracy in interval forecast.

Certain time series, for example, financial time series, always exhibits time varying volatility clustering, i.e., periods of swings followed by periods of relative calm. Whenever there is a reason to believe that, at any time point in a series, the terms will have a characterized variance/volatility as a function of time \(t\), the normal constant variance assumption of ARIMA model fails.

Engle(1982)[26] started the effort in modeling evolving volatility by assuming the variance of the current error term or innovation to be a function of the actual size of the previous time period’s error terms; often the variance is related to the squares of the previous innovations. If an autoregressive moving average model is assumed for the error variance instead of the pure autoregressive model as in ARCH model, the model becomes a generalized autoregressive conditional heteroscedasticity (GARCH) model, proposed by Bollerslev(1986)[7].

\(^2\)Here, volatility means conditional variance of the underlying asset return.
Other models include the exponential GARCH (EGARCH) model of Nelson(1991)[58]; the conditional heteroscedastic autoregressive moving average (CHARMA) model of Tsay(1987)[76]; the random coefficient autoregressive (RCA) model of Nicholls and Quinn(1982)[60]; and the stochastic volatility (SV) models of Melino and Turnbull(1990)[54], Harvey, Ruiz and Shephard(1994)[41] and Jacquier, Polson and Rossi(1994)[48].

In this thesis, we mainly focus on the ARIMA models driven by GARCH noise and the definitions of the ARCH and GARCH models are as follow.

Recall that $\varepsilon_t$ denotes the error term in equation 1.4, and these $\varepsilon_t$s are split into a stochastic piece $\epsilon_t$ and a time-dependent variance $g_t$ such that

$$\varepsilon_t = \sqrt{g_t}\epsilon_t,$$  \hspace{1cm} (1.18)

where $\epsilon_t$ is a random variable drawn from a Gaussian distribution with mean 0 and standard deviation 1, i.e., $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$.

**For ARCH($q$) Model**

The series $g_t$ is modeled by

$$g_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2.$$ \hspace{1cm} (1.19)

**For GARCH($p,q$) Model**

The series $g_t$ is modeled by

$$g_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 g_{t-1} + \cdots + \beta_p g_{t-p}$$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i g_{t-i}.$$ \hspace{1cm} (1.20)
1.4 Seasonal Autoregressive Integrated Moving Average Model

Seasonality has often been perceived as a phenomenon that generates peaks and troughs within a particular season, for example, year after year. Since the periodicity and largely repetitive pattern observed in time series data over the course of a period, it is highly predictable.

Although seasonality is a dominant of period-to-period fluctuations in many time series, some consider it uninteresting and cause a series to be ambiguous and they intend to remove the seasonal elements and focus on other components. One famous example is the treatment of the unemployment rate by United States Census Bureau where they make an official seasonal adjustment. However, when it comes to the prediction perspective, modeling the seasonal pattern itself becomes the central issue.

The analysis of seasonal time series has a long and storied history beginning with Gilbert(1854)[30], Gilbert(1856)[31], Gilbert(1865)[32]. See Hylleberg(1992)[46] for an extensive historical review. Of course, there exist many characterizations of seasonality in time series data: they range from stochastic seasonality as discussed by Box and Jenkins(1976)[8] to the unobserved component seasonality as given by Harvey(1989)[39] to deterministic seasonality vis-a-vis dummy variables in multiple regression as given by Johnston(1984)[49] to stable seasonal pattern models as given by Oliver(1987)[61]. However, the most famous stochastic seasonality model, seasonal autoregressive integrated moving average (SARIMA) model retains an important role as a forecasting benchmark.

When working with non-stationary seasonal data, both seasonal changes and
the changes between adjacent observations are important concepts. This motivated Box and Jenkins (1976)[8] to propose the seasonal autoregressive integrated moving average (SARIMA) model \(SARIMA(p, d, q) \times (P, D, Q)^s\) as follows,

\[
\phi(B)\Phi(B^s)\Delta^d\Delta^D_sX_t = \theta(B)\Theta(B^s)\epsilon_t,
\]

where \(s\) is the seasonal lag; \(\Delta^D_s = (1 - B^s)^P\), \(\Delta^d = (1 - B)^d\); \(\Phi(B^s) = 1 - \Phi_1B^s - \cdots - \Phi_PB^{Ps}\), \(\phi(B) = 1 - \phi_1B - \cdots - \phi_pB^p\), \(\Theta(B) = 1 + \Theta_1B^s + \cdots + \Theta_QB^{Qs}\) and \(\theta(B) = 1 + \theta_1B + \cdots + \theta_qB^q\) are polynomials in the back-shift operator \(B\), defined as \(BX_t = X_{t-1}\), \(B\epsilon_t = \epsilon_{t-1}\), \(B^sX_t = X_{t-s}\) and \(B^s\epsilon_t = \epsilon_{t-s}\); \(\epsilon_t\) is the innovation term. We shall require that all zeros of \(\phi(B), \Phi(B), \theta(B)\) and \(\Theta(B)\) are outside the unit circle.

Seasonal autoregressive model (SAR) is a special case of model (1.21) when we set \(\Theta(B^s) = 1\) while seasonal moving average model (SMA) is another special case when we set \(\Phi(B^s) = 1\) and \(\Delta^D_s = 1\).

More specifically, the special case of (1.21) where

\[
(1 - B)(1 - B^s)X_t = (1 - \theta_1B)(1 - \theta_sB^s)\epsilon_t
\]

with \(|\theta_1| < 1, |\theta_s| < 1\) retains an important position. This is known as the airline model because Box and Jenkins (1970)[8] found it appropriate for monthly airline passenger data. Subsequently, the model has been shown to provide robust forecasts for many observed seasonal time series, and hence it often provides a benchmark for forecast accuracy comparisons.

Note that the use of seasonal differences, as in (1.22) or through the multiplicative filter as in (1.21), makes rather strong assumptions about the stochastic properties of the time series under analysis. It has, therefore, become common practice to examine the nature of the stochastic seasonal properties of the
data via seasonal unit root tests. In particular, Hylleberg, Engle, Granger and Yoo (1990) [47] proposed a test for the null hypothesis of seasonal integration in quarterly data, which is a seasonal generalization of the Dickey Fuller test by Dickey and Fuller (1979) [22]. Their procedure has since been extended to the monthly case by Beaulieu and Miron (1993) [2] and Taylor (1998) [72], and was generalized to any periodicity, by Smith and Taylor (1999) [71].

The seasonal autoregressive integrated moving average (SARIMA) model (a.k.a., multiplicative seasonal model) is often found in textbooks, see for instance Brockwell and Davis (1986) [11] and Harvey (1993) [40]. Franses (1996) [28] fitted SARIMA models to various real macroeconomic time series.

1.5 Topics in this Thesis

Provided all the preliminaries introduced in this chapter, we have prepared necessary tools to explore the content of this thesis. The first two topics are closely related to the order identification schemes for the autoregressive integrated moving average (ARIMA) model. Topic 1 Model Identification for Time Series with Dependent Innovations suggests new tools for determining the orders when the errors (innovations) are not iid distributed. Topic 2 Automatic Order Identification for Seasonal ARIMA Series relates to a very common type of time series sequences: seasonal series. It proposes a novel method for selecting the seasonal and nonseasonal orders simultaneously and automatically. Topic 3 End-of-Day Stock Trading Volume Prediction with a Two-Component Hierarchical Model is an application of the fitting and predicting procedure of ARIMA model and it intensively uses the tools set up in the first two topics. It is associated with the trading execution industry: suppose we are at a certain time during a trading
day, using methods discussed in this topic, we can provide better prediction of the total daily stock volume given historical volume information and the present day accumulated volume information. The three topics are discussed in a detailed manner in the following chapters one by one.

For simplicity, we employ the following notations and terms throughout the content: \( \overset{\text{dist}}{\rightarrow} \) means convergence in distribution; \( \overset{P}{\rightarrow} \) means convergence in probability; \( \overset{a.s.}{\rightarrow} \) means almost sure convergence; \( \overset{\Delta}{=} \) means denotation; term errors, innovations and shocks, multiplicative seasonal model, seasonal ARIMA model and SARIMA model, EACF table and EACF matrix, order identification and order determination are used interchangeably.
Chapter 2
Model Identification for Time Series with Dependent Innovations

2.1 Introduction

The first step and also one of the key steps in building a time series model is order determination. Most of the existing order determination methods (See literature review in section 1.2.) assume that the innovation $\{\varepsilon_t\}$s, are iid and/or have constant conditional variance, i.e., $\mathbb{E}[\varepsilon_t^2|\mathcal{F}_{t-1}] = \sigma^2$, see for example Box and Jenkins(1976)[8] and Brockwell and Davis(1986)[11]. In this case, asymptotic properties of the partial sums have been extensively studied. It is hard to compile a complete list and we only mention some representatives: Davydov(1970)[19], Gorodetskii(1977)[33], Hall and Heyde(1980)[35], Phillip and Solo(1992)[66], Yokoyama(1995)[82] and Hosking(1996)[44]. However, this setting excludes many important processes such as the generalized autoregressive conditional heteroscedastic (GARCH) model by Bollerslev(1986)[7] and the stochastic volatility (SV) model by Harvey, Ruiz, and Shephard(1994)[41] and Jacquier, Polson, and Rossi(1994)[48], which are often used in time series modeling. In these models, the innovation series are not auto-correlated but auto-dependent. Romano and Thombs(1996)[67] pointed out that the traditional large sample inference on autocorrelations under assumption of iid innovations is misleading if the underlying $\{\varepsilon_t\}$s are not actually dependent.
Hence, it becomes interesting and important to extend the classical order determination schemes to accommodate auto-dependent but uncorrelated innovations.

Specifically, in this chapter we study the order determination schemes for MA, AR, ARMA and ARIMA models at 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, extended autocorrelation function (EACF) has proven to be very useful in identifying the AR and MA orders and it also works for differenced ARIMA process. We investigate how those schemes are impacted by the new type of innovations.

Several studies have been done in the pursuit of weakening the iid assumption. For instance, Diananda(1953)[20] assumed \( m \)-dependent \( \{\varepsilon_t\} \); Hannan and Heyde(1972)[37] investigated the situation when \( \{\varepsilon_t\} \)s are martingale difference sequence. Recently, Wang et al.(2002)[79] considered invariance principles for iid or martingale difference innovations, however, it seems that the proofs in the paper were not rigorous (See remark in [80]). Chung(2002)[16] and He(1996)[43] considered linear processes with martingale difference innovations having constant conditional variance, which nevertheless is a quite restrictive assumption that excludes the widely used GARCH models still. Yang and Zhang(2008)[81] discussed the effect on unit root test of one specific type of dependent but uncorrelated innovation: the GARCH(1,1) innovation. On the other hand, Min(2004)[56] further weakened the innovations by allowing a large class of nonlinear processes, which substantially relaxed iid and/or the martingale difference assumptions and investigated the effect of dependent innovations on pure AR and MA time series sequences. Our study extends these findings to broader conclusions.

The chapter is organized as follows. In section 2.2, we present a detailed study on how dependent but uncorrelated innovations affect the classical order
identification procedures. Based on these findings, we propose suitable correction methods and investigate their theoretical validities. Simulation and real examples are used to illustrate the findings in section 2.3. At last, Proofs are given in section 2.4.

2.2 Methodology

2.2.1 Assumptions on the Innovations

Recall that ARMA$(p, q)$ model for a univariate time series $X_t$ takes the form (1.4)

$$
\Phi(B)X_t = \Psi(B)\varepsilon_t,
$$

with $\Phi(B) = \phi(B)U(B)$ where $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p$, $U(B) = 1 - U_1 B - \cdots - U_d B^d$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_{p-d} B^{p-d}$ and $\Psi(B) = 1 + \psi_1 B + \cdots + \psi_q B^q$ are polynomials in the back-shift operator $B$. We shall require that all zeros of $U(B)$ are on, those of $\phi(B)$ and $\Psi(B)$ are outside the unit circle.

Let us first specify the exact assumptions of the dependent but uncorrelated innovation we investigate. Let $\{\varepsilon_n\}$ be iid random variables and $F$ be a measurable function such that innovation $\varepsilon_n = F(\varepsilon_n, \varepsilon_{n-1}, \cdots)$ is a well-defined random variable.

Assume

$$
\mathbb{E}(\varepsilon_n|\mathcal{F}_{n-1}) = 0, \quad \mathbb{E}(\varepsilon_n^2) = \sigma^2, \quad \forall n,
$$

(2.1)

where $\mathcal{F}_n \triangleq \sigma(\varepsilon_n, \varepsilon_{n-1}, \cdots)$ is the $\sigma$-field generated by sequence $\{\varepsilon_n\}$, representing the information available up to time $n$. 

The error series \( \{ \varepsilon_n \} \) is uncorrelated since \( \mathbb{E}(\varepsilon_n \varepsilon_{n-1}) = \mathbb{E}(\mathbb{E}(\varepsilon_n \varepsilon_{n-1} | F_{n-1})) = \mathbb{E}(\varepsilon_{n-1} \mathbb{E}(\varepsilon_n | F_{n-1})) = 0 \). The second part of condition (2.1) is weaker than the traditional condition \( \mathbb{E}(\varepsilon_n^2 | F_{n-1}) = \sigma^2 \). It can be seen from the fact that if \( \mathbb{E}(\varepsilon_n^2 | F_{n-1}) = \sigma^2 \), \( \mathbb{E}(\varepsilon_n^2) = \mathbb{E}(\mathbb{E}(\varepsilon_n^2 | F_0) | F_{n-1}) = \mathbb{E}(\mathbb{E}(\varepsilon_n^2 | F_{n-1}) | F_0) = \mathbb{E}(\sigma^2 | F_0) = \sigma^2 \).

Under the traditional condition \( \mathbb{E}(\varepsilon_n^2 | F_{n-1}) = \sigma^2 \), the second order correlation \( \text{Cov}(\varepsilon_n^2, \varepsilon_{n-1}^2) \) remains zero. However, our condition \( \mathbb{E}(\varepsilon_n^2) = \sigma^2 \), allows for second order correlations within \( \{ \varepsilon_n \} \)’s and thus dependence may occur.

Note that condition (2.1) incorporates those situation where there is no correlation in innovation sequence itself but in square sequence of the innovation.

Another important assumption of the innovation \( \{ \varepsilon_n \} \) relies on the following definition: \( P_i \varepsilon_i \varepsilon_j \Delta \overset{\Delta}{=} \mathbb{E}(\varepsilon_i \varepsilon_j | F_t) - \mathbb{E}(\varepsilon_i \varepsilon_j | F_{t-1}) \), specifically \( P_1 \varepsilon_i \varepsilon_j \overset{\Delta}{=} \mathbb{E}(\varepsilon_i \varepsilon_j | F_1) - \mathbb{E}(\varepsilon_i \varepsilon_j | F_0) \) and \( \| \cdot \| \) as \( L_2 \) norm. Based on the above definition, we assume our innovation \( \{ \varepsilon_n \} \) to satisfy

\[
\sum_{i,j=0}^{\infty} ||P_1 \varepsilon_i \varepsilon_j|| < \infty. \tag{2.2}
\]

**Remark.** The intuition of condition (2.2) is that the projection of the future \( \varepsilon_i \varepsilon_j \) to the space \( \mu_1 \otimes \mu_0 = \{ Z : \sum_{i,j=0}^{\infty} ||P_1 Z|| < \infty, Z \text{ is } F_1 \text{ measurable and } \mathbb{E}(Z | F_0) = 0 \} \) has a small magnitude; namely, the future states depend weakly on the current states or the current states depend weakly on the previous states.

This setup includes many interesting stochastic processes, such as \( ARCH(p) \) model, \( GARCH(p, q) \) model, \( EGARCH(p, q) \) model and stochastic volatility Model. For example, Min(2004)[56] showed that \( GARCH(p, q) \) model satisfies condition (2.2).

### 2.2.2 Moving Average Model
One of the fundamental building block of time series models is the moving average (MA) model (1.2). Its unique feature is that the autocorrelation function (1.6) of an MA(q) process cuts off at lag q, hence ACF can be used effectively to determine the order q.

The sample autocorrelation function ˆρ(h) is defined in (1.7). According to Bartlett formula (Box and Jenkins (1976)[8]), the asymptotic variance of ˆρ(h) is

\[
\text{Var}(\hat{\rho}(h)) = \frac{1}{n} \sum_{k=-\infty}^{\infty} \{ \rho^2(k + h) + \rho(k - h)\rho(k + h) + 2\rho^2(h)\rho^2(k) \\
-4\rho(h)\rho(k)\rho(k + h) \}. \tag{2.3}
\]

Due to the symmetry of the time series involved (i.e., ρ(h) = ρ(−h)), we have

\[
\text{Var}(\hat{\rho}(h)) = \frac{1}{n} \sum_{k=1}^{\infty} \{ \rho(k + h) + \rho(k - h) - 2\rho(h)\rho(k) \}. \tag{2.4}
\]

Since \{ɛ_t\}s are uncorrelated, we have \(γ(h) = \sigma^2 \sum_{i=0}^{q} ψ_iψ_{i+h}\), where ψ_j = 0 for \(j > q\). Hence \(γ(q+i) = ρ(q+i) = 0, \forall i > 0\). This is the cut-off property of the autocorrelation function of the MA(q) process.

If we further assume \(ɛ_t \sim i.i.d. N(0, \sigma^2)\), it is well known that \(\hat{γ}(h), h = 1, 2, \ldots\) has an asymptotic joint normal distribution. In particular the Bartlett formula (2.4) implies that

\[
\text{Var}(\hat{\rho}(h)) = \frac{1}{n}(1 + 2\rho^2(1) + \cdots + 2\rho^2(q)), \quad \text{if } h > q. \tag{2.5}
\]

Note that we only consider \(\text{Var}(\hat{\rho}(h))\) for \(h > q\) since this is related to identifying a MA(q) process by testing sample autocorrelation function due to the cut-off property.

However, when \{ɛ_t\}s are dependent, the validity of the above procedure needs
to be reexamined. Specifically, we investigate the following questions under the condition (2.1) and (2.2) of \( \{\varepsilon_t\} \): Whether the cut-off property still hold for MA series with dependent but uncorrelated innovations? Whether the sample covariance function \( \hat{\gamma}(h) \) still follows the asymptotic joint normal distribution? Whether the variance formula for sample correlation function \( \text{Var}(\hat{\rho}(h)) \) changes? If it changes, what is the variance formula? How to get the variance estimator \( \text{Var}(\hat{\rho}(h)) \) if variance does change?

In the following content, we state that the cut-off property (by Lemma 1) and the asymptotic joint normality of \( \hat{\gamma}(h) \) (following the same arguments in Theorem 7.2.1 of [11]) still hold, but the asymptotic variance IS different (by Theorem 1).

**Lemma 1.** Let \( \{X_t\} \) be a MA\((q)\) process with innovation \( \{\varepsilon_t\} \) following conditions (2.1) and (2.2). For the autocorrelation function \( \rho(h) \) defined in (1.6), we have

\[
\rho(q + 1) = \rho(q + 2) = \cdots = 0.
\]  

(2.6)

**Remark.** The proof is trivial since the cut-off properties is only related to correlation rather than dependence.

**Theorem 1.** Let \( \{X_t\} \) be a MA\((q)\) process with innovation \( \{\varepsilon_t\} \) following conditions (2.1) and (2.2); autocorrelation function \( \rho(h) \) is defined as (1.6) and sample autocorrelation \( \hat{\rho}(h) \) is defined as (1.7), then we have

\[
\text{Var}(\hat{\rho}(h)) \geq \frac{1}{n}(1 + 2 \sum_{k=1}^{q} \rho^2(k)).
\]  

(2.7)

**Remark.** The theorem implies that, if the dependence in \( \{\varepsilon_t\} \) is ignored, the inferences based on Bartlett formula is misleading.

Noticing the difference due to the introduction of dependent but uncorrelated
innovations, another question raises: How to estimate the \( \text{Var}(\hat{\rho}(i)) \)? Specifically, we want to estimate the asymptotic variance of the autocorrelation \( \hat{\rho}(h) \) with lag \( h \) satisfying \( h > q \).

Recall (1.7), we have \( \hat{\rho}(h) = \hat{\gamma}(h) / \hat{\gamma}(0) \) and \( \hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} x_t x_{t+h} \). By Taylor Expansion, we have \( \text{Var}(\hat{\rho}(h)) = \text{Var}(\hat{\gamma}(h) / \hat{\gamma}(0)) \approx \text{Var}(\hat{\gamma}(h)) / \hat{\gamma}^2(0) \), where \( \text{Var}(\hat{\gamma}(h)) = \text{Var}(\frac{1}{n} \sum_{t=1}^{n-h} x_t x_{t+h}) = \frac{1}{n^2} \sum_{t=1}^{n-h} \text{Cov}(x_t x_{t+h}, x_{t+h} x_{t+h}) \).

\[
\text{Var}(\hat{\gamma}(h)) = \frac{1}{n^2} \left( \sum_{t=1}^{n-h} \text{Var}(x_t x_{t+h}) + 2 \sum_{t=1}^{n-h} \sum_{s=1, s \neq t}^{n-h} \text{Cov}(x_t x_{t+h}, x_s x_{s+h}) \right) \rightarrow_{n \to \infty} \frac{1}{n} \left( \text{Var}(x_0 x_h) + 2 \sum_{t=1}^{n-h} \text{Cov}(x_0 x_h, x_t x_{t+h}) \right)
\]

\[
\text{Symmetry} = \frac{1}{n} \sum_{|d| \leq n-h} \text{Cov}(x_0 x_h, x_d x_{d+h}) = \frac{1}{n} \sum_{|d| \leq n-h} \mathbb{E}(x_0 x_h x_d x_{d+h}) - \mathbb{E}(x_0 x_h) \mathbb{E}(x_d x_{d+h})
\]

\[
\text{E}(x_0 x_h) = 0 \quad \frac{1}{n} \sum_{|d| \leq n-h} \mathbb{E}(x_0 x_h x_d x_{d+h}) \triangleq \frac{1}{n} \sum_{|d| \leq n-h} \sigma(d, h). \quad (2.8)
\]

**Remark.**

1. *Formula (2.8) works for any general series, not specifically for MA series.*

2. *For MA(\( q \)) series, we have*

\[
\sigma(d, h) = 0 \quad \text{if } |d| > q. \quad (2.9)
\]

*Thus \( \text{Var}(\hat{\rho}(h)) \approx \frac{1}{n \hat{\gamma}^2(0)} (\sigma(0, h) + 2\sigma(1, h) + \cdots + 2\sigma(q, h)) \).*

3. *Our estimator for the asymptotic variance should be*

\[
\hat{\text{Var}}(\hat{\rho}(h)) = \frac{1}{n \hat{\gamma}^2(0)} \left( \hat{\sigma}(0, h) + 2\hat{\sigma}(1, h) + \cdots + 2\hat{\sigma}(q, h) \right), \quad (2.10)
\]

*where \( \hat{\sigma}(d, h) = \hat{\mathbb{E}}(x_0 x_h x_d x_{d+h}) = \frac{1}{n} \sum_{i=1}^{n-d-h} x_i x_{h+i} x_{d+i} x_{d+h+i} \).*
4. There is still one difficulty in estimating (2.10): we do not know the order of the MA process $q$, so we use the following formula instead,

$$\hat{\text{Var}}(\hat{\rho}(h)) = \frac{1}{n\hat{\gamma}^2(0)}(\hat{\sigma}(0, h) + 2\hat{\sigma}(1, h) + \cdots + 2\hat{\sigma}(h - 1, h)).$$

(2.11)

The idea is that according to (2.9), we must have

$$\sigma(q + 1, h) = \sigma(q + 2, h) = \cdots = \sigma(h - 1, h) = 0 \quad \text{for} \quad h > q,$

and thus $\hat{\sigma}(q + 1, h) = \hat{\sigma}(q + 2, h) = \cdots = \hat{\sigma}(h - 1, h)$ should be around zero and not influence the estimator of the asymptotic variance much.

At this stage, we have established the whole order determination process for $MA(q)$ process with dependent but uncorrelated innovations. The illustration of the different inferences on ACF process due to different innovation terms is represented in section 2.3.1.

Remark. Please be advised that although the whole process is built for dependent but uncorrelated innovations, it applies to iid innovations automatically. This remark applies to following models too.

2.2.3 Autoregressive Model

Another important fundamental class of time series models is autoregressive model(1.1), and the corresponding order identification tool is partial autocorrelation function(1.8). Autoregressive model is very frequently used since it has much more broader appearance in econometrics and the identification tool related can be adapted to classical linear regression theory and thus easy to figure out. For decades, people have been using partial autocorrelation function (PACF) to identify the $AR(p)$ process by testing the significance of the sample PACF of an
observed time series. More specifically, the PACF of an AR($p$) process cuts off at lag $p$.

It was shown (See Brockwell and Davis[11].) that the partial autocorrelation $\alpha(k)$ at lag $k$ could be regarded as the correlation between $X_t$ and $X_k$, adjusted for the intervening observations $X_{t+1}, \ldots, X_{t+k-1}$, namely $\alpha(k) = \phi_{k,k} = \text{Cor}(X_t, X_{t+k}|X_{t+1}, \ldots, X_{t+k-1})$. Intuitively, $\phi_{k,k}$ is the correlation between $X_t$ and $X_{t-k}$ with the linear effect of $\{X_{t-1}, \ldots, X_{t-k+1}\}$ on each removed.

Since an AR($p$) model purposes the linear relationship between $X_t$ and $X_{t+1}, \ldots, X_{t-p}$, when we set $k > p$, $X_t$ and $X_{t-k}$ should have no correlation with the linear effect of $\{X_{t-1}, \ldots, X_{t-k+1}\}$ removed. This explains the cut-off property of the partial autocorrelation function of the autoregressive time series.

Under further assumption of gaussian white noise, i.e. $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$, it is well known that $\forall s \geq 1, \sqrt{n}\hat{\phi}_{p+1,p+1}, \cdots, \sqrt{n}\hat{\phi}_{p+s,p+s}$ has asymptotic joint normal distribution and with equal variance $\frac{1}{n}$. Namely, we have

$$\sqrt{n}(\hat{\phi}_{p+1,p+1}, \cdots, \hat{\phi}_{p+s,p+s})' \overset{dist}{\to} N\left(\mathbf{0}_{s \times 1}, \mathbf{I}_{s \times s}\right),$$

where $I$ is the diagonal identity matrix.

Note that we only consider $\forall \text{Var}(\hat{\rho}(h))$ for $h > p$ since this is related to identifying an AR($p$) process by testing sample autocorrelation function due to the cut-off property.

Now, as what we have done in last subsection, we need to re-exam the validity of the above procedure with assumption of condition (2.1) and (2.2) on $\{\varepsilon_t\}$ rather than iid assumption or constant conditional variance assumption. Specifically, we need to go through the same steps in this innovation setting: Whether the cut-off property still hold for dependent but uncorrelated innovations? Whether the
sample partial autoregressive function $\hat{\phi}_{h,h}$ still follows the asymptotic joint nor-
mal distribution? Whether the variance formula for sample partial autoregressive
function $\text{Var}(\hat{\phi}_{h,h})$ changes? If it changes, what is the variance formula? How to
get the variance estimator $\hat{\text{Var}}(\hat{\phi}_{h,h})$ if variance does change?

The sample PACF are the same no matter the innovations are iid or dependent
but uncorrelated. But the inference of PACF derived from time series driven by
iid normal innovation could be very misleading when applied to time series whose
innovations $\{\varepsilon_t\}$s are in fact serially dependent. Like what we did earlier, we start
from the validity of the cut-off property first.

**Lemma 2.** Assume $\{X_t\}$ to be an $AR(p)$ process with innovation $\{\varepsilon_t\}$ following
conditions (2.1) and (2.2). For the partial autocorrelation function $\phi_{h,h}$ defined
as (1.8), we have

$$\phi_{p+1,p+1} = \phi_{p+2,p+2} = \cdots = 0. \quad (2.12)$$

The next theorem concerns the properties of sample PACF $\hat{\phi}_{h,h}$ of an $AR(p)$
process with dependent innovations, based on which we can tell the impact on
sample PACF distribution due to different kinds of innovations.

**Theorem 2** Let $\hat{\phi}_{k,k}$ be the lag $k$ sample PACF of $AR(p)$ time series $\{X_t\}$ defined
by (1.9), where the innovations $\{\varepsilon_t\}$s satisfy (2.1) and (2.2). Then $\forall s \geq 1$,

$$\sqrt{n}(\hat{\phi}_{p+1,p+1}, \cdots, \hat{\phi}_{p+s,p+s})' \xrightarrow{\text{dist}} N \left( 0_{s \times 1}, \Xi_s \right). \quad (2.13)$$

If further assume $\text{Cov}(\varepsilon_t^2, \varepsilon_{t+k}^2) > 0$ for all $k \in \mathbb{N}$ and $\mathbb{E}(\varepsilon_t^2 \varepsilon_{t-i} \varepsilon_{t-j}) = 0$, $\forall i \neq j$,
then

$$n\text{Var}(\hat{\phi}_{k,k}) > 1, \quad \text{where} \quad k > p. \quad (2.14)$$
Remark. This theorem has immediate implication. The dependent but uncorrelated innovations such as GARCH innovations render a large variance of sample PACF compared with the case of iid innovations under which the asymptotic variance is $\frac{1}{n}$. So if we still use the traditional 95% confidence interval $\pm 1.96/\sqrt{n}$ to test for PACF zero lag, the type I error would be larger than 5%. This over-rejection leads to a specification of AR($k$) model with $k$ greater than the true order $p$.

We can estimate the variance of sample lag $k$ PACF in the following manner. Suppose we fit an AR($h$) model to $\{X_t\}$ defined in equation (1.1) with $h > p$. By writing $\{X_t\}$ as AR($h$) model with $\phi_{p+1} = \cdots = \phi_{h} = 0$, we can refer to the above theorem to draw inferences on $\phi_1, \cdots, \phi_p$. This reduces to a two-step procedure.

1. Compute the Yule-Walker estimate $\hat{\phi}^{yw}$ and obtain the residuals

$$
\hat{\varepsilon}_t = x_t - \phi_1^{yw} x_{t-1} - \cdots - \phi_m^{yw} x_{t-h}.
$$

2. From the sample estimate of $\Gamma_h$ and $\Omega_h = Cov(\varepsilon_1 x_0, \varepsilon_1 x_{-1}, \cdots, \varepsilon_1 x_{1-m})'$, we obtain an estimate of $\Sigma_h = \Gamma_h^{-1} \Omega_h \Gamma_h^{-1}$. An asymptotic confidence region for $\phi$ can be estimated through $(\hat{\phi} - \phi)' \Sigma_h^{-1} (\hat{\phi} - \phi) \sim \chi^2_h$.

The illustration of the different inferences on PACF process due to different error terms is represented in section 2.3.1.

2.2.4 Autoregressive and Moving Average Model

Autoregressive and moving average model is a hybrid of autoregressive model and moving average model which was first introduction by G. E. Box and quickly
become a widely accepted time series model. Currently, the order determination schemes of ARMA model have been researched extensively. Tsay and Tiao (1984) [74] proposed the extended autocorrelation function (EACF) technique which can tentatively identify the orders of a stationary or non-stationary ARMA process based on iterated least squares estimates of the autoregressive parameters. See chapter 1 for intensive literature review and introduction.

Since the major improvement of this scheme occurs when dealing with moving average sequences which we have already elaborated in subsection 2.2.2, we skip the discussion here and the illustration of the different inferences on EACF process due to different innovation terms is represented in section 2.3.1.

2.2.5 Autoregressive Integrated Moving Average Model

Autoregressive integrated moving average (ARIMA) model can be regarded as an extension of the ARMA model. Dealing with this type of time series sequences, Dickey Fuller test is of critical importance since satisfying the null hypothesis of the test implies that it is appropriate to transform the time series by differencing, i.e., degrade the ARIMA model to an ARMA model and thereby all the order determination tools of ARMA models can simply apply.

The definition of the Dickey Fuller Test is provided in the first chapter. Recently, for simplicity, the limiting distribution is also represented by function of Brownian Motion

\[ \hat{T}_n = n(\hat{\rho} - 1) \xrightarrow{\text{dist}} \frac{1}{2}(1 - \sum_{k=1}^{p-1} a_k) (W^2(1) - 1) \int_0^1 W^2(t) dt, \]  

(2.16)

where \( W \) is a standard Brownian Motion.

The equivalence of formula (1.17) and (2.16) is proved in Theorem 3. We use
the formula (2.16) for the following proof.

Now our question is what is the asymptotic distribution if the innovation \( \{\varepsilon_t\} \) is not a sequence of iid random variables but dependent and uncorrelated random variables satisfying condition (2.1) and (2.2)? We claim that with a new condition on the innovation, there is no change to the limiting distribution of the test statistics \( \hat{T}_n \).

Before we go to the proof, we need the following definition and lemmas.

**Definition 1.** The process \( Y_n = g(\mathcal{F}_n) \), 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

\[
\sum_{n=1}^{\infty} n^r \mathbb{E} \left\| P_1 Y_n \right\|_p < \infty. \tag{2.17}
\]

If (2.17) holds with \( r = 0 \), then \( Y_n \) is said to be \( \mathcal{L}^p \) weakly dependent.

Based on the definition, we require our innovation \( \{\varepsilon_t\} \) to be

\( \mathcal{L}^\alpha(\alpha > 2) \) weakly dependent with order 1. \tag{2.18}

This assumption of innovation allows a large class of nonlinear processes, which substantially relaxes the iid or martingale difference assumption. In particular, this condition is satisfied if the innovations are GARCH, random coefficient AR, bilinear AR, etc. See Wu and Min (2005) [80].

**Lemma 3** Assume \( \{\varepsilon_t\} \) follows condition (2.1) and (2.2), let \( S_n = \sum_{t=1}^{n} \varepsilon_t \), \( V_n^2 = \sum_{t=1}^{n} \varepsilon_t^2 \), \( \|S_n\|_2 = \mathbb{E}(\sum_{t=1}^{n} \varepsilon_t)^2 = \mathbb{E} \sum_{t=1}^{n} \varepsilon_t^2 \). then

\[
\frac{V_n^2}{\|S_n\|_2^2} \overset{a.s.}{\to} 1. \tag{2.19}
\]
For $k = 0, \cdots, n$, set $W_n(t) = \begin{cases} \frac{S_k}{\|S_n\|_2}, & t = k/n; \\ \frac{S_k}{\|S_n\|_2} + (nt - k)\frac{\varepsilon_{k+1}}{\|S_n\|_2}, & k/n \leq t \leq \frac{k+1}{n}. \end{cases}$

**Lemma 4** For time series $\{X_t\}$ following (1.14), let $\theta_i$ be the coefficients at lag $i$ of the MA infinity representation of $\{X_t\}$, $\Theta_n = \sum_{i=0}^n \theta_i$, $B_n^2 = \sum_{i=0}^{n-1} \Theta_i^2$. Assume that the error series $\{\varepsilon_t\}$ satisfies condition (2.18), $\sum_{i=1}^{\infty} (\Theta_{n+i} - \Theta_i)^2 = o(B_n^2)$ and $B_n \to \infty$, then

$$W_n \xrightarrow{\text{dist}} W.$$

**Remark.** 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.

The theorem indicating that the limiting distribution of the test statistics won’t change if we use dependent but uncorrelated errors are stated as follows.

**Theorem 4** If $\rho = 1$, for time series $\{X_t\}$ following (1.14) and errors $\{\varepsilon_t\}$s under the condition (2.1), (2.2) and (2.18), we have

$$W_n \xrightarrow{\text{dist}} W, \quad (2.20)$$

and therefore

$$\widehat{T}_n \xrightarrow{\text{dist}} (1 - \sum_{k=1}^{p-1} a_k)\frac{\int_0^1 W(t)dW(t)}{\int_0^1 W^2(t)dt}. \quad (2.21)$$

**Remark.** Intuitively, the reason why the introduction of dependent but uncorrelated innovations won’t change the result of the dickey fuller test statistics is as follows: $\rho$ is a division of two variances and the variance of $\rho$ should then have fourth order and random walk should dominate(eliminate the effect of) the dependence within innovations in fourth cumulant.
Remark. According to theorem 8.2.1 in Fuller (1995)[29], \( \hat{a}_k \) is a consistency estimator of \( a_k \) and we can use \( \hat{a}_k \) in estimating the limiting distribution, where \( (\hat{\rho}, \hat{a}_1, \cdots, \hat{a}_k)' \) is the vector of regression coefficients obtained by regressing \( X_t \) on \( X_{t-1}, \Delta X_{t-1}, \cdots, \Delta X_{t-p+1} \).

2.3 Application

In this section, we use both simulation and real data to demonstrate the conclusions and methods we developed in the previous section. Simulation results for different schemes respectively are provided first to validate the theorems and thereafter we go through the whole order determination process using real data sets. We call the EACF proposed by Tsay and Tiao (1984)[74] as original EACF and the adjusted EACF procedure for uncorrelated and dependent innovations as modified EACF.

2.3.1 Simulation

In the simulation, we use the \( GARCH(1,1) \) innovations as the dependent but uncorrelated errors since GARCH is among the most common and useful models used to model this type of errors. Namely, we have the following model, (See (1.3) and (1.20) for generalized models.)

\[
\begin{align*}
X_t &= \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \cdots + \psi_q \varepsilon_{t-q}, \\
\varepsilon_t &= \sqrt{g_t} \epsilon_t, \quad g_t = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \beta g_{t-1},
\end{align*}
\]  

(2.22)

\(^1\)Only part of the results are presented; results with different parameters of the GARCH model and results with other innovation models such as stochastic volatility model can be given upon request.
where $\{\epsilon_t\}$s are iid normal random sequences.

**Moving Average Model**

We use the following $MA(3)$ model to simulate observations.

\[
\begin{align*}
X_t &= \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \psi_3 \epsilon_{t-3}, \\
\epsilon_t &= \sqrt{g_t} \epsilon_t, \quad g_t = \alpha_0 + \alpha \epsilon_{t-1}^2 + \beta g_{t-1},
\end{align*}
\]  

(2.23)

with $\psi_1 = 0.8, \psi_2 = -0.8, \psi_3 = 0.8$ and two different GRACH effect settings: i) $\alpha_1 = 0.1, \beta_1 = 0.8$; ii) $\alpha_1 = 0.5, \beta_1 = 0.2$.

We consider two formulae for $\text{Var}(\hat{p}(h))$:

1. $\frac{1}{n}(1 + 2\rho^2(1) + \cdots + 2\rho^2(h))$;

2. $\frac{1}{n\gamma^2(0)}(\sigma(0) + 2\sigma(1) + \cdots + 2\sigma(h))$.

Correspondingly, we denote the statistics $\hat{p}^2(h)/\text{Var}(\hat{p}(h))$ as $T(h)$ and $T^*(h)$ respectively:

1. $T(h) = \frac{\hat{p}^2(h)}{n}(1 + 2\rho^2(1) + \cdots + 2\rho^2(h))$;

2. $T^*(h) = \frac{\hat{p}^2(h)}{n\gamma^2(0)}(\sigma(0) + 2\sigma(1) + \cdots + 2\sigma(h))$.

We simulate 2000 replications of the model with time series of length 1000 each. The empirical percentile of $T(h)$ and $T^*(h)$ with that of $\chi^2_1$ is in the Table 2.1.

**Remark.** From Table 2.1, we conclude that statistics $T^*(h)$ follows the chi-square distribution strictly while the percentile of the statistics $T(h)$ is always larger than that of the chi-square distribution which indicates that $T^*(h)$ outperforms the $T(h)$ in mimicking the asymptotic variance.
Table 2.1: Empirical percentile of $\hat{\rho}^2(h)/\text{Var}(\hat{\rho}(h))$.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T(2)$</td>
<td>1.852505</td>
<td>2.820467</td>
<td>0.804859</td>
<td>2.340128</td>
<td>4.828887</td>
<td>7.465305</td>
<td>13.01293</td>
</tr>
<tr>
<td>$T(3)$</td>
<td>1.821693</td>
<td>2.938174</td>
<td>0.786355</td>
<td>2.30885</td>
<td>4.917539</td>
<td>7.197477</td>
<td>12.23115</td>
</tr>
<tr>
<td>$\chi^2_1$</td>
<td>1</td>
<td>1.414214</td>
<td>0.454936</td>
<td>1.323304</td>
<td>2.705543</td>
<td>3.841459</td>
<td>6.634897</td>
</tr>
<tr>
<td>$T^*(2)$</td>
<td>1.003886</td>
<td>1.402895</td>
<td>0.478995</td>
<td>1.334047</td>
<td>2.656405</td>
<td>3.731442</td>
<td>6.553786</td>
</tr>
<tr>
<td>$T^*(3)$</td>
<td>1.007435</td>
<td>1.332494</td>
<td>0.484573</td>
<td>1.330523</td>
<td>2.806145</td>
<td>3.776094</td>
<td>6.084727</td>
</tr>
</tbody>
</table>

Having confirmed that the variance of $\hat{\rho}(h)$ may change when faced with dependent innovations, we need to use a new order identification scheme.

Since MA model is a specific case of 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 list the result table here (Table 2.2). All the interpretation in section 2.3.1 can be applied here though.

**Autoregressive Model**

We use the following $AR(3)$ model to simulate observations.

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

(2.24)

with $\phi_1 = 0.8, \phi_2 = -0.8, \phi_3 = 0.8$ and two different GRACH effect settings: i) $\alpha = 0.1, \beta = 0.8$; ii) $\alpha = 0.5, \beta = 0.2$.

We face similar situation on autoregressive models as moving average models, thus we just briefly provide the result by Table 2.3 and leave the interpretation into the following subsection.

**Autoregressive Moving Average Model**
Table 2.2: EACF results for a simulated MA(3) series with different type of innovations: $X_t = \varepsilon_t + 0.8 \varepsilon_{t-1} - 0.8 \varepsilon_{t-2} + 0.8 \varepsilon_{t-3}$ where innovation $\varepsilon_t$’s types are: (1) iid; (2) GARCH(1,1) with $\alpha = 0.1, \beta = 0.8$; (3) GARCH(1,1) with $\alpha = 0.5, \beta = 0.2$. The first half are results from original EACF procedure using Bartlett formula while the second half are results from modified EACF procedure.

<table>
<thead>
<tr>
<th>Model specified by <em>Bartlett</em> EACF</th>
<th>Candidate Models</th>
<th>Innovation (1)</th>
<th>Innovation (2)</th>
<th>Innovation (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA(2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MA(3)</td>
<td>851</td>
<td>85.10%</td>
<td>765</td>
<td>76.50%</td>
</tr>
<tr>
<td>MA(4)</td>
<td>66</td>
<td>6.60%</td>
<td>93</td>
<td>9.30%</td>
</tr>
<tr>
<td>MA(5)</td>
<td>26</td>
<td>2.60%</td>
<td>61</td>
<td>6.10%</td>
</tr>
<tr>
<td>ARMA(1,3)</td>
<td>2</td>
<td>0.20%</td>
<td>3</td>
<td>0.30%</td>
</tr>
<tr>
<td>ARMA(1,4)</td>
<td>21</td>
<td>2.10%</td>
<td>30</td>
<td>3.00%</td>
</tr>
<tr>
<td>ARMA(1,5)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(2,3)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(2,4)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(2,5)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(3,4)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Others</td>
<td>34</td>
<td>3.40%</td>
<td>48</td>
<td>4.80%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model specified by <em>Modified</em> EACF</th>
<th>Candidate Models</th>
<th>Innovation (1)</th>
<th>Innovation (2)</th>
<th>Innovation (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA(2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MA(3)</td>
<td>898</td>
<td>89.80%</td>
<td>877</td>
<td>87.70%</td>
</tr>
<tr>
<td>MA(4)</td>
<td>41</td>
<td>4.10%</td>
<td>59</td>
<td>5.90%</td>
</tr>
<tr>
<td>MA(5)</td>
<td>19</td>
<td>1.90%</td>
<td>22</td>
<td>2.20%</td>
</tr>
<tr>
<td>ARMA(1,3)</td>
<td>0</td>
<td>0.00%</td>
<td>1</td>
<td>0.10%</td>
</tr>
<tr>
<td>ARMA(1,4)</td>
<td>16</td>
<td>1.60%</td>
<td>5</td>
<td>0.50%</td>
</tr>
<tr>
<td>ARMA(1,5)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(2,3)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(2,4)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(2,5)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>ARMA(3,4)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Others</td>
<td>26</td>
<td>2.60%</td>
<td>36</td>
<td>3.60%</td>
</tr>
</tbody>
</table>
Table 2.3: EACF results for a simulated AR(3) series with different type of innovations: $X_t = 0.8X_{t-1} - 0.8X_{t-2} + 0.8X_{t-3} + \varepsilon_t$ where innovation \( \varepsilon_t \)’s types are: (1) iid; (2) GARCH(1,1) with \( \alpha = 0.1, \beta = 0.8 \); (3) GARCH(1,1) with \( \alpha = 0.5, \beta = 0.2 \). The first half are results from original EACF procedure using Bartlett formula while the second half are results from modified EACF procedure.

<table>
<thead>
<tr>
<th>Candidate Models</th>
<th>Innovation (1)</th>
<th>Innovation (2)</th>
<th>Innovation (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>AR(3)</td>
<td>579</td>
<td>57.90%</td>
<td>484</td>
</tr>
<tr>
<td>ARMA(3,1)</td>
<td>85</td>
<td>8.50%</td>
<td>100</td>
</tr>
<tr>
<td>ARMA(3,2)</td>
<td>53</td>
<td>5.30%</td>
<td>77</td>
</tr>
<tr>
<td>AR(4)</td>
<td>7</td>
<td>0.70%</td>
<td>9</td>
</tr>
<tr>
<td>ARMA(4,1)</td>
<td>182</td>
<td>18.20%</td>
<td>193</td>
</tr>
<tr>
<td>ARMA(4,2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>AR(5)</td>
<td>12</td>
<td>1.20%</td>
<td>13</td>
</tr>
<tr>
<td>ARMA(5,1)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(5,2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>Others</td>
<td>82</td>
<td>8.20%</td>
<td>124</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Candidate Models</th>
<th>Innovation (1)</th>
<th>Innovation (2)</th>
<th>Innovation (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>AR(3)</td>
<td>648</td>
<td>64.80%</td>
<td>692</td>
</tr>
<tr>
<td>ARMA(3,1)</td>
<td>104</td>
<td>10.40%</td>
<td>86</td>
</tr>
<tr>
<td>ARMA(3,2)</td>
<td>50</td>
<td>5.00%</td>
<td>41</td>
</tr>
<tr>
<td>AR(4)</td>
<td>6</td>
<td>0.60%</td>
<td>5</td>
</tr>
<tr>
<td>ARMA(4,1)</td>
<td>137</td>
<td>13.70%</td>
<td>131</td>
</tr>
<tr>
<td>ARMA(4,2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>AR(5)</td>
<td>3</td>
<td>0.30%</td>
<td>3</td>
</tr>
<tr>
<td>ARMA(5,1)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(5,2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>Others</td>
<td>52</td>
<td>5.20%</td>
<td>42</td>
</tr>
</tbody>
</table>

Model specified by **Bartlett EACF**

Model specified by **Modified EACF**
In this subsection, our task is to validate the difference of the time series order determination procedure EACF for ARMA sequences with uncorrelated but dependent innovations and iid innovations. The simulation goes this way,

1. Simulate time series with different types of innovations (iid, GARCH(with different parameters), stochastic volatility(SV)). The length of the time series is set to be 1000 while the replication number is also set to be 1000.

2. Apply the generated time series to the original extended autocorrelation function (EACF) procedure using Bartlett formula.

3. Identify the model of the generated time series by the EACF table using automatic algorithm 1 and 2 discussed in the preliminary.

4. Count the frequency of different models selected.

5. Test whether the original EACF procedure treat the purely ARMA model (with iid errors) and ARMA with dependent but uncorrelated innovations model differently.

In the simulation, we use $ARMA(1,1)$ as the mean equation with AR and MA term $(0.8, 0.5)$ respectively. We test the EACF procedure with four models of errors: (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 $\alpha = 0.5$.

The reason why we list two settings of GARCH model is that the GARCH effects differ with different parameters. It is stated in Min(2004)[56] that $Var(\hat{\rho})$ is significantly impacted if $Cov(\varepsilon_0^2, \varepsilon_{q-k+i}^2)/E^2(\varepsilon_0^2)$ is large. For instance, if $\{\varepsilon_t\}$ is a GARCH(1,1) process as in our simulation, then $Cov(\varepsilon_0^2, \varepsilon_1^2)/\sigma^4 = 2\alpha + (6\alpha^2(\alpha + \beta/3))/(1 - 2\alpha^2 - (\alpha + \beta)^2)$. Given $\alpha = 0.5$ and $\beta = 0.2$ the ratio is 86; given $\alpha = 0.1$ and $\beta = 0.8$ the ratio is 0.3294118.
The simulation results are in Table 2.4. The mean equation is listed with parameter values in the caption followed by all the innovation we consider and with parameter values if applicable. The table then tabulate all the candidate models with chosen counts and frequency under each innovation type.

Table 2.4: Original EACF results for a simulated ARMA(1,1) series with different type of innovations: \(X_t = 0.8X_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}\) where innovation \(\varepsilon_t\)'s types 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 \(\alpha = 0.5\).

<table>
<thead>
<tr>
<th>Candidate Models</th>
<th>Innovation (1)</th>
<th>Innovation (2)</th>
<th>Innovation (3)</th>
<th>Innovation (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>white noise</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MA(2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MA(3)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>814</td>
<td>81.40%</td>
<td>769</td>
<td>76.90%</td>
</tr>
<tr>
<td>ARMA(1,2)</td>
<td>99</td>
<td>9.90%</td>
<td>115</td>
<td>11.50%</td>
</tr>
<tr>
<td>ARMA(1,3)</td>
<td>41</td>
<td>4.10%</td>
<td>53</td>
<td>5.30%</td>
</tr>
<tr>
<td>AR(2)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>2</td>
<td>0.20%</td>
<td>4</td>
<td>0.40%</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>27</td>
<td>2.70%</td>
<td>36</td>
<td>3.60%</td>
</tr>
<tr>
<td>ARMA(2,3)</td>
<td>8</td>
<td>0.80%</td>
<td>8</td>
<td>0.80%</td>
</tr>
<tr>
<td>AR(3)</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(3,1)</td>
<td>0</td>
<td>0.00%</td>
<td>1</td>
<td>0.10%</td>
</tr>
<tr>
<td>ARMA(3,2)</td>
<td>1</td>
<td>0.10%</td>
<td>1</td>
<td>0.10%</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>5</td>
<td>0.50%</td>
<td>7</td>
<td>0.70%</td>
</tr>
<tr>
<td>Other Models</td>
<td>3</td>
<td>0.30%</td>
<td>6</td>
<td>0.60%</td>
</tr>
</tbody>
</table>

From Table 2.4, the original EACF procedure does not provide as good order determination for ARMA+GARCH Model as for pure ARMA model, say 76.9%(66.2%) compared to 81.4%.

We launch a test to make the conclusion statistically clear. For example, we pick up two simulations in Table 2.4: \(ARMA(1, 1)\) and \(ARMA(1, 1)+GARCH(1, 1)\) with \(\alpha = 0.1\) and \(\beta = 0.8\). Set \(p_1\) as the percentile that EACF picks up the right AR and MA order for model 1: \(ARMA(1, 1)\); \(p_2\) as the percentile that EACF
picks up the right AR and MA order for model 2: \( \text{ARMA}(1,1) + \text{GARCH}(1,1) \).

In our case, \( \hat{p}_1 = 0.814, \hat{p}_2 = 0.769 \).

Our null hypothesis and alternative one are respectively

\[
H_0 : p_1 = p_2 \quad H_a : p_1 \neq p_2.
\]

This is exactly the case of comparing two population proportions in independent sampling. The test statistics
\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n_1} + \frac{\hat{p}(1 - \hat{p})}{n_2}}}
\]
where \( \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \). Thus
\[
z = \frac{0.814 - 0.769}{\sqrt{0.7915 \times (1 - 0.7915) \times \frac{1}{1000}}} = 2.47696.
\]
Since \(|z| > z_{\alpha/2}\) for reasonable \(\alpha\) level, we can reject the null hypothesis and claim that original EACF procedure is less effective in identifying the correct AR and MA order for uncorrelated but dependent innovations than for normal innovations. When we compare \( \text{ARMA}(1,1) \) with another set of \( \text{ARMA}(1,1) + \text{GARCH}(1,1) \) with \( \alpha = 0.1 \) and \( \beta = 0.8 \), the test statistic is even more significant. In conclusion, we need to adjust the original EACF procedure with new variance derived in (2.8). With the modified EACF procedure, we have Table 2.5.

Comparing the new table generated from the modified EACF procedure to the old one generated from original EACF procedure, we can identify easily that the modified EACF works more efficiently than the original EACF procedure.

**Independently identical distributed innovation (column 1)**

In this case, the modified EACF works as well as the Bartlett EACF procedure. This is exactly the case for iid innovation cause the two algorithms are equivalent for iid error terms.

**GARCH innovation and the dependence is not strong (column 2)**

In this case, the improvement from modified EACF compared to Bartlett EACF is clear enough but not strong (76.9% to 82.7%) since the garch effect
Table 2.5: Modified EACF results for a simulated ARMA(1,1) series with different type of innovations: $X_t = 0.8X_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$ where innovation $\varepsilon$’s’ types 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 $\alpha = 0.5$.

<table>
<thead>
<tr>
<th>Candidate Models</th>
<th>Innovation (1)</th>
<th>Innovation (2)</th>
<th>Innovation (3)</th>
<th>Innovation (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>white noise</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MA(2)</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>MA(3)</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>824</td>
<td>824</td>
<td>876</td>
<td>843</td>
</tr>
<tr>
<td>ARMA(1,2)</td>
<td>85</td>
<td>76</td>
<td>44</td>
<td>74</td>
</tr>
<tr>
<td>ARMA(1,3)</td>
<td>39</td>
<td>45</td>
<td>30</td>
<td>37</td>
</tr>
<tr>
<td>AR(2)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>31</td>
<td>33</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>ARMA(2,3)</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>AR(3)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(3,1)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(3,2)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Other Models</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
is not very strong.

**GARCH innovation and the dependence is strong (column 3)**

In this case, the improvement of the modified EACF compared to Bartlett EACF is very clear and strong (66.2% to 87.6%) due to the strong garch effect involved.

**Stochastic volatility type of innovation**

The effect is quite similar to that of GARCH situation which indicates that the improvement of the accuracy and efficiency of the modified EACF procedure does not depend on the innovation type.

**Autoregressive Integrated Moving Average Model**

In this subsection, we study the Dickey Fuller test for ARIMA process with dependent but uncorrelated innovations. As we discussed in the previous section, the Dickey Fuller test should still work and we illustrate the conclusions by a simulation.

The mean equation used is a $ARIMA(1,1,1)$ model with $AR$ coefficient and $MA$ coefficient (0.8, 0.5) correspondingly.

\[
\begin{align*}
\Delta X_t &= 0.8\Delta X_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}, \\
\varepsilon_t &= \sqrt{g_t} \varepsilon_t, \quad g_t = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \beta g_{t-1}.
\end{align*}
\] (2.25)

The innovations used follow GARCH model with two parameter sets: $\alpha = 0.8, \beta = 0.1$ and $\alpha = 0.5, \beta = 0.2$. The results when we set $p$ value as 0.05 are in the Table 2.6.
Table 2.6: Results of Dickey Fuller test (1); replication = 10^5.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number and frequency of time series that test indicated no difference required</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1,0,1) AR = 0.8, MA = 0.5</td>
<td>100000 100.00%</td>
</tr>
<tr>
<td>ARIMA(1,1,1) AR = 0.8, MA = 0.5</td>
<td>4929 4.93%</td>
</tr>
<tr>
<td>ARIMA(1,1,1)+GARCH(1,1) AR = 0.8, MA = 0.5</td>
<td>5132 5.13%</td>
</tr>
<tr>
<td>ARIMA(1,1,1)+GARCH(1,1) AR = 0.8, MA = 0.5</td>
<td>5180 5.18%</td>
</tr>
<tr>
<td>α = 0.5, β = 0.2</td>
<td></td>
</tr>
</tbody>
</table>

The test results should be independent of the AR and MA coefficients chosen. In order to validate this claim, we change the AR and MA coefficients and have another try, namely AR = 0.5 and MA = 0.8. See Table 2.7 for the results.

Table 2.7: Results of Dickey Fuller test (2) with different set of parameters; replication=10^5.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number and frequency of time series that test indicated no difference required</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1,0,1) AR = 0.5, MA = 0.8</td>
<td>100000 100.00%</td>
</tr>
<tr>
<td>ARIMA(1,1,1) AR = 0.5, MA = 0.8</td>
<td>5631 5.63%</td>
</tr>
<tr>
<td>ARIMA(1,1,1)+GARCH(1,1) AR = 0.8, MA = 0.8</td>
<td>6146 6.15%</td>
</tr>
<tr>
<td>ARIMA(1,1,1)+GARCH(1,1) AR = 0.8, MA = 0.8</td>
<td>5886 5.89%</td>
</tr>
<tr>
<td>α = 0.8, β = 0.1</td>
<td></td>
</tr>
</tbody>
</table>

**Remark.** Although the replication has been enlarged to 10^5, the difference between different innovations on the ADF test is not evident which validates our conclusion that the introduction of dependent but uncorrelated innovations would not change the result of the ADF test.
2.3.2 Real Data

We analyze two real data sets utilizing our modified EACF procedure in this subsection: (1) General Motor Stock Price; (2) Big City Traffic Volume.

General Motor Stock Price Data Set

Time series from the financial world have been widely researched and General Motor (GM) data set’s result as we show later perfectly validates our modified EACF procedure on determining ARIMA orders for financial time series such as stock price. We see promptly that the fitting result of the modified EACF procedure is much better than that of the classical one.

The data set (downloaded from Yahoo Finance) contains the stock transaction record on GM within the period from Jan 2nd 1996 to May 8th 2006. Information contained in the data set includes daily opening price, closing price, daily high price, daily low price, daily transaction volume and adjusted closing price while we only use the adjusted closing price in our analysis. In order that the return series to be a stationary one, we use the log-return series. See Figure 2.1 for the series before and after logarithm. (We also provide the plots of the original stock price series without adjustment for dividends and splits.) After transformation, we apply order determination schemes on the log-return series: we first try the classical EACF procedure whose result is shown on the left-hand-side of Table 2.8 and the corresponding modified EACF procedure’s result is shown on the right-hand-side.

From Table 2.8, the modified EACF procedure simplifies the mean equation from ARMA(2,2) to a white noise. Should we fit the log-return series with ARMA(2,2)+GARCH(1,1) model, the fitting result would indicate that all the
Figure 2.1: General motor stock price series.

Table 2.8: Comparison of EACF tables for GM data set.

<table>
<thead>
<tr>
<th></th>
<th>Original EACF Table</th>
<th>Modified EACF Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>0 2 0 0 2 0 0 0</td>
<td>1 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>[2]</td>
<td>2 2 0 0 2 0 0 0</td>
<td>2 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>[3]</td>
<td>2 2 0 0 0 0 0 0</td>
<td>3 2 2 0 0 0 0 0</td>
</tr>
<tr>
<td>[4]</td>
<td>2 0 2 0 0 0 0 0</td>
<td>4 2 0 2 0 0 0 0</td>
</tr>
<tr>
<td>[5]</td>
<td>2 0 2 2 0 0 0 0</td>
<td>5 2 0 2 2 0 0 0</td>
</tr>
<tr>
<td>[6]</td>
<td>2 2 2 2 2 0 0 0</td>
<td>6 2 2 2 2 2 0 0</td>
</tr>
</tbody>
</table>

Table 2.8: Comparison of EACF tables for GM data set.
AR and MA coefficients are insignificant (See Table 2.9), which confirms the white noise model specified by modified EACF procedure to be the correct one.

Table 2.9: The fitting results of GM data set using ARMA(2,2)+GARCH(1,1) model.

\[
\begin{align*}
X_t &= \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2}, \\
\varepsilon_t &= \sqrt{g_t} \varepsilon_t, \quad g_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 g_{t-1}.
\end{align*}
\]

| Parameter | Estimate | Std. Error | t value | Pr(>|t|) | Significance |
|-----------|----------|------------|---------|----------|--------------|
| \( \mu \) | 2.37E-04 | 1.43E-04   | 1.658   | 0.0973   | .            |
| \( \phi_1 \) | 5.55E-01 | 3.69E-01   | 1.506   | 0.1321   | .            |
| \( \phi_2 \) | 9.34E-02 | 3.34E-01   | 0.28    | 0.7794   | .            |
| \( \psi_1 \) | -5.57E-01| 3.67E-01   | -1.52   | 0.1284   | .            |
| \( \psi_2 \) | -1.42E-01| 3.37E-01   | -0.419  | 0.6749   | .            |
| \( \alpha_0 \) | 5.49E-07 | 3.14E-07   | 1.747   | 0.0806   | .            |
| \( \alpha_1 \) | 3.75E-02 | 6.41E-03   | 5.847   | 5.02E-09 | ***          |
| \( \beta_1 \) | 9.62E-01 | 6.45E-03   | 149.1   | <2.00E-16| ***          |

In conclusion, our modified EACF procedure dramatically decreases the order determination workload by figuring out the simpler model straightforwardly while with the classical one, we may end in a larger model and have to correct/decrease the orders via fitting results, which is more time and effort consuming.

**Big City Traffic Volume Data Set**

Another new and growing area casting huge demand for time series models is traffic control field. Huge amount of traffic information including speed, volume has been collected already while little statistical treatments have been proved successful for projection purpose.

We use our modified EACF order determination scheme on a traffic volume data set collected from a major city in China. Each observation in the time series records the traffic volume of a specific location (number of vehicles passing the location) within a time period (3 minutes). There are \( 20 \times 24 = 480 \) observations.
per day. The data set contains 8 weeks’ information and we use the first seven weeks for model identification and the last week to test the forecasting result. See Figure 2.2 for a general idea of the data set.

![Figure 2.2: Metropolitan traffic volume data set.](image)

Similarly, after properly transforming the original time series to a stationary one, we use both the original EACF procedure and modified EACF procedure on the stationary series and the order determination results are listed in Table 2.10.

![Table 2.10: Comparison of EACF tables for traffic data set.](image)

<table>
<thead>
<tr>
<th>Original EACF Table</th>
<th>Modified EACF Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>[1,]</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>[2,]</td>
<td>[2,]</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>[3,]</td>
<td>[3,]</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>[4,]</td>
<td>[4,]</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>[5,]</td>
<td>[5,]</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>[6,]</td>
<td>[6,]</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

According to Table 2.10, the modified EACF procedure has simplified the model from ARMA(1,4) to ARMA(1,1). The fitting result of model ARMA(1,4)+GARCH(1,1)
also illustrates the worthiness of the modified EACF procedure since the coefficients of redundant ARMA terms are all insignificant.

The prediction power of the model specified by modified EACF procedure compared to the model identified by the original EACF procedure are also generated in Table 2.11.

Table 2.11: Sum of prediction error squares: the numbers listed below are based on the summation of \( n = 3360 \) error squares gained from the last week’s transformed series.

<table>
<thead>
<tr>
<th>Model</th>
<th>Prediction Steps Ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>ARMA(1,4)+GARCH(1,1)</td>
<td>29.75342</td>
</tr>
<tr>
<td>ARMA(1,1)+GARCH(1,1)</td>
<td>29.43388</td>
</tr>
</tbody>
</table>

Via Table 2.11, the simpler model performs as well as the complex one, if not slightly better. Given the equivalent prediction and fitting power, the simpler the model, the better the procedure: parsimonious model would be more efficient and stable since it depends on less coefficients. The above arguments validate the superiority of our modified EACF order determination scheme.

2.4 Proof

**Lemma 1.** Let \( \{X_t\} \) to be a \( MA(q) \) process with innovation \( \{\varepsilon_t\} \) following conditions (2.1) and (2.2). For the autocorrelation function \( \rho(h) \) defined as (1.6), we have

\[
\rho(q + 1) = \rho(q + 2) = \cdots = 0.
\]

**Proof:** From the first part of condition (2.1), i.e. \( \mathbb{E}(\varepsilon_n|\mathcal{F}_{n-1}) = 0 \), we have
the uncorrelated condition as follows. For \( i > 0 \), we have

\[
E(\varepsilon_n \varepsilon_{n-i}) = E(E(\varepsilon_n \varepsilon_{n-i}) | \mathcal{F}_{n-1}) = E(\varepsilon_{n-i} E(\varepsilon_n | \mathcal{F}_{n-1})) = E(\varepsilon_{n-i} \times 0) = 0. \tag{2.27}
\]

Thus

\[
\text{Cov}(x_t, x_{t-h}) = E \left[ \left( \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \right) \left( \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-h-j} \right) \right] = E \left( \sum_{i,j=0}^{\infty} \psi_i \psi_j \varepsilon_{t-i} \varepsilon_{t-h-j} \right) \\
\text{by (2.27)} \Rightarrow \sum_{j=0}^{\infty} \psi_{j+h} \psi_j E(a_{t-h-j}^2) \text{ by (2.1)} = \sigma^2 \sum_{j=0}^{\infty} \psi_{j+h} \psi_j \quad \forall h > q. \tag{2.28}
\]

Since \( X_t \) is a MA(q) process, we have \( \psi_h = 0 \), and accordingly

\[
\gamma(h) = \text{Cov}(x_t, x_{t-h}) = \sigma^2 \times 0 = 0 \Rightarrow \rho(h) = 0 \quad \forall h > q.
\]

**Theorem 1.** Let \( \{X_t\} \) be a MA(q) process with innovation \( \{\varepsilon_t\} \) following conditions (2.1) and (2.2); autocorrelation function \( \rho(h) \) is defined as (1.6) and sample autocorrelation \( \hat{\rho}(h) \) is defined as (1.7), then we have

\[
\text{Var}(\hat{\rho}(h)) \geq \frac{1}{n} (1 + 2 \sum_{k=1}^{q} \rho^2(k)). \tag{2.29}
\]

**Remark.** The original proof is provided in Min(2004)[56], and it is reiterated here for reader’s convenience.

**Proof:** Note that we have

\[
\hat{\gamma}(0) \rightarrow \gamma(0) \quad \text{and} \quad \sqrt{n} \hat{\gamma}(h) \rightarrow N(0, \|\xi\|^2),
\]

where \( \xi = \sum_{i=1}^{\infty} P x_{t-h} x_i = \varepsilon_1 \sum_{i=0}^{q} \psi_i x_{1-h+i} \). Moreover, asymptotically, \( \text{Var}(\hat{\gamma}(h)) = \frac{1}{n} \|\xi\|^2 \) and \( \text{Var}(\hat{\rho}(h)) = \frac{1}{n} \|\xi\|^2 / \gamma^2(0) \).
We have the following calculation form,

\[
\|\xi\|^2 = \mathbb{E}(\varepsilon_1 \sum_{i=0}^{q} \psi_i x_{1-h+i})^2 \geq \mathbb{E}(\varepsilon_1^2) \mathbb{E}(\sum_{i=0}^{q} \psi_i x_{1-h+i})^2 \\
= \sigma^2(\sum_{i=0}^{q} \psi_i^2 \gamma(0) + 2 \sum_{k=1}^{q} \sum_{i=0}^{q} \psi_i \psi_{i+k} \gamma(k)) = \gamma^2(0) + 2 \sum_{k=1}^{q} \gamma^2(k).
\]

Thus

\[
\text{Var}(\hat{\rho}(h)) = \frac{1}{n} \|\xi\|^2/\gamma^2(0) \geq \frac{1}{n} \frac{1}{\gamma^2(0)}(\gamma^2(0) + 2 \sum_{k=1}^{q} \rho^2(k)) = \frac{1}{n} (1 + 2 \sum_{k=1}^{q} \rho^2(k))
\]

Remark. When \(\mathbb{E}(\varepsilon_n^2 | F_{n-1}) = \sigma^2\), we have the strict equality and this is just the situation of iid case.

Lemma 2. Assume \(\{X_t\}\) to be an AR\((p)\) process with innovation \(\{\varepsilon_t\}\) following conditions (2.1) and (2.2). For the partial autocorrelation function \(\alpha(h) = \phi_{h,h}\) defined as (1.8), we have

\[
\phi_{p+1,p+1} = \phi_{p+2,p+2} = \cdots = 0.
\]

Proof: We need to show that the PACF of a AR\((p)\) process is zero for lags greater than \(p\).

For \(h \geq p\), the best linear predictor of \(X_{h+1}\) in terms of \(X_1, \cdots, X_h\) is

\[
\hat{X}_{h+1} = \phi_1 X_h + \phi_2 X_{h-1} + \cdots + \phi_p X_{h+1-p}.
\]

Since by definition, \(\phi_{hh}\) is the last component of \(\phi_h = \Gamma_h^{-1} \gamma_h\) where \(\Gamma_h = [\gamma(i-j)]_{i,j=1}^{h}\) and \(\gamma_h = [\gamma(1), \gamma(2), \cdots, \gamma(h)]'\), the coefficient \(\phi_{hh}\) of \(X_1\) is \(\phi_p\) if \(h = p\) and 0 if \(h > p\), we conclude that the PACF \(\alpha(\cdot)\) of the process \(\{X_t\}\) has
the properties

\[ \alpha(p) = \phi_p \text{ and } \alpha(h) = 0 \text{ for } h > p. \]

**Theorem 2.** Let \( \hat{\phi}_{k,k} \) be the lag \( k \) sample PACF of AR(\( p \)) time series \( \{X_t\} \) defined by (1.9), where the innovations \( \{\varepsilon_t\} \) satisfies (2.1) and (2.2). Then \( \forall s \geq 1, \)

\[ \sqrt{n}(\hat{\phi}_{p+1,p+1}, \ldots, \hat{\phi}_{p+s,p+s})' \xrightarrow{\text{dist}} N(0, \Xi), \tag{2.31} \]

If further assume \( \text{Cov}(\varepsilon_t^2, \varepsilon_{t+k}^2) > 0 \) for all \( k \in \mathbb{N} \) and \( \mathbb{E}(\varepsilon_t^2 \varepsilon_{t-i} \varepsilon_{t-j}) = 0, \forall i \neq j, \) then

\[ n \text{Var}(\hat{\phi}_{k,k}) > 1 \quad \text{where} \quad k > p. \tag{2.32} \]

**Remark.** The original proof is provided in Min(2004)[56], and the first part of the proof is reiterated here for reader’s convenience.

**Proof:** Notice that 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 value of \( \phi_{p+1}, \ldots, \phi_k \) are all zero. For each \( p < k \leq p+h, \) \( \sqrt{n}\hat{\phi}_{k,k} = [(X'X/n)^{-1}]_{k} (X'\varepsilon)/\sqrt{n} \) has the same limit law as \( [(\Gamma_k^{-1})_{k}]_{k} (X'\varepsilon)/\sqrt{n} \), where \( X \in \mathbb{R}^{n \times k} \) and \( \varepsilon \in \mathbb{R}^{n \times 1} \) are defined as

\[
X = \begin{pmatrix}
x_0 & x_{-1} & \cdots & x_{1-k} \\
x_1 & x_0 & \cdots & x_{2-k} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n-1} & x_{n-2} & \cdots & x_{n-k}
\end{pmatrix}, \quad \varepsilon = \begin{pmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\vdots \\
\varepsilon_n
\end{pmatrix},
\]

and \( [(\Gamma_k^{-1})_{k}]_k \) is the \( k \)th row of \( \Gamma_k^{-1} \) where \( \Gamma_k \triangleq (\gamma(i-j))_{1 \leq i,j \leq k}. \)

Clearly, \( [(\Gamma_k^{-1})_{k}]_{k} (X'\varepsilon)/\sqrt{n} \) is a linear combination of \( X'_{p+h} \varepsilon/\sqrt{n} \) where \( X_{p+h} \in \mathbb{R}^{n \times (p+h)} \) has similar structure as \( X. \) \( X'_{p+h} \varepsilon/\sqrt{n} \) has an asymptotic \((h+p)\)-variate
normal distribution with zero mean, therefore its linear combination \([\Gamma^{-1}_k]_k(X^t\varepsilon)/\sqrt{n}(p < k \leq p + h)\) also have a joint normal distribution with zero mean, which in turn implies the asymptotic joint normal distribution of \(\sqrt{n}(\hat{\phi}_{p+1,p+1}, \cdots)'\) with zero mean.

To show (2.32), \(n\text{Var}(\hat{\phi}_{k,k}) = \sum_k(k,k) = (\Gamma^{-1}_k\Omega_k\Gamma^{-1}_k)(k,k)\), where \(\sum_k(k,k)\) stands for the \((k,k)\) entry of the matrix \(\sum_k\). We show \(\Omega_k > \sigma^2\Gamma_k\), i.e., \(\Omega_k - \sigma^2\Gamma_k\) is positive definite. To this end, consider any nonrandom \(C = (c_0, c_{-1}, \cdots, c_{-k})' \in \mathbb{R}^k\) with a positive \(L_2\) norm,

\[
C'\Omega_k C = C'\text{Cov}(\varepsilon_1 x_0, \varepsilon_1 x_{-1}, \cdots, \varepsilon_1 x_{-k})C = \mathbb{E}[\varepsilon_1^2(c_0 x_0 + c_{-1} x_{-1} + \cdots + c_{-k} x_{-k})^2]
\]

\[
= \mathbb{E}[\varepsilon_1^2(\sum_{i=0}^{\infty} \kappa_i \varepsilon_{-i})^2] \quad \text{MA representation of } x_t
\]

\[
= \sum_{i,j=0}^{\infty} \kappa_i \kappa_j \mathbb{E}(\varepsilon_1^2 \varepsilon_{-i} \varepsilon_{-j}) = \sum_{i=0}^{\infty} \kappa_i^2 \mathbb{E}(\varepsilon_1^2 \varepsilon_{-i}) \quad i \neq j \text{ terms vanish}
\]

\[
> \sum_{i=0}^{\infty} \kappa_i^2 \mathbb{E}(\varepsilon_1^2) \mathbb{E}(\varepsilon_{-i}^2) \quad \text{ACF of } \varepsilon_t^2 \text{ are positive.}
\]

\[
= \sigma^2 \sum_{i=0}^{\infty} \mathbb{E}(\varepsilon_{-i}^2) = \sigma^2 C' \Gamma_k C > 0
\]

So \(\Gamma^{-1}_k\Omega_k\Gamma^{-1}_k > \Gamma^{-1}_k\sigma^2\Gamma_k\Gamma^{-1}_k = \sigma^2\Gamma^{-1}_k\). It is known that the \((k,k)\) entry of \(\Gamma^{-1}\) is \(1/\sigma^2\) for \(k > p\). The result follows.

**Theorem 3.** Assume \((\Gamma, T) = (\sum_{i=1}^{\infty} \gamma_i^2 Z_i^2, \sum_{i=1}^{\infty} \sqrt{2} \gamma_i Z_i)\) where \(\gamma_i^2 = 4[(2i - 1) \pi]^{-2}\) and \(\{Z_i\}_{i=1}^{\infty}\) is a sequence of iid \(N(0, 1)\) random variables; also assume \(W\) to be a standard Brownian Motion, we have the following equation

\[
\frac{(W^2(1) - 1)}{\int_0^1 W^2(t)dt} = (\Gamma^{-1}(T^2 - 1)).
\]
Proof: Expand the Brownian motion by the reproducing kernel method

\[ W(t) = \sum_{n=0}^{\infty} \frac{2\sqrt{2}}{(2n+1)\pi} (\sin(n + \frac{1}{2})\pi t) Z_n, \]

where \( Z_n \) are iid standard normal random variables.

Note that \( \sin(n + \frac{1}{2})\pi t \) belongs to a set of orthogonal basis which have the property that

\[ \int_{-1}^{1} (\sin(n + \frac{1}{2})\pi t) (\sin(m + \frac{1}{2})\pi t) \, dt = \delta_{mn} = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases}. \]

Then

\[ \int_{0}^{1} W^2(t) \, dt = \sum_{n=0}^{\infty} \left( \frac{2\sqrt{2}}{(2n+1)\pi} \right)^2 Z_n^2 \int_{0}^{1} (\sin(n + \frac{1}{2})\pi t)^2 \, dt = \sum_{n=0}^{\infty} \left( \frac{2\sqrt{2}}{(2n+1)\pi} \right)^2 Z_n^2 = \Gamma, \]

and

\[ W(1) = \sum_{n=0}^{\infty} \frac{2\sqrt{2}}{(2n+1)\pi} (\sin(n + \frac{1}{2})\pi) Z_n = \sum_{n=0}^{\infty} \sqrt{2} \frac{2(-1)^n}{(2n+1)\pi} Z_n = \sum_{n=0}^{\infty} \sqrt{2}\gamma_n Z_n = T. \]

Thus

\[ \frac{(W^2(1) - 1)}{\int_{0}^{1} W^2(t) \, dt} = \frac{T^2 - 1}{\Gamma}. \]

In conclusion, we equate (1.17) and (2.16). Thus these two formula are equivalent.

Lemma 3. Assume \( \{\varepsilon_t\} \) follows condition (2.1) and (2.2), let \( S_n = \sum_{t=1}^{n} \varepsilon_t. \)
\[ V_n^2 = \sum_{t=1}^{n} \varepsilon_t^2, \|S_n\|_2^2 = \mathbb{E}(\sum_{t=1}^{n} \varepsilon_t)^2 = \mathbb{E} \sum_{t=1}^{n} \varepsilon_t^2. \text{ then} \]
\[ \frac{V_n^2}{\|S_n\|_2^2} \overset{a.s.}{\to} 1. \] (2.33)

**Proof:** Since \( \|S_n\|_2^2 = \mathbb{E}(\sum_{t=1}^{n} \varepsilon_t)^2 = \mathbb{E} \sum_{t=1}^{n} \varepsilon_t^2 = n\sigma^2 \), (2.33) changes to \( \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2 \overset{P}{\to} \sigma^2 \). Using Birkhoff’s Ergodic Theorem in Durrett[24], we get the convergence immediately.

**Lemma 4.** For time series \( \{X_t\} \) following (1.14), let \( \theta_i \) to be the coefficients at lag \( i \) of the MA infinity representation of \( X_t \), \( \Theta_n = \sum_{i=0}^{n} \theta_i \), \( B_n^2 = \sum_{i=0}^{n-1} \Theta_i^2 \). Assume that the error series \( \{\varepsilon_t\} \) satisfies condition (2.18), \( \sum_{i=1}^{\infty} (\Theta_{n+i} - \Theta_i)^2 = o(B_n^2) \) and \( B_n \to \infty \), then
\[ W_n \overset{\text{dist}}{\longrightarrow} W. \]

**Proof:** See theorem 1 of Wu and Min(2005) [80].

**Theorem 4.** If \( \rho = 1 \), for time series \( \{X_t\} \) following (1.14) and errors \( \{\varepsilon_t\} \) under the condition (2.1),(2.2) and (2.18) , we have
\[ W_n \overset{\text{dist}}{\longrightarrow} W, \] (2.34)

and therefore
\[ \hat{T}_n \overset{\text{dist}}{\longrightarrow} (1 - \sum_{k=1}^{p-1} a_k) \frac{\int_0^1 W(t)dW(t)}{\int_0^1 W^2(t)dt}. \] (2.35)

**Proof:** Recall that \( S_n = \sum_{i=1}^{n} \varepsilon_i \) and \( V_n^2 = \|S_n\|_2^2 = \sum_{i=1}^{n} \varepsilon_i^2 \). \( \hat{\rho}_n \) is the least square estimator of \( \rho \)
\[ \hat{\rho}_n = \frac{\sum_{t=p}^{n} X_{t-1}(X_t - \sum_{k=1}^{p-1} \hat{a}_k \Delta X_{t-k})}{\sum_{t=p}^{n} X_{t-1}^2}. \]
and we immediately have

\[
\hat{T}_n = n(\hat{\rho}_n - 1) = n\frac{\sum_{t=p}^{n} X_{t-1}(X_t - X_{t-1} - \sum_{k=1}^{p-1} \hat{a}_k \Delta X_{t-k})}{\sum_{t=p}^{n} X_t^2} = \frac{1}{n} \frac{\sum_{t=p}^{n} X_{t-1} \varepsilon_t}{\sum_{t=p}^{n} X_t^2} + O_p(1).
\]

Using the polynomial decomposition given by Phillips and Solo[66], we get

\[
X_t = \psi_n(1) \sum_{k=1}^{t} \varepsilon_k + \eta_t - \eta_0,
\]

where \( \psi_n(1) \xrightarrow{p} \psi(1) = (1 - \sum_{k=1}^{p-1} a_k)^{-1} \), and \( \eta_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j} \) where \( \alpha_j = -\sum_{i=j+1}^{\infty} \psi_i \).

Notice that for the numerator,

\[
\frac{1}{n} \sum_{t=p}^{n} X_{t-1} \varepsilon_t = \frac{\psi_n(1)}{n} \sum_{t=p}^{n} (\varepsilon_j) \varepsilon_t + \frac{1}{n} \sum_{t=p}^{n} (\eta_{t-1} - \eta_0) \varepsilon_t \overset{A}{=} Q_1 + Q_2,
\]

where \( Q_1 = \frac{\psi_n(1)}{2n} \left( (\sum_{j=1}^{n} \varepsilon_j)^2 - \sum_{j=1}^{n} \varepsilon_j^2 \right) = \frac{\psi_n(1)}{2n} (S_n^2 - V_n^2) \) and \( \mathbb{E}Q_2 = 0 \) and \( \mathbb{E}Q_2^2 = O\left(\frac{1}{n}\right) \). Meanwhile, for the denominator, we have

\[
\frac{1}{n^2} \sum_{t=p}^{n} X_t^2 = \psi_n^2(1) \frac{1}{n^2} \sum_{t=p}^{n} (\varepsilon_j)^2 + o_p(1) = \psi_n^2(1) \frac{1}{n^2} \sum_{t=p}^{n} S_t^2 + o_p(1).
\]

So

\[
\hat{T}_n \overset{P}{\rightarrow} \psi_n^{-1}(1) \frac{1}{n} \sum_{t=p}^{n} S_t^2 - \psi_n^{-1}(1) \frac{1}{n} \left( (S_n/V_n)^2 - 1 \right) + o_p(1) \]

\[
\overset{\text{dist}}{\rightarrow} P \left( \frac{1}{n} \sum_{k=1}^{p-1} a_k \int_0^1 W(t) dW(t)\right) \text{ by lemma (3)}
\]

\[
\text{dist} \rightarrow (1 - \sum_{k=1}^{p-1} a_k \int_0^1 W(t) dW(t)) \text{ by lemma (4)}.
\]
Chapter 3
Automatic Order Identification for Seasonal ARIMA Series

3.1 Introduction

In the application of time series modeling and forecasting, seasonal patterns occur frequently and deserve detailed investigation. As is discussed in Chapter 1, the widely-used seasonal autoregressive moving average model will be the underlying model to discuss. Consider a general multiplicative seasonal model $SARIMA(p, d, q) \times (P, D, Q)^s$ as (1.21), whose mean equation is

$$\phi(B)\Phi(B^s)\Delta^d\Delta^D_s X_t = \theta(B)\Theta(B^s)\varepsilon_t$$ (3.1)

where $s$ is the seasonal lag\(^1\); $\Delta^D_s = (1 - B^s)^D$, $\Delta^d = (1 - B)^d$; $\Phi(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps}$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\Theta(B) = 1 + \Theta_1 B^s + \cdots + \Theta_Q B^{Qs}$ and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ are polynomials in the back-shift operator $B$, defined as $BX_t = X_{t-1}$, $B\varepsilon_t = \varepsilon_{t-1}$, $B^sX_t = X_{t-s}$ and $B^s\varepsilon_t = \varepsilon_{t-s}$; $\varepsilon_t$ is the innovation term. We shall require that all zeros of $\phi(B), \Phi(B), \theta(B)$ and $\Theta(B)$ are outside the unit circle.

Seasonal autoregressive model (SAR) is a special case of model (1.21) if we set $\Theta(B^s) = 1$ while seasonal moving average model (SMA) is another special case if

---

\(^1\)An example of $s$: within a monthly data series, if there exists a yearly pattern, then $s = 12$. Reasonably, we assume $s$ to be fixed and known to us. However, we can easily allow $s$ to be another parameter and compare models with different combinations of seasonal orders, nonseasonal orders and seasonal lag $s$. 
we set $\Phi(B^s) = 1$ and $\Delta^D_s = 1$.

Identification of tentative orders is the very beginning step in building a sophisticated time series model and has been found to be an important but difficult stage. The complexity is further compounded when seasonality presents in the time series.

In the literature, Brockwell and Davis (1986) [11] suggested using the behavior of autocorrelation function (ACF) and partial autocorrelation function (PACF) on seasonal lags ($s, 2s, \cdots$) to determine the seasonal orders and using their nonseasonal behavior on nonseasonal lags ($1, 2, \cdots$) to determine the nonseasonal orders. While ACF and PACF have clear patterns for pure (S)MA and (S)AR series, their strength in identifying ARIMA models are quite limited, not mention the mixture or seasonal and nonseasonal ARIMA models (SARIMA model). Tsay and Tiao (1984) [74] proposed the extended autocorrelation function (EACF) which is very powerful for order determination of ARIMA models. However, it is not specifically designed for seasonal time series: the mixture of non-seasonal ARIMA effect and seasonal ARIMA effect leads to an unclear and messy EACF table $^2$. Liu (1989) [51] presented a filtering method which works specifically for seasonal ARIMA models. Nevertheless, the intermediary model introduced in the paper is ad-hoc and heuristic.

In this chapter, we propose the unified approach for the tentative specification of both the seasonal and nonseasonal orders of a general multiplicative seasonal model (1.21). Our method treats seasonal patterns differently: if we assume the seasonal pattern of the series follows the seasonal autoregressive (SAR) model, an iterative regression procedure is given to produce consistent estimates of the seasonal autoregressive parameters while the nonseasonal orders are then determined by the extend autocorrelation function (EACF) of the remaining sequences

$^2$See detailed discussion on Section 3.2.
filtering the seasonal effects; on the other hand, if we assume the seasonal moving average (SMA) model and seasonal autoregressive integrated moving average (SARIMA) model, another iterative fitting and filtering procedure using maximum likelihood estimators is provided instead. The Automatic Seasonal Nonseasonal Order Identification System (ASNOIS) based on above identification tools is then set up and can be used to identify seasonal and nonseasonal orders. This new approach has the advantages of determining the seasonal and nonseasonal orders simultaneously and automatically: not only can the ASNOIS provide both seasonal and nonseasonal orders at the same time but also can it return the best order combinations by only inputting the time series to be identified without any other human instructions.

This chapter is organized as follows. In section 3.2, we present a detailed study on the methodology of the ASNOIS. Performances on simulations are represented in section 3.3, illustrating the strength of the ASNOIS, while proofs are left in section 3.4.

### 3.2 Methodology

For nonseasonal ARIMA models, extended autocorrelation function (EACF) is a very powerful order determination tool. It is very generic: works for $AR$, $MA$, $ARMA$ and $ARIMA$ models. The result is represented by an EACF matrix where numbers in each cell can be either 0 or 1, where 0 represents insignificant, and 1 indicates significant. Upper-right zero triangle starting from cell $(p, q)$ indicates an $ARMA(p - 1, q - 1)$ model. See also Table 1.1 and the discussion on Chapter 1. However, for multiplicative seasonal model, unfortunately, the mixture of non-seasonal effect and seasonal effect result in an unclear and messy
EACF matrix: significant $1, \cdots, q, s, s+1, \cdots, s+q, 2s, \cdots$ columns occur for series following $SMA$ model and significant $1, 2, \cdots, s, s+1, \cdots, s+p$ rows occur for series following $SAR$ model and more complicated structures are associated with $SARIMA$ series. Since $s$ is usually relatively large, we have to check a huge EACF table; even we can see the pattern from the huge EACF table, we are still not sure how to separate the seasonal pattern and nonseasonal pattern from the table directly.

If we can instead isolate the seasonal part of model (1.21) from the nonseasonal part, namely, transfer model (1.21) into

\[
\Phi(B^s)\Delta_{s}^D X_t = \Theta(B^s)Y_t,
\]

\[
\phi(B)\Delta_{d}^d Y_t = \theta(B)\varepsilon_t,
\]

(3.2)

and then treat these two equations independently by EACF method, we can gain seasonal orders from the first formula and nonseasonal orders from the second formula.

As you may see, this transformation of the original model (1.21) has technical difficulty lying in how to get the $Y_t$. Nevertheless, it does shed lights on how to solve this problem and an important feature of the seasonal models provides us a feasible path. In practice, $D$ is rarely more than one; $P$ and $Q$ typically add up less than two. In another word, a $SARMA(P, D, Q)$ with $P + D + Q \leq 2$ is sufficient usually and we rarely encounter a higher order multiplicative seasonal model. (See Brockwell and Davis[11].)

Based on above facts, it would not be too much workload if we enumerate all possible seasonal patterns, get the consistent estimates of the seasonal coefficients ($\Phi(B^s), \Delta_{s}^D$ and $\Theta(B^s)$) for every seasonal patterns, eliminate the seasonal impact by transforming $X_t$ into $Y_t$, and finally we can get the nonseasonal orders by
checking EACF matrix of the $Y_t$. This is our first step.

Afterwards, for each seasonal pattern (seasonal part), we have a corresponding EACF table (nonseasonal part). A comprehensive penalty score should be set up in our second step to select the best order combination from competitive alternatives balancing the seasonal and nonseasonal orders. Intuitively, correct seasonal assumption would come up with a clear nonseasonal EACF matrix with lower $p$ and $q$ orders.

**Remark.** We may also reversely separate the nonseasonal impact from $X_t$ and check EACF table for the remaining series concerning the seasonal orders. However, the task to enumerate all possible nonseasonal patterns is relative more consuming and less practical. This is exactly how practically small seasonal orders show its strength.

We discuss the two steps in the following subsections.

### 3.2.1 Separating Seasonal Impact and Getting Nonseasonal Orders

Since a $SARMA(P, D, Q)$ model with $P + D + Q \leq 2$ is sufficient for the seasonal pattern, we select $SAR(1)$, $SAR(2)$, $SMA(1)$, $SMA(2)$ and $SARMA(1, 1)$ as our candidates for seasonal models. For simplicity, we consolidate difference lags ($D$) into autoregressive lags ($P$). The methods to obtain the consistent estimates of the seasonal coefficients and thus filter the seasonal impacts are different: for the seasonal autoregressive (SAR) model, an iterative regression procedure is given to produce consistent estimates of the seasonal autoregressive parameters while the nonseasonal orders are then determined by the extend autocorrelation function (EACF) of the remaining sequences filtering the seasonal effects; for the
seasonal moving average (SMA) model and seasonal autoregressive integrated moving average (SARIMA) model, another iterative fitting procedure using maximum likelihood estimators is provided instead.

**Assume SAR as the seasonal model**

Let us take $SAR(1)$ model for demonstration purpose. A more detailed representation of equation (1.21) for $SAR(1)$ model is

$$(1 - \Phi B^s)(1 - \phi_1 B - \cdots - \phi_p B^p)X_t = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t,$$  

(3.3)

where the seasonal lag is $s$, the seasonal $AR$ parameter is $\Phi$, the nonseasonal $AR$ parameters are $\phi_1, \cdots, \phi_p$ and the nonseasonal $AR$ parameters are $\theta_1, \cdots, \theta_q$. We can reasonably assume $p < s$.

In the first place, we are interested in estimating $\Phi$. We can use the recursive formula (See Tsay and Tiao(1984)[74].) to get consistent estimators $\hat{\beta}_1, \cdots, \hat{\beta}_{p+s}$ where

$$
\hat{\beta}_i \xrightarrow{P} \phi_i \quad i = 1, 2, \cdots, p \quad n \rightarrow \infty, \\
\hat{\beta}_s \xrightarrow{P} \Phi \quad n \rightarrow \infty, \\
\hat{\beta}_{s+i} \xrightarrow{P} -\phi_i \Phi \quad i = 1, 2, \cdots, p \quad n \rightarrow \infty, \\
\hat{\beta}_i \xrightarrow{P} 0 \quad \text{otherwise} \quad n \rightarrow \infty. \quad (3.4)
$$

Thus $\hat{\beta}_s$ is a candidate for estimating $\Phi$ and we name it as recursive estimator.

Instead of using the estimators $\hat{\beta}_1, \cdots, \hat{\beta}_p, \hat{\beta}_s$ as the consistent estimators for
the AR coefficients, based on them, we can use nonlinear least square estimators

\[ \hat{\phi}_1, \cdots, \hat{\phi}_p, \hat{\Phi} = \arg \min \sum_{i=1}^{p} (\hat{\phi}_i - \hat{\beta}_i)^2 + (\hat{\Phi} - \hat{\beta}_s)^2 + \sum_{i=1}^{p} (-\hat{\phi}_i \hat{\Phi} - \hat{\beta}_i + s)^2. \quad (3.5) \]

We call the estimator \( \hat{\Phi} \) established by (3.5) the least square estimator and it is consistent too by the following Theorem. See the proof of Theorem 5 in Section 3.4.

**Theorem 5** If \( \hat{\beta}_1, \cdots, \hat{\beta}_p \) and \( \hat{\beta}_s \) are consistent estimators such that (3.4) holds, then the nonlinear least square estimators obtained from (3.5) are also consistent.

The superiority of the least square estimator over the recursive estimator is demonstrated in Table 3.1 and Table 3.2 which correspond two SAR(1) series: it has smaller variance and standard deviation; its mean is slightly closer to the true mean; its range is much tighter; it has less extreme observations (outliers).

**Table 3.1:** Comparison of the recursive and least square estimators for SAR(1) model \((1 - 0.8B^4)(1 - 0.8B)X_t = (1 + 0.8B)\varepsilon_t.\)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>variance</th>
<th>sd</th>
<th>mean</th>
<th>median</th>
<th>minimum</th>
<th>maximum</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recursive</td>
<td>2.503717</td>
<td>1.582314</td>
<td>0.8214474</td>
<td>0.8022436</td>
<td>-40.62024</td>
<td>47.33023</td>
<td>87.95047</td>
</tr>
<tr>
<td>Least Square</td>
<td>0.213317</td>
<td>0.461863</td>
<td>0.7995064</td>
<td>0.8043294</td>
<td>-23.98460</td>
<td>2.188101</td>
<td>26.17270</td>
</tr>
</tbody>
</table>

**Table 3.2:** Comparison of the recursive and least square estimators for SAR(1) model \((1 - 0.8B^7)(1 - 0.8B + 0.8B^2)X_t = (1 + 0.8B - 0.8B^2)\varepsilon_t.\)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>variance</th>
<th>sd</th>
<th>mean</th>
<th>median</th>
<th>minimum</th>
<th>maximum</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recursive</td>
<td>38.6375</td>
<td>6.215907</td>
<td>0.732029</td>
<td>0.7925728</td>
<td>-509.0917</td>
<td>125.0832</td>
<td>634.1749</td>
</tr>
<tr>
<td>Least Square</td>
<td>0.001169</td>
<td>0.03418</td>
<td>0.7914533</td>
<td>0.7941777</td>
<td>0.5889048</td>
<td>0.8819789</td>
<td>0.2930741</td>
</tr>
</tbody>
</table>

After obtaining the consistent least square estimator \( \hat{\Phi} \), we can make the transformation \( Y_t = X_t - \hat{\Phi}X_{t-s} \) and \( Y_t \) should follow a nonseasonal ARMA \((p,q)\) model if \( SAR(1) \) is the correct seasonal model and EACF table on the \( Y_t \) series should help us to figure out the nonseasonal orders.
It is trivial to extend our method to \(SAR(2)\) model from \(SAR(1)\) model: Instead of having \(\hat{\Phi}\), we right now estimate \(\Phi_1\) and \(\Phi_2\) and the transformation becomes \(Y_t = X_t - \hat{\Phi}_1 X_{t-s} - \hat{\Phi}_2 X_{t-2s}\).

**Assume SMA as the seasonal model**

The treatment for the seasonal moving average model is quite different from what we have done to the seasonal autoregressive model. Similarly, let us take \(SMA(1)\) model for demonstration purpose. A more detailed representation of equation (1.21) for \(SMA(1)\) model is

\[
(1 - \phi_1 B - \cdots - \phi_p B^p)X_t = (1 + \Theta B^s)(1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t, \tag{3.6}
\]

where the seasonal lag is \(s\), the seasonal MA parameter is \(\Theta\), the nonseasonal AR parameters are \(\phi_1, \cdots, \phi_p\) and the nonseasonal AR parameters are \(\theta_1, \cdots, \theta_q\). We can also reasonably assume \(q < s\).

The key difficulty lies in how to get the consistent estimator of \(\Theta\) in the seasonal MA model rather than \(\Phi\) in the seasonal AR model. Since

\[
\frac{(1 - \phi_1 B - \cdots - \phi_p B^p)}{1 + \Theta B^s} = (1 - \phi_1 B - \cdots - \phi_p B^p)(1 - \Theta B^s + \Theta^2 B^{2s} + (-\Theta)^n B^{ns} - \cdots)
\]

is an infinite series, we cannot fit it with a finite set of AR parameters and then use the recursive formula.

Alternatively, given \(SMA(1)\) model, we can use maximum likelihood method to fit for all combinations of non-seasonal order pairs \((p, q)\), utilize the estimated parameters \(\hat{\phi}_i\) and \(\hat{\Theta}\) to get the transferred series(moving average series if the order is correct), and then check the autocorrelation function of the remaining series. The result table is also a matrix and should have the same upper zero
triangle as the EACF table. The major criticism to this method is its slow speed and huge computation cost.

Instead, we can use the following iterative method which can dramatically improve efficiency:

1. Start from a relatively large nonseasonal orders \((p_0, q_0)\).\(^3\) Fit the \(SMA(1)\) model using maximum likelihood method and get the tentative seasonal moving average coefficient \(\hat{\Theta}_0\).

2. Filter the original series by Taylor expansion of \(Y_t = (1 + \hat{\Theta}_0 B^s)^{-1}X_t\) and analyze the EACF table on \(Y_t\) to get the updated tentative nonseasonal order \((p_1, q_1)\).

3. Use the nonseasonal order \((p_1, q_1)\) to fit the \(SMA(1)\) model using maximum likelihood method and get new seasonal moving average coefficient \(\hat{\Theta}_1\).

4. Repeat step 2 to step 3 to get corresponding \(\hat{\Theta}_i\) and \((p_i, q_i)\) until both \(p_i\) and \(q_i\) converge.

\textbf{Remark.} The validity of using relative large nonseasonal \((p_0, q_0)\) to start the iteration is based on the fact that the possible overfitting of nonseasonal orders would lead to little impact on the estimation of the seasonal coefficient(s).

\textbf{Remark.} Step 2 involves the process to identify the AR and MA orders from the EACF table which we implement by Algorithm 1 and 2. See Preliminary section for the details of this algorithm.

\(^3\)They are still less than the known seasonal moving average order \(s\). We can start from \(p_0 = q_0 = s - 1\) though it is highly unnecessary. \(p_0 = q_0 = s/2\) might be a good choice.
Remark. The filtering $Y_t = (1 + \hat{\Theta}_0 B^s)^{-1}X_t$ can be exercised by

$$Y_t = \frac{1}{1 + \hat{\Theta}_0 B^s}X_t = 1 - \hat{\Theta} B^s X_t + \cdots + (-1)^{n} \hat{\Theta}^n B^{ns} X_t + \cdots$$

$$= (1 - \hat{\Theta} B^s + \cdots + (-1)^{c} \hat{\Theta}^c B^{cs}) X_t$$

$$= X_t - \hat{\Theta} X_{t-s} + \cdots + (-1)^{c} \hat{\Theta}^c X_{t-cs},$$

where $c$ is the cut-off point such that $\hat{\Theta}^c$ is relatively small.

Remark. With reasonable anticipation, if the seasonal model we assume is correct, the algorithm should converge in several steps; if the seasonal model is incorrect, the algorithm hardly converges. Thus we set up a reasonable maximum iteration times to stop the algorithm and the computation cost dramatically decreased compared to the previous method.

Remark. Using the iterative method introduced above, we can directly get the nonseasonal orders corresponding to the seasonal pattern. However, it is better for us to keep the last step EACF table rather than the orders specified by it since the table contains more information than order pair itself.

It is straightforward to extend our method to $SMA(2)$ model from $SMA(1)$ model: instead of having $\hat{\Theta}$, we right now estimate $\Theta_1$ and $\Theta_2$ and the transformation becomes $Y_t = (1 + \hat{\Theta}_1 B^s + \hat{\Theta}_2 B^{2s})^{-1}X_t$, with some additional computational complexity in filtering.

Assume SARMA as the seasonal model

A more detailed representation of equation (1.21) for $SARMA(1, 1)$ model is

$$(1 - \Phi B^s)(1 - \phi_1 B - \cdots - \phi_p B^p)X_t = (1 + \Theta B^s)(1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t,$$

(3.7)
where the seasonal lag is $s$, the seasonal AR parameter is $\Phi$, the seasonal MA parameter is $\Theta$, the nonseasonal AR parameters are $\phi_1, \cdots, \phi_p$ and the nonseasonal MA parameters are $\theta_1, \cdots, \theta_q$. We can reasonably assume $p < s$ and $q < s$.

We only need to slightly change the iterative method introduced for $SMA(1)$ model to suit the $SARMA(1, 1)$ model. The new algorithm goes as follows.

1. Start from a relative large nonseasonal orders $(p_0, q_0)$. Fit the $SARMA(1, 1)$ model using maximum likelihood method and get $\hat{\Theta}_0$ as the seasonal moving average coefficient and $\hat{\Phi}_0$ as the seasonal autoregressive coefficient.

2. Filter the original series by Tyler expansion of $Y_t = \frac{1 + \hat{\Phi}_0 B^s}{1 + \hat{\Theta}_0 B^s} X_t$ and analyze the EACF table on $Y_t$ to get the updated tentative nonseasonal order $(p_1, q_1)$.

3. Use the nonseasonal order $(p_1, q_1)$ to fit the $SARMA(1, 1)$ model using maximum likelihood method and get $\hat{\Theta}_1$ as the seasonal moving average coefficient and $\hat{\Phi}_1$ as the seasonal autoregressive coefficient.

4. Repeat step 2 to step 3 to get corresponding $\hat{\Theta}_i, \hat{\Phi}_i$ and $(p_i, q_i)$ until both $p_i$ and $q_i$ converge.

Remark. The filtering $Y_t = \frac{1 + \hat{\Phi}_0 B^s}{1 + \hat{\Theta}_0 B^s} X_t$ can be implemented by two steps we have already been familiar with: (1)$Z_t = (1 + \hat{\Theta}_0 B^s)^{-1} X_t$; (2)$Y_t = (1 + \hat{\Phi}_0 B^s)Z_t$.

Remark. Although we can definitely assume higher order seasonal models and use same algorithms to separate the seasonal impact, we do hold the belief that the summation of the seasonal orders usually does not exceed 2.

3.2.2 Automatic Order Identification Strategy
Till this stage, for each time series (as an input), we have five competitive models corresponding to five seasonal model assumptions. In order to reflect our concerns of model accuracy and model parsimony in selecting winner, a penalty score system should be set up, taking consideration of the following three concerns:

1. It should penalize messy EACF table and reward clean EACF table;
2. it should penalize large orders and reward small orders no matter seasonal or nonseasonal;
3. it should give suitable weights in selecting winners to seasonal part and nonseasonal part.

We investigate those three aspects in details one by one.

**Seasonal Part Comparison**

This one is trivial. Assume other conditions to be the same, seasonal models with less orders are preferred: $SMA(1)$ is better than $SMA(2)$; $SAR(1)$ is better than $SAR(2)$; $SAR(2), SMA(2)$ and $SARMA(1, 1)$ are equivalent.

**Non-Seasonal Part Comparison**

One straightforward method similar to what we operate in seasonal part is:

1. Use the Algorithm 1 and/or Algorithm 2 to identify the nonseasonal order $(p, q)$ from the EACF table (Automatically done for $SMA$ and $SARMA$ models).

2. Assume other conditions to be the same, nonseasonal models with less orders are preferred.
Nevertheless, in order not losing much information from the EACF table, we can instead using the following continuous penalty function

\[ S \triangleq \sum_{i=1}^{n} \sum_{j=i}^{m} (i + j)^{-0.5} \times EACF(i,j) \] (3.8)

where \( EACF(i,j) \) is the \((i,j)^{th}\) entry of the EACF table, \(1 \leq i \leq n, i \leq j \leq m\).\(^4\) We give more weights to entries close to origin and less weights to the counterparts; we penalize messy entries (with significant symbol 1) and reward clean entries (with insignificant symbol 0). The model with smaller penalty \( S \) is preferred then.

The advantage of this method is that it avoids any errors introduced by the automatic Algorithm 1 and/or Algorithm 2 and use all the information generated by the EACF table.

**Balance between Seasonal Comparison and Non-seasonal Comparison**

One more concern is the balance between the seasonal penalty and nonseasonal penalty. For example, one model has smaller seasonal orders while another model has smaller nonseasonal penalty score. Our experience suggests the following trade-off:

1. If the nonseasonal penalty score (3.8) is different, we pick up the model with smaller score no matter what is the seasonal model.

2. If in previous step, multiple models share the same nonseasonal score, we choose the model with smaller seasonal orders.

We incorporate the identification and penalize methods introduced in the

\(^{4}\)Due to the property of EACF table, we are only interested in upper right zero triangle.
above two subsections into the Automatic Seasonal Nonseasonal Order Identification System (ASNOIS). Two key advantages of the ASNOIS are

**Simultaneity** in the sense that it would provide both seasonal orders and non-seasonal orders at the same time as a function of the time series investigated.

**Automation** in the sense that it does not need any human decision or eyeball check during the process.

### 3.3 Numerical Study

In this section we generate simulation results to demonstrate advantages of ASNOIS in identifying multiplicative seasonal model orders. For the following examples, we specify the seasonal and nonseasonal orders and coefficients, determine the mean equation, and based on which, generate time series of length 1000.

At the first place, we show the ASNOIS outputs using a single time series sequence as an input. Secondly, we test the ASNOIS accuracy by listing the results of ASNOIS for 100 replications of time series generated with same parameter set. Intuitively, the more frequently the ASNOIS identifies the correct model, the better the performance.

We discuss the outputs of the first example in details and briefly list results for another two examples.

**Remark.** *The outputs of ASNOIS are composed of two parts. The first part lists results for five different seasonal models: the seasonal model assumption, the EACF table for the nonseasonal series after filtering the seasonal affect, the nonseasonal orders gained from the EACF table using Algorithm 1 or 2 and the*
<table>
<thead>
<tr>
<th>If we assume SAR(1)</th>
<th>If we assume SMA(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ar1</strong></td>
<td><strong>ar2</strong></td>
</tr>
<tr>
<td>ma1</td>
<td>0</td>
</tr>
<tr>
<td>ma2</td>
<td>2</td>
</tr>
<tr>
<td>ma3</td>
<td>2</td>
</tr>
<tr>
<td>ma4</td>
<td>2</td>
</tr>
<tr>
<td>ma5</td>
<td>2</td>
</tr>
<tr>
<td>ma6</td>
<td>2</td>
</tr>
<tr>
<td>ma7</td>
<td>0</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is \((0, 0)\). The non-seasonal ARMA model is \((1, 0)\).

The running time is 0 minutes 8 seconds. The running time is 0 minutes 12 seconds.

<table>
<thead>
<tr>
<th>If we assume SAR(2)</th>
<th>If we assume SMA(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ar1</strong></td>
<td><strong>ar2</strong></td>
</tr>
<tr>
<td>ma1</td>
<td>0</td>
</tr>
<tr>
<td>ma2</td>
<td>0</td>
</tr>
<tr>
<td>ma3</td>
<td>1</td>
</tr>
<tr>
<td>ma4</td>
<td>2</td>
</tr>
<tr>
<td>ma5</td>
<td>2</td>
</tr>
<tr>
<td>ma6</td>
<td>2</td>
</tr>
<tr>
<td>ma7</td>
<td>2</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is \((0, 0)\). Reach maximum iteration number. Break!

The running time is 0 minutes 15 seconds. The running time is 1 minutes 25 seconds.

<table>
<thead>
<tr>
<th>If we assume SARMA(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ar1</strong></td>
</tr>
<tr>
<td>ma1</td>
</tr>
<tr>
<td>ma2</td>
</tr>
<tr>
<td>ma3</td>
</tr>
<tr>
<td>ma4</td>
</tr>
<tr>
<td>ma5</td>
</tr>
<tr>
<td>ma6</td>
</tr>
<tr>
<td>ma7</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is \((0, 0)\).

The running time is 0 minutes 20 seconds.

The best model is sar1.
computation time. The second part has only one line (last line), showing the best model determined by ASNOIS. In this example, the minimum penalty score occurs when choosing $SAR(1)$ or $SARMA(1, 1)$ model, and the former is preferable than the latter as it is parsimonious and thus the best model is $SAR(1) + ARMA(0, 0)$ which is exact the simulation assumption. Note that for $SMA(2)$ model, the iterative method does not converge and it returns a break information and move on.
Table 3.4: Accuracy Count for $\text{SAR}(1):(1 - 0.9B^7)X_t = \varepsilon_t$.

<table>
<thead>
<tr>
<th>sar1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
<th>sma1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>93</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sar1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
<th>sma2</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sarma11</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Remark.** The number in each entry tells the times out of 100 that ASNOIS identifies the simulated series with seasonal patterns as what is listed on the left top of the matrix and nonseasonal orders as row and column names corresponding to the entry. In this example, 93% of the time ASNOIS can identify the true model, indicating ASNOIS’s great performance on identifying orders.
Table 3.5: ASNOIS Output for SMA(1)+ARMA(1,1): \((1 - 0.9B)X_t = (1 - 0.3B^7)(1 + 0.3B)\varepsilon_t\).

<table>
<thead>
<tr>
<th>If we assume SAR(1)</th>
<th>If we assume SMA(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1 ar2 ar3 ar4 ar5 ar6 ar7</td>
<td>ar1 ar2 ar3 ar4 ar5 ar6 ar7</td>
</tr>
<tr>
<td>ma1 2 2 2 2 2 2 2</td>
<td>ma1 2 2 2 2 2 2 2</td>
</tr>
<tr>
<td>ma2 2 0 2 2 2 1 0</td>
<td>ma2 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>ma3 2 2 0 0 2 2 1</td>
<td>ma3 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>ma4 2 2 0 0 2 2 2</td>
<td>ma4 2 2 0 0 0 0 0</td>
</tr>
<tr>
<td>ma5 2 2 0 0 2 2 2</td>
<td>ma5 2 2 0 2 0 0 0</td>
</tr>
<tr>
<td>ma6 2 2 2 0 2 2 0</td>
<td>ma6 2 2 2 0 0 0 0</td>
</tr>
<tr>
<td>ma7 2 2 2 2 1 1 2</td>
<td>ma7 2 2 2 0 0 2 0</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is (6, 6). The non-seasonal ARMA model is (1, 1).

The running time is 0 minutes 13 seconds. The running time is 0 minutes 30 seconds.

<table>
<thead>
<tr>
<th>If we assume SAR(2)</th>
<th>If we assume SMA(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1 ar2 ar3 ar4 ar5 ar6 ar7</td>
<td>ar1 ar2 ar3 ar4 ar5 ar6 ar7</td>
</tr>
<tr>
<td>ma1 2 2 2 2 2 2 2</td>
<td>ma1 2 2 2 2 2 2 2</td>
</tr>
<tr>
<td>ma2 2 0 2 0 2 2 0</td>
<td>ma2 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>ma3 0 2 2 0 0 0 2</td>
<td>ma3 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>ma4 0 2 2 2 2 1 2</td>
<td>ma4 2 0 2 0 0 0 0</td>
</tr>
<tr>
<td>ma5 2 0 2 2 1 2 1</td>
<td>ma5 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>ma6 2 0 0 2 2 0 2</td>
<td>ma6 2 2 2 0 0 0 0</td>
</tr>
<tr>
<td>ma7 2 2 2 2 1 1 2</td>
<td>ma7 2 2 2 0 0 2 0</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is (6, 6). The non-seasonal ARMA model is (1, 1).

The running time is 0 minutes 25 seconds. The running time is 0 minutes 38 seconds.

<table>
<thead>
<tr>
<th>If we assume SARMA(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1 ar2 ar3 ar4 ar5 ar6 ar7</td>
</tr>
<tr>
<td>ma1 2 2 2 2 2 2 2</td>
</tr>
<tr>
<td>ma2 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>ma3 2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>ma4 2 2 0 0 0 0 0</td>
</tr>
<tr>
<td>ma5 0 2 0 2 0 0 0</td>
</tr>
<tr>
<td>ma6 0 2 2 0 0 0 0</td>
</tr>
<tr>
<td>ma7 2 0 2 2 0 2 0</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is (1, 1).

The running time is 0 minutes 33 seconds.

The best model is sma1.
Table 3.6: Accuracy Count for SMA(1)+ARMA(1,1): \((1 - 0.9B)X_t = (1 - 0.3B^7)(1 + 0.3B)\varepsilon_t\).

<table>
<thead>
<tr>
<th>sar1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
<th>sma1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar2</td>
<td>0</td>
<td>74</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sar1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
<th>sma2</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar2</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sarma1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3.7: ASNOIS Output for SARMA(1,1)+ARMA(1,2): \((1 - 0.5B^7)(1 - 0.9B)X_t = (1 + 0.5B^7)(1 - 0.2B + 0.3B^2)\varepsilon_t\).

<table>
<thead>
<tr>
<th>If we assume SAR(1)</th>
<th>If we assume SMA(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>ar2</td>
</tr>
<tr>
<td>ma1</td>
<td>2</td>
</tr>
<tr>
<td>ma2</td>
<td>2</td>
</tr>
<tr>
<td>ma3</td>
<td>2</td>
</tr>
<tr>
<td>ma4</td>
<td>2</td>
</tr>
<tr>
<td>ma5</td>
<td>2</td>
</tr>
<tr>
<td>ma6</td>
<td>2</td>
</tr>
<tr>
<td>ma7</td>
<td>2</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is \((6, 6)\). Reach maximum iteration number. Break!
The running time is 0 minutes 14 seconds. The running time is 1 minutes 17 seconds.

<table>
<thead>
<tr>
<th>If we assume SAR(2)</th>
<th>If we assume SMA(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>ar2</td>
</tr>
<tr>
<td>ma1</td>
<td>2</td>
</tr>
<tr>
<td>ma2</td>
<td>2</td>
</tr>
<tr>
<td>ma3</td>
<td>2</td>
</tr>
<tr>
<td>ma4</td>
<td>2</td>
</tr>
<tr>
<td>ma5</td>
<td>1</td>
</tr>
<tr>
<td>ma6</td>
<td>2</td>
</tr>
<tr>
<td>ma7</td>
<td>2</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is \((6, 6)\). Reach maximum iteration number. Break!
The running time is 0 minutes 25 seconds. The running time is 3 minutes 32 seconds.

If we assume SARMA(1,1)

<table>
<thead>
<tr>
<th>ar1</th>
<th>ar2</th>
<th>ar3</th>
<th>ar4</th>
<th>ar5</th>
<th>ar6</th>
<th>ar7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ma1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>ma2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ma3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ma4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ma5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>ma6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ma7</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

The non-seasonal ARMA model is \((1, 2)\).
The running time is 0 minutes 30 seconds.

The best model is sarma11.
Table 3.8: Accuracy Count for SARMA$(1,1)+$ARMA$(1,2)$: 
\[(1 - 0.5B^7)(1 - 0.9B)X_t = (1 + 0.5B^7)(1 - 0.2B + 0.3B^2)\varepsilon_t.\]

<table>
<thead>
<tr>
<th>sar1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
<th>sma1</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sar11</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
<th>ma5</th>
<th>ma6</th>
<th>ma7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar2</td>
<td>0</td>
<td>79</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ar7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
3.4 Proof

Definition 2. Uniform convergence in probability: \( Q_n(\theta) \) converges uniformly in probability to \( Q_0(\theta) \) means

\[
\sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{P} 0.
\] (3.9)

Lemma 5. (From Newey and McFadden[59].) If there is a function \( Q_0(\theta) \)^5 such that

1. \( Q_0(\theta) \) is uniquely maximized(minimized) at \( \theta_0 \);

2. The parameter space \( \Theta \) is compact;

3. \( Q_0(\theta) \) is continuous;

4. \( Q_n(\theta) \) converges uniformly in probability to \( Q_0(\theta) \),

and assume \( \hat{\theta} \) maximizes(minimizes) \( \hat{Q}_n(\theta) \) subject to \( \theta \in \Theta \), then we have

\[
\hat{\theta} \xrightarrow{P} \theta_0.
\] (3.10)

Theorem 5. If \( \hat{\beta}_1, \ldots, \hat{\beta}_p \) and \( \hat{\beta}_s \) are consistent estimators such that (3.4) holds, then the nonlinear least square estimators obtained from (3.5) are also consistent.

Proof: Based on the fact that \( \hat{\beta}_1, \ldots, \hat{\beta}_p \) and \( \hat{\beta}_s \) are consistent estimators of \( \phi_1, \ldots, \phi_p \) and \( \Phi \), in order to show that \( \hat{\phi}_1, \ldots, \hat{\phi}_p \) and \( \hat{\Phi} \) are also consistent, we

---

5"\theta" could be a vector here.
set \( \theta = (\theta_1, \ldots, \theta_p, \theta_s) \) and

\[
Q_0(\theta) = \sum_{i=1}^{p} (\theta_i - \phi_i)^2 + (\theta_s - \Phi)^2 + \sum_{i=1}^{p} (-\theta_i \theta_s + \phi_i \Phi)^2;
\]  

(3.11)

\[
Q_n(\theta) = \sum_{i=1}^{p} (\theta_i - \hat{\beta}_i)^2 + (\theta_s - \hat{\beta}_s)^2 + \sum_{i=1}^{p} (-\theta_i \theta_s - \hat{\beta}_{s+i})^2.
\]  

(3.12)

Since \( \hat{\phi}_1, \ldots, \hat{\phi}_p \) and \( \hat{\Phi} \) minimizes \( \hat{Q}_n(\theta) \) by (3.5), we only need to verify the four conditions of Lemma 5 to prove consistency of the least square estimators.

For condition (1), since \( Q_0(\theta) \) is the summation of several quadratic functions, the minimum parameter set is unique, and then \( \theta_0 = (\theta_1, \ldots, \theta_p, \theta_s) = (\phi_1, \ldots, \phi_p, \Phi) \).

Conditions (2) and (3) are trivial.

For condition (4), since \( \hat{\beta}_1, \ldots, \hat{\beta}_p \) and \( \hat{\beta}_s \) are consistent estimators of \( \phi_1, \ldots, \phi_p \) and \( \Phi \), we have \( \sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{P} 0 \), and thus \( Q_n(\theta) \) converges uniformly in probability to \( Q_0(\theta) \).

As a result, by Lemma 5, we have

\[
\hat{\theta} = (\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\Phi}) \xrightarrow{P} \theta_0 = (\phi_1, \ldots, \phi_p, \Phi).
\]

We conclude that \( \hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\Phi} \) are also the consistent estimators of \( \phi_1, \ldots, \phi_p, \Phi \).
Chapter 4
End-of-Day Stock Trading Volume Prediction with a Two-Component Hierarchical Model

4.1 Introduction

Both human traders and algorithmic trading engine designers have a profound interest in high quality prediction of the volume that will be traded in the remainder of the trading day. This volume represents the liquidity against which orders can be transacted while the available liquidity determines the market impact of working any order. Indeed, with inadequate liquidity, it may not be feasible or it may be too expensive to execute a large order in the remainder of the trading day. The commonly used guaranteed execution algorithms such as the Initiation Price face the challenge of working parent orders across the day. To minimize market impact, the algorithm must keep its volume participation rate as low as possible. On the other hand, the algorithm must also ensure that its volume participation rate is high enough in order to complete the order within the day. Thus, it is crucial to have an accurate prediction of the volume in the remainder of the day to effectively execute such algorithms.

Remainder day’s volume prediction also plays an important role in analyzing transaction cost. Berkowitz(1988)[63] introduced the concept of daily Volume Weighted Average Price (VWAP) and used the difference between the average execution price and the recorded VWAP to measure the cost of each trade. Minimizing such cost is among critical goals for institutional investors and having
better knowledge of reminder day’s volume would definitely help to make efficient execution decision and thus favorable cost.

Recently, both academia and practitioners have been pursuing better models forecasting the trading volumes. Lo and Wang(2000)[52] analyzed behavior of equity trading volume using Capital Asset Pricing Model (CAPM). Hautsch(2002)[42] modeled the intra-day volume activity based on volume durations using Autoregressive Conditional Duration (ACD) models, which was originally introduced by Engle and Russell(1998)[27]. Darolles and Le Fol(2003)[18] proposed a methodology of decomposing trading volume and Bialkowski, Darolles and Le Fol(2008)[5] extended the previous work into intra-day data, decomposing intra-day trading volume into two components: one reflects volume change associated with market evolutions and the other represents stock specific volume pattern. It used historical VWAP curve to estimate market component and ARMA and SETAR models to estimate stock specific component.

Although our aim is to predict the volume to be traded in the reminder of the day, we instead investigate the total volume accrued throughout the day, or the end-of-day volume, for simplicity. Given the fact that the volume that has been accumulated from the beginning of the day to the time of prediction is known to us, the two volumes are equivalent.

Intuitively, there are two sources of useful information for projecting the end of day volume. Since we have observed the partial volume observed up to the time of prediction, and if the distribution of total volume throughout the day is relatively stable hence can be estimated using historical data, then the total end-of-day volume prediction can be made with the partial volume and the estimated proportion it should assume for the total volume. We call this method *intra-day* prediction. The second source of information is the dynamics of daily volume changing over time. If such dynamics is properly modeled, daily volume can be
predicted as well. We denote this method as daily prediction.

Since the intra-day method utilizes the volume accrued during the trading day while the daily method uses daily volume series, these two methods provide independent predictions. They can be improved by combining both sources of information. In this paper we propose a hierarchical model for such a combination. It extends to the stable seasonal pattern model of Chen & Fomby(1999)[14] where the model was used to predict end-of-year total number of tourists. Similar idea was used in Oliver(1987)[61] as well. This approach is different from that of Bialkowski, Darolles & Le Fol(2008)[18] who decomposed the trading volume into two components that reflect volume changes due to market evolutions and the stock specific volume pattern.

This chapter is organized as follows: in Section 4.2, the two component hierarchical model is presented. Its prediction procedures and some extensions are shown in Section 4.3. Section 4.4 shows an empirical study using Dow Jones Industrial Average component stocks, comparing out-of-sample prediction performance of different methods.

### 4.2 Two Component Hierarchical Model

Consider a trading volume series \( \cdots, x_{t1}, x_{t2}, \cdots, x_{td}, \cdots \) where \( t = 1, 2, \cdots, n \) corresponds to different trading days and \( d \) denotes the number of trading periods used per day. In this paper we use \( d = 13 \), corresponding in total 13 30-minute trading intervals of US equity market. The end-of-day volume for day \( t \) is then given by \( y_t = x_{t1} + \cdots + x_{td} \).

We assume that the daily pattern is stable across different days, i.e., given the end-of-day volume total \( y_t, (x_{t1}, \cdots, x_{td}) \) are conditionally independent for
Namely

\[ p[x_{11}, \cdots, x_{1d}, \cdots, x_{n1}, \cdots, x_{nd} | y_1, \cdots, y_n] = \prod_{t=1}^{n} p[x_{t1}, \cdots, x_{td} | y_t] \]  \hspace{1cm} (4.1)

where \( p[\cdot | \cdot] \) denotes the conditional density.

This assumption allows us to model the series \( \cdots, x_{t1}, x_{t2}, \cdots, x_{td}, \cdots \) with a two-component hierarchical structure,

\[ p[x_{t1}, \cdots, x_{td} | y_t] \sim M_1(y_t, \theta) \quad \text{and} \quad \{y_t\} \sim M_2(y_{t-1}, \ldots, \beta), \]  \hspace{1cm} (4.2)

where the intra-day model \( M_1 \) is a \((d-1)\)-dimensional distribution and the daily model \( M_2 \) is a time series model for the daily volume series \( \{y_t\} \). \( \theta \) and \( \beta \) are parameters to be estimated.

There are several important characteristics associated with this hierarchical model. First, it is assumed that the dependency between different days is only through the total volume \( \{y_t\} \), not the individual observations within the days. Second, given the total volume, the intra-day distribution of the total volume throughout the day is the same. Third, the two models, one for intra-day distribution and the other for daily volume series, can be modeled separately. Those three properties are ideal in terms of making intra-day prediction of the end-of-day volume.

In the following steps, we discuss the possible models for \( M_1 \) and \( M_2 \). Since \( \{x_{tk}\} \) is integer-valued, one immediate choice for model \( M_1 \) is the multinomial distribution, with the total count being the daily total volume. However, trading volume is often large enough to be treated as a continuous variable. Treating it as a continuous variable also provides an advantage of more flexibility in the modeling of total volume series \( y_t \). Nevertheless, we still want its distribution to possess the proportional interpretation of the multinomial distribution. Chen
& Fomby(1999)[14] proposed a continuous analogue of the multinomial distribution called the Gaussian multinomial distribution. Specifically, a $d$–dimensional random variable $(x_{t1}, \cdots, x_{td})$ is said to follow Gaussian multinomial (G-MN) distribution $(y_t, \theta_1, \cdots, \theta_d, \sigma^2)$ if

$$p[x_{t1}, \cdots, x_{t(d-1)}] \sim N(\mu, \sigma^2\sum) \quad \text{and} \quad x_{td} = y_t - \sum_{i=1}^{d-1} x_{ti},$$

(4.3)

where $0 < \theta_i < 1, \sum_{i=1}^{d} \theta_i = 1$ and

$$\mu = \begin{pmatrix} \theta_1 y_t \\ \theta_2 y_t \\ \vdots \\ \theta_{d-1} y_t \end{pmatrix}, \quad \sum = \begin{pmatrix} \theta_1(1 - \theta_1) & -\theta_1 \theta_2 & \cdots & -\theta_1 \theta_{d-1} \\ -\theta_1 \theta_2 & \theta_2(1 - \theta_2) & \cdots & \theta_2 \theta_{d-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\theta_1 \theta_{d-1} & -\theta_2 \theta_{d-1} & \cdots & \theta_{d-1}(1 - \theta_{d-1}) \end{pmatrix}.$$

Note that $\mu$ and $\sum$ do not depend on $t$ due to our assumption of stable daily pattern.

This distribution can be viewed as a continuous version of the multinomial distribution. The mean and correlation coefficient matrix are the same as the multinomial distribution, except for the extra variance parameter $\sigma^2$. In the multinomial distribution, the variance varies with the total volume in a relatively restricted way. In the Gaussian multinomial distribution, the variance is constant. It is possible to allow for varying variance (depending on observable variables) similar to the weighted regression setting or a GARCH type of heteroscedasticity. Here we choose to assume constant variance.

It can be shown that the Gaussian multinomial distribution also has the combination property of the multinomial distribution. For example, if $(x_{t1}, \cdots, x_{td}) \sim G-MN(y, \theta_1, \cdots, \theta_d, \sigma^2)$, then $(x_{t1} + x_{t2}, x_{t3}, \cdots, x_{td}) \sim G-MN(y, \theta_1 + \theta_2, \theta_3 \cdots, \theta_d, \sigma^2)$. 
More critically, this property implies that for $1 \leq k < d$,

$$
\sum_{i=1}^{k} x_{ti} \sim N(\gamma_k y_t, \sigma^2 \gamma_k (1 - \gamma_k)), \quad \text{where} \quad \gamma_k = \sum_{i=1}^{k} \theta_i.
$$

This feature allows us to estimate $\gamma_k$ much more easily than estimating $\theta_1, \cdots, \theta_k$ in high dimensional space. The parameter $\gamma_k$ is of critical interest as our prediction procedure involves only $\gamma_k$, instead of the individual parameters $\theta_1, \cdots, \theta_k$. The maximum likelihood estimator of $\gamma_k$ in our setting is

$$
\hat{\gamma}_k = \frac{\sum_{t=1}^{n} (\sum_{i=1}^{k} x_{ti}) y_t}{\sum_{t=1}^{n} y_t^2}.
$$

(4.4)

with standard error $\hat{\sigma} [\hat{\gamma}_k (1 - \hat{\gamma}_k)]^{1/2} / \sum y_t^2$. The maximum likelihood estimator for $\sigma^2$ is

$$
\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} (x_t - \hat{\theta} y_t)' \sum_{t=1}^{-1} (x_t - \hat{\theta} y_t)}{n}.
$$

To model the daily volume dynamics, we use a Gaussian model where

$$
p[y_t | y_{t-1}] \sim N(\mu_t, \sigma_t^2),
$$

(4.5)

where $y_{t-1} = (y_{t-1}, \cdots, y_1)$, $\mu_t = f(y_{t-1}, \theta)$, and $\sigma_t = g(y_{t-1}, \theta)$. This model includes all standard ARIMA models of Box and Jenkins(1976)[8] and Brockwell and Davis(1986)[11], with or without GARCH errors (Bollerslev(1986)[7]).

### 4.3 Prediction based on the two-component hierarchical models

At a given time $k$ of the trading day $t$, we are interested in estimating the end of day volume $y_t$ using volume accumulated up to time $k$, namely $x_{t1}, \cdots, x_{tk}$ and the
historical daily volume series \( y_{t-1}, \ldots, y_1 \). Specifically, let \( x_t^{(k)} = (x_{t1}, \ldots, x_{tk}) \), \( \theta = (\theta_1, \ldots, \theta_k)' \) and \( \sum_k = (v_{ij})_{k \times k} \), where \( v_{ij} = -\theta_i \theta_j \) for \( i \neq j \) and \( v_{ii} = \theta_i (1 - \theta_i) \) and \( i, j \leq k \), then

\[
p[y_t|x_t,\ldots,x_{tk},y_{t-1}] \propto p[x_t,\ldots,x_{tk}|y_t]p[y_t|y_{t-1}] \sim N(\mu_{t,k},\sigma^2_{t,k}),
\]

where

\[
\sigma^2_{t,k} = \left( \frac{\theta_k' \sum_k^{-1} \theta_k}{\sigma^2} + \frac{1}{\sigma_t^2} \right)^{-1} = \left( \frac{\gamma_k}{(1 - \gamma_k)\sigma^2} + \frac{1}{\sigma_t^2} \right)^{-1}
\]

and

\[
\mu_{t,k} = \left( \frac{x_t^{(k)} \sum_k^{-1} \theta_k}{\sigma^2} + \frac{\mu_t}{\sigma_t^2} \right) \sigma^2_{t,k} = \left( \frac{\sum_{i=1}^{k} x_{ti}}{(1 - \gamma_k)\sigma^2} + \frac{\mu_t}{\sigma_t^2} \right) \sigma^2_{t,k}.
\]

Hence, the least squares prediction of \( y_t \) is \( \mu_{t,k} \). In fact, letting \( c_t = \sigma^2/\sigma_t^2 \), (4.6) can be written as

\[
\mu_{t,k} = \frac{1}{\gamma_k + (1 - \gamma_k)c_t} \sum_{i=1}^{k} x_{ti} + \frac{c_t}{\gamma_k + (1 - \gamma_k)c_t} (1 - \gamma_k)\mu_t
\]

\[
\Delta = w_{tk} \frac{\sum_{i=1}^{k} x_{ti}}{\gamma_k} + (1 - w_{tk})\mu_t.
\]

Here, \( \sum_{i=1}^{k} x_{ti} \) and \( \mu_t \) are the predictions of the end of day volume based on \( M_1 \) model alone and \( M_2 \) model alone, respectively. The weights \( w_{tk} \) and \( 1 - w_{tk} \) dictate the contribution of models \( M_1 \) and \( M_2 \) to the prediction of the total volume.

Note that, when the model for the total volume series and the model for volumes within the day have equal precision, that is, \( \sigma^2 = \sigma_t^2 \), then \( u_{t,k} = \sum_{i=1}^{k} x_{ti} + (1 - \gamma_k)\mu_t = \sum_{i=1}^{k} x_{ti} + \sum_{i=k+1}^{d} \theta_i \mu_t \). Hence, the adjustment procedure is simply to replace the predicted contribution of the individual observations \( \theta_i \mu_t \)
by their observed values $x_{ti}$. When $c < 1$, $\sum_{i=1}^{k} x_{ti}$ bears more weight in prediction, because the individual volume observation has less variance (more accuracy) than the daily volume model.

An alternative is to treat the combination weights $w_{tk}$ and $1 - w_{tk}$ in (4.7) as unknown parameters. If one assumes homoscedasticity in the daily volume series, the weights are constant over time. They can be estimated with least squares as if the daily total volume follows a regression model with $\sum_{i=1}^{k} x_{ti}$ and $\mu_t$ as explanatory variables. This model has a nonparametric modeling flavor as it bypasses the assumptions of $M_1$ and $M_2$ and optimize the linear combination directly. We call this model the ‘regression approach’. One can further extend this model by including other informative variables in the combination, such as market trading volume up to period $k$.

### 4.4 Empirical Study

In this section, we apply the two component hierarchical model to historical volume profile of the 30 selected stocks in the Dow Jones Industrial Average from January 2010 to September 2010. Detailed results are shown using the volume profile of Apple Inc.$^1$. There are in total 185 trading days in the study. We use the first 155 days for modeling and parameter estimation and the last 30 days for out-of-sample prediction performance comparison. Figure 4.1 shows the volume series of Apple Inc.

To avoid dealing with micro-structure noise and for the ease of computation, we aggregate the minute-by-minute volume data to 30-minute intervals, with 13 periods per day. Here we use 1:00pm as our time for the end-of-day volume prediction, that is, we have observed the trading volume up to 1:00pm (thus $k = 7$)

---

$^1$Arbitrarily chosen, not included in DJIA though.
Figure 4.1: Apple Inc. volume series: the volumes to the left of vertical dash line are used in fitting while the right hand side volumes are used for prediction and testing purpose.

and want to estimate the end-of-day volume using the accumulated volume that day and the historical daily volumes before that day.

We use the Gaussian multinomial distribution for Model $M_1$. Figure 4.2 depicts the estimated $\theta_i, 1 \leq i \leq 13$ for Apple Inc., the intra-day distribution of the total volume into the 13 intra-day periods. Regularity of the pattern for different stocks is also demonstrated and verified. Furthermore, Table 4.1 shows the estimates and standard errors of $\gamma_k = \sum_{i=1}^{k} \theta_k$.

Table 4.1: Maximum likelihood estimates (standard errors) of the gamma.

<table>
<thead>
<tr>
<th></th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
<th>$\gamma_5$</th>
<th>$\gamma_6$</th>
<th>$\gamma_7$</th>
<th>$\gamma_8$</th>
<th>$\gamma_9$</th>
<th>$\gamma_{10}$</th>
<th>$\gamma_{11}$</th>
<th>$\gamma_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.165</td>
<td>0.280</td>
<td>0.363</td>
<td>0.434</td>
<td>0.490</td>
<td>0.537</td>
<td>0.580</td>
<td>0.630</td>
<td>0.681</td>
<td>0.741</td>
<td>0.813</td>
<td>0.888</td>
</tr>
<tr>
<td>Std. err.</td>
<td>0.011</td>
<td>0.013</td>
<td>0.014</td>
<td>0.015</td>
<td>0.015</td>
<td>0.015</td>
<td>0.015</td>
<td>0.015</td>
<td>0.014</td>
<td>0.013</td>
<td>0.012</td>
<td>0.010</td>
</tr>
</tbody>
</table>

We fit the historical daily volume series using Gaussian ARMA models and
Figure 4.2: Intraday volume distribution of different stocks: x-axis corresponds to the 13 trading periods and y-axis represents the volume percentage $\theta$. Volumes would accumulate to 100%.

ARMA-GARCH models. The autocorrelation function (ACF) and partial autocorrelation function (PACF) of Apple Inc.’s daily volume series are shown in Figure 4.3. They appear to be quite stationary. Further study using the extended autocorrelation function of Tsay and Tiao(1984)[74] and the adjusted extended autocorrelation function we discussed in previous chapter (also see Chen, Min and Chen(2010)[15]) and the AIC criterion of Akaike(1974)[1] identifies an AR(2) model as an appropriate model for the daily volume series of Apple Inc. To include heteroscedasticity, we also model the error term of the AR(2) model as a GARCH(2,0) process, after carrying out a model selection procedure. Table 4.2 shows the estimated parameters for AR(2) and AR(2)-GARCH(2,0) model.

The GARCH component is marginally significant. We also fit other volume series individually and identify the best ARMA and ARMA-GARCH models. They are used as Model $M_2$ in the prediction exercises.
Figure 4.3: ACF and PACF for Apple Inc.’s daily volume series.

Table 4.2: Fitting results for candidate ARMA models.

<table>
<thead>
<tr>
<th></th>
<th>AR(2)</th>
<th>AR(2)-GARCH(2,0)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>value</td>
<td>sd</td>
</tr>
<tr>
<td>intercept</td>
<td>0.9966</td>
<td>0.0703</td>
</tr>
<tr>
<td>ar1</td>
<td>0.4599</td>
<td>0.0714</td>
</tr>
<tr>
<td>ar2</td>
<td>0.2239</td>
<td>0.0714</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0943</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log likelihood</td>
<td>-44.35</td>
<td></td>
</tr>
</tbody>
</table>
Given the estimated parameters of $M_1$ and $M_2$, we compare following six different prediction methods.

- **arma** prediction using the daily volume series only, under ARMA model;

- **garch** prediction using the daily volume series only, under ARMA-GARCH model;

- **intraday** prediction using intraday model $M_1$ only;

- **new-arma** prediction using the new hierarchical model with ARMA as model $M_2$;

- **new-garch** prediction using the new hierarchical model with ARMA-GARCH as model $M_2$;

- **reg** prediction using combination (4.7) with least squares optimized weights.

We first show the detailed results of Apple Inc. Table 4.3 numerates the estimated variance $\sigma^2$ of the inra-day Gaussian-Multinomial distribution ($M_1$) and $\sigma_t^2$ for the ARMA model ($M_2$) and the corresponding weights $w_{tk}$ and $1 - w_{tk}$ in (4.7). Note that if we assume ARMA model for $M_2$, $c_t$ and $\omega_{tk}$ are indeed independent of time $t$. In this case, the prediction provided by the daily series is relatively more accurate (smaller variance) hence having a large weight in the combined prediction. The estimated weights under least square criterion is also presented and it tends to put relatively more weights on intra-day prediction which is not optimal in this case.

Table 4.4 shows the estimated variance $\sigma_t^2$ under ARMA-GARCH model for the 30 days in our prediction period and their corresponding ratio $c_t$, comparing to the estimated $\sigma^2$ listed in Table 4.3. It does change quite significantly, from 0.3 to almost 3.

Table 4.5 shows the actual prediction of the volume (in millions of shares) of the 30 day prediction period, under different methods. The true observation,
Table 4.3: Variance Comparison I.

<table>
<thead>
<tr>
<th>intra-day ( (\sigma^2) )</th>
<th>ARMA ( (\sigma^2) )</th>
<th>( w_{tk} )</th>
<th>( 1 - w_{tk} )</th>
<th>( \text{reg } w_1 )</th>
<th>( \text{reg } w_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.41E+13</td>
<td>4.22E+13</td>
<td>0.48</td>
<td>0.52</td>
<td>0.77</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Table 4.4: Variance Comparison II.

<table>
<thead>
<tr>
<th>( t )</th>
<th>( \sigma_i^2 )</th>
<th>( \sigma^2 / \sigma_i^2 )</th>
<th>( t )</th>
<th>( \sigma_i^2 )</th>
<th>( \sigma^2 / \sigma_i^2 )</th>
<th>( t )</th>
<th>( \sigma_i^2 )</th>
<th>( \sigma^2 / \sigma_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.32E+13</td>
<td>1.93</td>
<td>11</td>
<td>6.03E+13</td>
<td>1.06</td>
<td>21</td>
<td>3.57E+13</td>
<td>1.80</td>
</tr>
<tr>
<td>2</td>
<td>2.46E+13</td>
<td>2.60</td>
<td>12</td>
<td>4.71E+13</td>
<td>1.36</td>
<td>22</td>
<td>2.43E+13</td>
<td>2.64</td>
</tr>
<tr>
<td>3</td>
<td>2.99E+13</td>
<td>2.14</td>
<td>13</td>
<td>2.55E+13</td>
<td>2.51</td>
<td>23</td>
<td>2.86E+13</td>
<td>2.24</td>
</tr>
<tr>
<td>4</td>
<td>2.51E+13</td>
<td>2.56</td>
<td>14</td>
<td>3.84E+13</td>
<td>1.67</td>
<td>24</td>
<td>2.18E+13</td>
<td>2.94</td>
</tr>
<tr>
<td>5</td>
<td>3.22E+13</td>
<td>1.99</td>
<td>15</td>
<td>2.72E+13</td>
<td>2.36</td>
<td>25</td>
<td>3.04E+13</td>
<td>2.11</td>
</tr>
<tr>
<td>6</td>
<td>2.54E+13</td>
<td>2.53</td>
<td>16</td>
<td>2.75E+13</td>
<td>2.34</td>
<td>26</td>
<td>4.86E+13</td>
<td>1.32</td>
</tr>
<tr>
<td>7</td>
<td>3.43E+13</td>
<td>1.87</td>
<td>17</td>
<td>2.79E+13</td>
<td>2.30</td>
<td>27</td>
<td>2.70E+13</td>
<td>2.38</td>
</tr>
<tr>
<td>8</td>
<td>2.65E+13</td>
<td>2.42</td>
<td>18</td>
<td>2.45E+13</td>
<td>2.62</td>
<td>28</td>
<td>8.58E+13</td>
<td>0.75</td>
</tr>
<tr>
<td>9</td>
<td>3.95E+13</td>
<td>1.62</td>
<td>19</td>
<td>2.15E+13</td>
<td>2.99</td>
<td>29</td>
<td>1.77E+14</td>
<td>0.36</td>
</tr>
<tr>
<td>10</td>
<td>3.10E+13</td>
<td>2.07</td>
<td>20</td>
<td>2.52E+13</td>
<td>2.54</td>
<td>30</td>
<td>7.23E+13</td>
<td>0.89</td>
</tr>
</tbody>
</table>

labelled as real is shown at the bottom line for comparison. Figure 4.4 shows
the predictions of arma, garch, intra-day, reg, new-arma and new-garch, with the
true observations marked as dots. Day 28 is an unusual observation. The actual
volume is significantly larger than normal. The new prediction method is able
to capture such a large movement, while the daily model under-predicts and the
intra-day model over-predicts.

Table 4.6 summaries the prediction performance by showing the root mean
squared prediction error

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Pred_i - True_i)^2}. \]

In the following we present the prediction performance comparison for the 30
stocks in the Dow Jones Industrial Average Index. The first two columns of Table
4.7 provide the company information; the following two columns store the ARMA
and GARCH orders of the daily volume series. Mean squared prediction errors
of the six prediction methods are then listed, followed by the ratio comparison
Figure 4.4: Comparison of different prediction methods.

Table 4.5: Prediction results from various methods. (in millions of shares)

<table>
<thead>
<tr>
<th></th>
<th>day 1</th>
<th>day 2</th>
<th>day 3</th>
<th>day 4</th>
<th>day 5</th>
<th>day 6</th>
<th>day 7</th>
<th>day 8</th>
<th>day 9</th>
<th>day 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>arma</td>
<td>16.05</td>
<td>15.30</td>
<td>15.83</td>
<td>19.32</td>
<td>20.62</td>
<td>18.10</td>
<td>18.88</td>
<td>16.55</td>
<td>15.84</td>
<td>20.93</td>
</tr>
<tr>
<td>garch</td>
<td>15.78</td>
<td>14.67</td>
<td>15.48</td>
<td>19.22</td>
<td>20.15</td>
<td>17.30</td>
<td>18.64</td>
<td>15.74</td>
<td>15.47</td>
<td>21.10</td>
</tr>
<tr>
<td>intraday</td>
<td>15.68</td>
<td>12.39</td>
<td>13.00</td>
<td>22.41</td>
<td>20.13</td>
<td>15.70</td>
<td>21.65</td>
<td>13.38</td>
<td>12.79</td>
<td>20.65</td>
</tr>
<tr>
<td>newarma</td>
<td>15.87</td>
<td>13.89</td>
<td>14.46</td>
<td>20.82</td>
<td>20.38</td>
<td>16.94</td>
<td>20.22</td>
<td>15.01</td>
<td>14.36</td>
<td>20.80</td>
</tr>
<tr>
<td>real</td>
<td>14.38</td>
<td>11.88</td>
<td>13.87</td>
<td>19.94</td>
<td>20.31</td>
<td>15.27</td>
<td>18.62</td>
<td>12.88</td>
<td>13.64</td>
<td>23.27</td>
</tr>
<tr>
<td></td>
<td>day 11</td>
<td>day 12</td>
<td>day 13</td>
<td>day 14</td>
<td>day 15</td>
<td>day 16</td>
<td>day 17</td>
<td>day 18</td>
<td>day 19</td>
<td>day 20</td>
</tr>
<tr>
<td>arma</td>
<td>18.01</td>
<td>17.46</td>
<td>15.49</td>
<td>17.72</td>
<td>17.25</td>
<td>15.89</td>
<td>15.38</td>
<td>15.76</td>
<td>16.09</td>
<td>19.70</td>
</tr>
<tr>
<td>garch</td>
<td>17.00</td>
<td>17.20</td>
<td>14.88</td>
<td>17.70</td>
<td>16.69</td>
<td>15.49</td>
<td>15.03</td>
<td>15.47</td>
<td>15.75</td>
<td>19.69</td>
</tr>
<tr>
<td>newarma</td>
<td>16.30</td>
<td>17.47</td>
<td>14.38</td>
<td>18.04</td>
<td>15.62</td>
<td>15.17</td>
<td>14.77</td>
<td>15.31</td>
<td>14.89</td>
<td>17.77</td>
</tr>
<tr>
<td>newgarch</td>
<td>15.56</td>
<td>17.35</td>
<td>14.27</td>
<td>18.01</td>
<td>15.64</td>
<td>15.08</td>
<td>14.68</td>
<td>15.25</td>
<td>15.05</td>
<td>18.26</td>
</tr>
<tr>
<td></td>
<td>day 21</td>
<td>day 22</td>
<td>day 23</td>
<td>day 24</td>
<td>day 25</td>
<td>day 26</td>
<td>day 27</td>
<td>day 28</td>
<td>day 29</td>
<td>day 30</td>
</tr>
<tr>
<td>garch</td>
<td>18.77</td>
<td>21.06</td>
<td>21.62</td>
<td>19.79</td>
<td>23.64</td>
<td>21.63</td>
<td>17.94</td>
<td>27.65</td>
<td>20.83</td>
<td>20.54</td>
</tr>
<tr>
<td>intraday</td>
<td>20.36</td>
<td>22.81</td>
<td>17.79</td>
<td>21.29</td>
<td>25.20</td>
<td>22.55</td>
<td>16.19</td>
<td>47.18</td>
<td>15.47</td>
<td>25.91</td>
</tr>
<tr>
<td>newarma</td>
<td>19.86</td>
<td>22.00</td>
<td>20.02</td>
<td>20.83</td>
<td>24.44</td>
<td>22.47</td>
<td>17.41</td>
<td>36.89</td>
<td>18.83</td>
<td>23.19</td>
</tr>
<tr>
<td>reg</td>
<td>20.25</td>
<td>22.57</td>
<td>18.89</td>
<td>21.20</td>
<td>25.00</td>
<td>22.64</td>
<td>16.83</td>
<td>42.82</td>
<td>17.07</td>
<td>24.83</td>
</tr>
<tr>
<td>real</td>
<td>17.76</td>
<td>22.74</td>
<td>22.51</td>
<td>19.22</td>
<td>27.14</td>
<td>21.30</td>
<td>16.00</td>
<td>35.75</td>
<td>15.98</td>
<td>22.60</td>
</tr>
</tbody>
</table>
Table 4.6: Root mean square error from various methods.

<table>
<thead>
<tr>
<th></th>
<th>arma</th>
<th>garch</th>
<th>intra-day</th>
<th>new-arma</th>
<th>new-garch</th>
<th>reg</th>
<th>$\text{garch}_{\text{arma}}$</th>
<th>$\text{new-garch}_{\text{garch}}$</th>
<th>$\text{new-garch}_{\text{intra-day}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15.85</td>
<td>13.76</td>
<td>15.62</td>
<td>9.72</td>
<td>9.72</td>
<td>11.95</td>
<td>86.84%</td>
<td>70.69%</td>
<td>62.24%</td>
</tr>
</tbody>
</table>

statistics.

We make the following observations:

1. In most of the cases, the new two component hierarchical model performs much better than using the intra-day model alone and using the daily series dynamics alone. In more than half of the cases, the improvement is over 20%, some significantly higher.

2. Although the ARMA-GARCH model may be marginally better than ARMA model for modeling the daily volume series dynamics (in the Apple Inc. case), ARMA model works almost as well as ARMA-GARCH model in terms of prediction.

3. The alternative combination method with least squares optimized weights outperforms the intra-day model alone and the daily series alone. Its overall performance is not as good as the one based on the hierarchical model.

4. The phenomenon observed on the 28-th day of Apple Inc. volume series is quite important. It shows that, the hierarchical model can indeed effectively combine two independent sources of information and produce more accurate prediction. This phenomenon is also observed in other stocks as well.

5. The idea of including other factors such as market volume in the combination has been tried, without success. More research needs to be done to find more appropriate factors.

In conclusion, our empirical study shows that the proposed two component hierarchical model and its associated prediction method are effective in making
<table>
<thead>
<tr>
<th>company</th>
<th>symbol</th>
<th>arma orders</th>
<th>garch orders</th>
<th>arma</th>
<th>garch</th>
<th>intra day</th>
<th>new arma</th>
<th>new garch</th>
<th>reg</th>
<th>garch arma</th>
<th>new garch arma</th>
<th>new garch garch</th>
<th>new garch intra day</th>
</tr>
</thead>
<tbody>
<tr>
<td>3M</td>
<td>MMM</td>
<td>1 0</td>
<td>1 0</td>
<td>3.40</td>
<td>3.23</td>
<td>1.62</td>
<td>1.73</td>
<td>1.64</td>
<td>1.43</td>
<td>94.95%</td>
<td>95.05%</td>
<td>50.98%</td>
<td>101.80%</td>
</tr>
<tr>
<td>Alcoa</td>
<td>AA</td>
<td>1 0</td>
<td>2 0</td>
<td>28.33</td>
<td>21.58</td>
<td>12.45</td>
<td>11.75</td>
<td>10.55</td>
<td>10.90</td>
<td>76.18%</td>
<td>89.74%</td>
<td>48.87%</td>
<td>84.71%</td>
</tr>
<tr>
<td>American Express</td>
<td>AXP</td>
<td>1 0</td>
<td>1 0</td>
<td>5.93</td>
<td>4.65</td>
<td>5.57</td>
<td>3.96</td>
<td>3.73</td>
<td>4.96</td>
<td>78.42%</td>
<td>94.28%</td>
<td>80.19%</td>
<td>66.99%</td>
</tr>
<tr>
<td>ATT</td>
<td>T</td>
<td>1 1</td>
<td>1 0</td>
<td>16.24</td>
<td>16.12</td>
<td>12.94</td>
<td>11.29</td>
<td>11.10</td>
<td>11.56</td>
<td>99.31%</td>
<td>98.27%</td>
<td>68.78%</td>
<td>85.74%</td>
</tr>
<tr>
<td>Boeing</td>
<td>BA</td>
<td>1 1</td>
<td>1 1</td>
<td>6.34</td>
<td>6.08</td>
<td>3.64</td>
<td>3.07</td>
<td>3.27</td>
<td>2.80</td>
<td>95.93%</td>
<td>106.67%</td>
<td>53.81%</td>
<td>89.98%</td>
</tr>
<tr>
<td>Caterpillar</td>
<td>CAT</td>
<td>1 1</td>
<td>1 1</td>
<td>7.36</td>
<td>6.49</td>
<td>5.40</td>
<td>4.43</td>
<td>4.36</td>
<td>4.74</td>
<td>88.07%</td>
<td>98.55%</td>
<td>67.26%</td>
<td>80.79%</td>
</tr>
<tr>
<td>Chevron Corporation</td>
<td>CVX 1 1</td>
<td>1 0</td>
<td>5.08</td>
<td>4.76</td>
<td>3.84</td>
<td>3.14</td>
<td>2.92</td>
<td>3.32</td>
<td>93.84%</td>
<td>93.12%</td>
<td>61.32%</td>
<td>76.02%</td>
<td></td>
</tr>
<tr>
<td>Cisco Systems</td>
<td>CSCO</td>
<td>1 0</td>
<td>1 0</td>
<td>40.18</td>
<td>31.66</td>
<td>34.90</td>
<td>23.55</td>
<td>22.90</td>
<td>28.57</td>
<td>78.79%</td>
<td>97.21%</td>
<td>72.33%</td>
<td>65.61%</td>
</tr>
<tr>
<td>Coca-Cola</td>
<td>KO</td>
<td>1 0</td>
<td>0 0</td>
<td>7.51</td>
<td>7.51</td>
<td>4.88</td>
<td>3.81</td>
<td>3.81</td>
<td>4.01</td>
<td>100.00%</td>
<td>100.00%</td>
<td>50.78%</td>
<td>78.09%</td>
</tr>
<tr>
<td>DuPont</td>
<td>DD</td>
<td>1 1</td>
<td>1 0</td>
<td>4.87</td>
<td>4.91</td>
<td>3.63</td>
<td>3.12</td>
<td>3.10</td>
<td>3.09</td>
<td>100.83%</td>
<td>99.40%</td>
<td>63.05%</td>
<td>85.23%</td>
</tr>
<tr>
<td>ExxonMobil</td>
<td>XOM</td>
<td>1 1</td>
<td>1 1</td>
<td>13.20</td>
<td>11.53</td>
<td>7.95</td>
<td>7.14</td>
<td>7.14</td>
<td>7.08</td>
<td>87.32%</td>
<td>100.04%</td>
<td>61.94%</td>
<td>89.84%</td>
</tr>
<tr>
<td>General Electric</td>
<td>GE</td>
<td>2 0</td>
<td>0 0</td>
<td>53.57</td>
<td>53.57</td>
<td>31.28</td>
<td>24.82</td>
<td>24.82</td>
<td>23.53</td>
<td>100.00%</td>
<td>100.00%</td>
<td>46.33%</td>
<td>79.34%</td>
</tr>
<tr>
<td>Hewlett-Packard</td>
<td>HPQ</td>
<td>1 1</td>
<td>1 1</td>
<td>27.03</td>
<td>26.97</td>
<td>20.82</td>
<td>16.28</td>
<td>17.95</td>
<td>16.50</td>
<td>99.79%</td>
<td>110.26%</td>
<td>66.53%</td>
<td>86.21%</td>
</tr>
<tr>
<td>The Home Depot</td>
<td>HD</td>
<td>1 0</td>
<td>0 0</td>
<td>9.05</td>
<td>9.05</td>
<td>7.80</td>
<td>6.28</td>
<td>6.28</td>
<td>6.94</td>
<td>100.00%</td>
<td>100.00%</td>
<td>48.78%</td>
<td>50.66%</td>
</tr>
<tr>
<td>Intel</td>
<td>ITC</td>
<td>1 0</td>
<td>0 0</td>
<td>53.88</td>
<td>53.88</td>
<td>51.88</td>
<td>26.28</td>
<td>26.28</td>
<td>37.96</td>
<td>100.00%</td>
<td>100.00%</td>
<td>48.78%</td>
<td>50.66%</td>
</tr>
<tr>
<td>IBM</td>
<td>IBM</td>
<td>1 1</td>
<td>1 1</td>
<td>4.05</td>
<td>3.32</td>
<td>3.73</td>
<td>2.79</td>
<td>2.53</td>
<td>3.25</td>
<td>81.90%</td>
<td>90.56%</td>
<td>76.24%</td>
<td>67.77%</td>
</tr>
<tr>
<td>Johnson Johnson</td>
<td>JNJ</td>
<td>1 1</td>
<td>1 0</td>
<td>6.33</td>
<td>6.43</td>
<td>5.29</td>
<td>3.90</td>
<td>3.77</td>
<td>4.94</td>
<td>101.53%</td>
<td>96.64%</td>
<td>58.69%</td>
<td>71.38%</td>
</tr>
<tr>
<td>JPMorgan Chase</td>
<td>JPM</td>
<td>1 1</td>
<td>2 0</td>
<td>25.91</td>
<td>17.45</td>
<td>17.37</td>
<td>12.92</td>
<td>11.41</td>
<td>14.45</td>
<td>67.34%</td>
<td>88.32%</td>
<td>65.42%</td>
<td>65.07%</td>
</tr>
<tr>
<td>Kraft Foods</td>
<td>KFT</td>
<td>1 0</td>
<td>2 0</td>
<td>8.95</td>
<td>6.70</td>
<td>5.48</td>
<td>4.29</td>
<td>4.36</td>
<td>4.56</td>
<td>74.91%</td>
<td>101.42%</td>
<td>65.01%</td>
<td>79.52%</td>
</tr>
<tr>
<td>McDonald’s</td>
<td>MCD</td>
<td>1 0</td>
<td>1 1</td>
<td>4.77</td>
<td>4.21</td>
<td>5.62</td>
<td>2.39</td>
<td>3.55</td>
<td>5.08</td>
<td>88.19%</td>
<td>148.64%</td>
<td>84.36%</td>
<td>63.13%</td>
</tr>
<tr>
<td>Merck</td>
<td>MRK</td>
<td>1 1</td>
<td>1 1</td>
<td>8.31</td>
<td>6.58</td>
<td>6.46</td>
<td>5.53</td>
<td>5.37</td>
<td>5.31</td>
<td>79.15%</td>
<td>97.19%</td>
<td>81.65%</td>
<td>83.11%</td>
</tr>
<tr>
<td>Microsoft</td>
<td>MSFT</td>
<td>1 0</td>
<td>1 1</td>
<td>40.50</td>
<td>37.43</td>
<td>51.02</td>
<td>35.37</td>
<td>37.10</td>
<td>46.32</td>
<td>92.41%</td>
<td>104.90%</td>
<td>99.11%</td>
<td>72.71%</td>
</tr>
<tr>
<td>Pfizer</td>
<td>PFE</td>
<td>1 1</td>
<td>1 1</td>
<td>35.56</td>
<td>35.36</td>
<td>21.79</td>
<td>20.07</td>
<td>19.46</td>
<td>18.70</td>
<td>99.41%</td>
<td>96.97%</td>
<td>55.05%</td>
<td>89.30%</td>
</tr>
<tr>
<td>Procter Gamble</td>
<td>PG</td>
<td>1 1</td>
<td>1 1</td>
<td>6.48</td>
<td>5.76</td>
<td>5.27</td>
<td>3.96</td>
<td>3.67</td>
<td>4.79</td>
<td>88.85%</td>
<td>92.77%</td>
<td>63.75%</td>
<td>69.65%</td>
</tr>
<tr>
<td>Travelers</td>
<td>TRV</td>
<td>1 1</td>
<td>1 0</td>
<td>4.01</td>
<td>4.03</td>
<td>3.22</td>
<td>2.45</td>
<td>2.44</td>
<td>2.94</td>
<td>100.43%</td>
<td>99.40%</td>
<td>60.48%</td>
<td>75.67%</td>
</tr>
<tr>
<td>UTX</td>
<td>UTX</td>
<td>1 0</td>
<td>1 1</td>
<td>3.63</td>
<td>3.31</td>
<td>2.63</td>
<td>2.32</td>
<td>2.18</td>
<td>2.44</td>
<td>91.14%</td>
<td>94.13%</td>
<td>65.98%</td>
<td>82.95%</td>
</tr>
<tr>
<td>VZ</td>
<td>VZ</td>
<td>1 1</td>
<td>2 0</td>
<td>14.15</td>
<td>14.09</td>
<td>9.14</td>
<td>8.79</td>
<td>8.38</td>
<td>7.74</td>
<td>99.54%</td>
<td>95.36%</td>
<td>59.48%</td>
<td>91.71%</td>
</tr>
<tr>
<td>Wal-Mart</td>
<td>WMT</td>
<td>1 1</td>
<td>1 1</td>
<td>9.13</td>
<td>8.35</td>
<td>6.21</td>
<td>5.15</td>
<td>4.98</td>
<td>5.25</td>
<td>91.40%</td>
<td>96.59%</td>
<td>59.62%</td>
<td>80.17%</td>
</tr>
<tr>
<td>Walt Disney</td>
<td>DIS</td>
<td>1 1</td>
<td>2 0</td>
<td>6.92</td>
<td>4.86</td>
<td>5.66</td>
<td>4.30</td>
<td>3.27</td>
<td>4.83</td>
<td>70.21%</td>
<td>76.01%</td>
<td>67.27%</td>
<td>57.72%</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90.34%</td>
<td>98.67%</td>
<td>64.57%</td>
<td>77.65%</td>
</tr>
</tbody>
</table>
predictions of the end-of-day volume, and is more accurate compared to that using the intra-day model alone and daily volume series alone.
References


Vita

Shuhao Chen

• EDUCATION

2006 B. S. in Mathematics, Fudan University, Shanghai, China.
2010 M. S. in Mathematical Finance, Rutgers University, Piscataway, NJ.
2011 Ph.D. candidate in Statistics, Rutgers University, Piscataway, NJ.

• PROFESSIONAL EXPERIENCE

2006 Fellowship, Department of Statistics and Biostatistics, Rutgers University.
2007-2010 Teaching assistant, Department of Statistics and Biostatistics, Rutgers University.