5-2011

ARIMA models for bank failures: Prediction and comparison

Fangjin Cui
University of Nevada, Las Vegas

ARIMA MODELS FOR BANK FAILURES: PREDICTION AND COMPARISON

by

Fangjin Cui

Bachelor of Engineering
Beijing University of Chemical Technology
2003

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science in Mathematical Sciences
Department of Mathematical Sciences
College of Sciences

Graduate College
University of Nevada, Las Vegas
May 2011

THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Fangjin Cui

entitled

ARIMA Models for Bank Failures: Prediction and Comparison

be accepted in partial fulfillment of the requirements for the degree of

Master of Science in Mathematical Sciences

Chih-Hsiang Ho, Committee Chair

Kaushik Ghosh, Committee Member

Amei Amei, Committee Member

Stephen M. Miller, Committee Member

Nasser Daneshvary, Graduate Faculty Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

May 2011

ABSTRACT

ARIMA Models for Bank Failures: Prediction and Comparison

by

Fangjin Cui

Dr. Chih-Hsiang Ho, Examination Committee Chair
Professor of Mathematical Sciences
University of Nevada, Las Vegas

The number of bank failures has increased dramatically over the last twenty-two years. A common notion in economics is that some banks can become “too big to fail.” Is this still a true statement? What is the relationship, if any, between bank sizes and bank failures? In this thesis, the proposed modeling techniques are applied to real bank failure data from the FDIC. In particular, quarterly data from 1989:Q1 to 2010:Q4 are used in the data analysis, which includes three major parts: 1) pairwise bank failure rate comparisons using the conditional test (Przyborowski and Wilenski, 1940); 2) development of the empirical recurrence rate (Ho, 2008) and the empirical recurrence rates ratio time series; and 3) the Autoregressive Integrated Moving Average (ARIMA) model selection, validation, and forecasting for the bank failures classified by the total assets.

ACKNOWLEDGEMENTS

Approaching the completion of this thesis and study at University of Nevada, Las Vegas, I am always reminded that such a large endeavor is not accomplished by a single person. Rather it is the accomplishment of all the unselfish help rendered by many others. Without their contribution, it would have been impossible to get to where I am now and to achieve what I have done.

Any listing of supporters must start with Dr. Chih-Hsiang Ho, the chair of the MS - Math degree Examination Committee and member of MA - Econ degree Examination Committee, and Dr. Stephen M. Miller, the chair of the MA - Econ degree Examination Committee and member of MS - Math. I could not be at this point in my development, nor could I have ever finished this work, were it not for their effort.

I would give my deep appreciation to Dr. Chih-Hsiang Ho. It was his help and support that got me admitted to the MS program. His encouragement, guidance, and support kept me on the right track. Even though he was not my instructor in any courses; he is still a great mentor for me in both research and study. This thesis benefited greatly from his keen insight.

I am proud of being a master student of Dr. Stephen M. Miller. His guidance in class made my integration into class easier, when I first came to the U.S.A. My deep appreciation also goes out for his assistance in investing the time to review my work and pointing me in the right direction on several occasions. His patience while directing each step of my thesis was beyond all my expectations.

I also would like to thank the other members of my committee: Dr Amei Amei, Dr. Kaushik Ghosh, and Dr. Nasser Daneshvary for their support and guidance. Their

insightful comments helped shape to the fundamental arguments in this thesis.

I want to extend my appreciation to Dr. Lein-lein Chen. Although she is not a member of my committee, I consider her my personal mentor. Her motherly patience and strong will in both my studies and my life gave me great inspiration.

I also want to thank my friends: Siqi Tan, Qing Chen, Zhou Wang, Shilei Zhou, Fang Lin, and Colleen Carroll. They all contributed to my effect through course assistance, guidance and friendly discussion.

Finally, I want to thank my parents, Wei Gao and Ping Cui for providing an atmosphere of love and respect for other. Rather than questioning my decision to continue graduate study in China in 2007, they were fully supportive no matter in finance or in daily life and asked only how they could help. They were a source of inspiration throughout my lifetime and in every step of this work. In many ways, this venture is an elaboration of the many risks that they have taken.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Since September 25, 2008, when Washington Mutual Inc., became the biggest bank failure on record, almost 300 banks have collapsed. During the last 2 years, the number of bank failures significantly increased compared to the previous 6 years, during which period only around 40 banks failed. In retrospect, the number of bank failures has increased dramatically over the last twenty-five years. Out of 3879 total bank failures since 1934, when the Federal Deposit Insurance Corporation (FDIC) was established, nearly 3000 occurred between 1985 and 2010. A bank fails when it can no longer cover its obligations (liabilities) with its assets and must file for bankruptcy. The increase in bank failures is typically accompanied by high unemployment and reduced liquidity. Moreover, the survivors collect the market power by reducing competition and potentially harming consumers in the future.

To reduce the risk of bank failures, the FDIC, which guaranteed to pay the first $100,000 deposit in full to each account if the bank failed since 1980, raised the amount to $250,000 temporarily during the Financial Crisis in 2008. Additionally, the Congress passed the Emergency Economic Stabilization Act to assist the banking industry during the Financial Crisis. Thus, the United States Secretary of the Treasury spent up to $700 billion to support distressed assets from banks, which injected new capital into the banking system. Despite the aforementioned events, the number of bank failures increased. As more and more analysts focus their attention on the banking industry, a widespread question emerges: Will the situation worsen in the future? The key point raised is: Can we forecast bank failures in the future?

A common notion in economics is that some banks can become “too big to fail.” If it is true, then people who deposit in a relatively large bank face less risk than those who put their money in a smaller bank. Is this still a true statement? What is the relationship, if any, between bank sizes and bank failures?

In this study, the following proposed modeling techniques are applied to real bank failure data from the FDIC. First, the data of bank failures will be divided into three groups, based on the total assets held by the banks, as follows: Group 1, banks with assets under $300 million; Group 2, banks with between $300 million and $1 billion in assets; Group 3, banks with more than $1 billion in assets. In particular, quarterly data from 1989:Q1 to 2010:Q4 are used in the data analysis, which includes three major parts: 1) pairwise bank failure rate comparisons using the conditional test (Przyborowski and Wilenski, 1940); 2) development of the empirical recurrence rate (Ho, 2008) and the empirical recurrence rates ratio time series; and 3) the Autoregressive Integrated Moving Average (ARIMA) model selection, validation, and forecasting for the bank failures classified by the total assets.

Specifically, the fundamental tools of ARIMA are introduced in Chapter 2. Bank data are introduced in Chapter 3. Chapter 4 illustrates the ARIMA modeling techniques using the empirical recurrence rate time series converted from the Group 2 bank failures. Pairwise bank failure rate comparisons using the conditional test and the empirical recurrence rates ratio will be presented in Chapter 5. Chapter 6 concludes our work.

CHAPTER 2

FUNDAMENTAL THEORIES AND METHODS

2.1 Poisson Process

A point process is a sequence of real numbers $\{t_1, t_2, \ldots\}$ with properties

$$t_1 < t_2 < \cdots \quad \text{and} \quad \lim_{i\to\propto} t_i = +\propto.$$

Generally, at time point t_i a certain event happens. Hence, the $t_i's$ are called event times. Frequently, the event times are of less interest than the number of events, which occur in an interval $(0, \mathrm{t}]$, $\mathrm{t} > 0$. Let *N(t)* be the random variable that denotes the number of events in the interval (0, t]. For obvious reasons, $\{N(t), t \geq 0\}$ is said to be the counting process belonging to the point process $\{t_1, t_2, \ldots\}$. The intensity function of the process is defined as $\lambda(t) = \lim_{\Delta t\to 0} \frac{P(N(t, t+\Delta t] = 1)}{\Delta t}$. A counting process *N(t)* is called a Poisson process, if and only if it satisfies the following conditions: (1) N(0) = 0; (2) The random variables N(a, b] and N(c, d] are independent, for any a < b ≤ c < d; And (3) for any a < b, N(a, b] has the Poisson distribution with mean $\int_a^b \lambda(x)dx$. If $\lambda(t)$ is constant over t, the process is referred to as a homogeneous Poisson process. For a homogeneous Poisson Processes, λ is treated as the rate of occurrences.

2.2 Empirical Recurrence Rate

A key parameter desired by the economists is the recurrence rate of failures of the targeted bank group. Let t_1, K, t_n be the times of the n-ordered bank failures during an observation period $(t_0, 0)$, where t_0 is the time-origin and 0 is the present time. If h is the time-step, a discrete time series $\{z_l\}$ is generated sequentially at equidistant time

intervals $t_0 + h,\ t_0 + 2h, \ldots, t_0 + lh, \ldots, t_0 + Nh$ (= present time). Using the empirical recurrence rate (ERR) (Ho, 2008) as follows:

$$z_l = \frac{n_l}{lh} = \frac{\text{total number of bank failures in } (t_0, t_0 + lh)}{lh}.$$

where l = 1, 2, …, N. z_l can be regarded as the observation at time $t\ (= t_0 + l\,h)$, for the bank failures to be modeled. Note that z_l evolves over time and is simply the maximum likelihood estimator (MLE) of the mean, if the underlying process observed over $(t_0,$ $t_0 + l\,h)$ is a homogeneous Poisson process. The time-plot of the empirical recurrence rate (ERR-plot) offers the possibility of further insights into the data. If we have data up to time T, the value $z_{T+k},\ k \geq 1$ needs to be predicted based on the sample observation $(z_1, \mathrm{K}\ , z_T)$ of an ERR time series. We will apply the ARIMA class of models to handle our ERR time series because it is a process that evolves over time. ARIMA models are introduced next.

2.3 ARIMA Models

The Autoregressive Moving Average (ARMA) model, also called Box-Jenkins model, was introduced by Box and Jenkins (1976). The basic processes of the Box–Jenkins ARMA (p,q) model may be thought of in following ways: the autoregressive process, and the moving average process. The autoregressive model is analogous to the regression model, based on the idea that the current value of the series X_t. Autoregressive model, (AR(p) model), which constructs the present value based on a linear function of its past values and a noise term, according to

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Z_t$$

X_t is mean-zero stationary, $\phi_1, \ldots, \phi_p$ are the autoregressive coefficients for p order process. The autoregressive operator is defined to be

$$\varphi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p$$

The other one is moving average model, (MA(q) model), which describes the present term by a linear function of its past error term and a noise term, as follow:

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

The moving average operator is

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$$

A sequence, $\{Z_t\}$, of uncorrelated random variables, each with zero mean and variance σ^2, is referred to as white noise. This is indicated by the notation

$$\{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$

$\{X_t\}$ is an ARMA(p, q) process, if $\{X_t\}$ is stationary and can be written as

$$X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where $\{Z_t\} \sim WN\ (0, \sigma^2)$ and the polynomials $(1 - \varphi_1 z - \cdots - \varphi_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factor (Brockwell and Davis, 2002).

Thus, the general ARMA models are a combination of the AR operators and MA operators. Note that Z_t is a white noise sequence with zero mean and constant variance (σ^2).

Autoregressive Integrated Moving Average (ARIMA) generalizes ARMA and incorporates a wide range of nonstationary series, which are reduced to ARMA processes when differenced finite number of times. Differencing will be discussed in Section 2.4.2.

Additionally, ARIMA modeling involves three stages: model exploration, estimation, and diagnostics. The first step, model exploration, is to identify the appropriate model

and the orders of model, which are normally achieved by plots of the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF). Also, the identification can be done by fitting different possible model structures and orders, then using a goodness-of-fit statistic to select the best model, which is an auto fit procedure. The second step, estimation, is to estimate the coefficients of the model. The maximum likelihood estimation method is used for this part. The last step is a diagnostic check of the selected model. As with the linear regression model, a key element in this step is to make sure that the residuals of the selected model are normally distributed. Also, all the parameters in the model are statistically significant. The best model is the one that has the fewest parameters among all models that fit the data, which is usually guided by the principle of parsimony (Cryer, and Chan, 2008; Box and Jenkins, 1976; Shumway and Stoffer, 2005).

2.4 Data Transformation

ARMA model requires that the realized date follow a stationary process which means the statistical properties such as mean, variance, autocorrelations, etc. keep constant over time. Some mathematical transformations will be employed, if the process is not stationary. Two common transformations that will be discussed are the following.

2.4.1 Box-Cox Transformation

The Box-Cox procedure automatically identifies a transformation from the family of power transformations on Y. If the variability of the data set increases or decreases over time, the Box-Cox transformation will be employed to make the variance constant. This transformation converts original observations $Y_1, Y_2, \dots, Y_n$ to $f_\lambda(Y_1), f_\lambda(Y_2), \dots, f_\lambda(Y_n)$, where:

$$f_\lambda(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, \lambda \neq 0 \\ \log(y), \lambda = 0 \end{cases}.$$

Suitable value of λ, will be chosen to make the variability of $f_\lambda(y)$ a constant.

2.4.2 Differencing

Differencing is a data-processing technique used to remove trends or seasonal components. In this, one simply considers the difference between pairs of observations with appropriate time separations, such as, the first difference, which is denoted as:

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t,$$

where B is the backward shift operator. Differencing of order d is

$$\nabla^d X_t = (1 - B)^d X_t.$$

Furthermore, single differencing is used to remove linear trend, while double differencing is to eliminate quadratic trend. As mentioned earlier, ARIMA processes can be reduced to ARMA processes by differencing a time series.

The differencing technique adopted to deal with the seasonality of period d is the lag d difference operator ∇_d, which is defined as:

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d) X_t.$$

For example, differencing at lag 4 will remove the seasonal effect in a quarterly time series.

2.4.3 Subtracting the Mean

A zero-mean ARMA process is denoted as ARMA process in ITSM2000 (Brockwell and Davis, 2002). Therefore, the sample mean of the transformed data is subtracted from each observation, once the apparent deviations from stationarity of the data have been

removed by differencing.

2.5 Model Diagnostics and Comparison

The AR and MA terms are determined after correcting any autocorrelation that remains in the differenced series.

2.5.1 The Sample ACF/PACF of the Residuals

If the sample size *n* is large enough, the autocorrelation of residuals sequence $Y_1, \dots,$ Y_n with finite variance is approximately independent and identically distributed (iid) with distribution N(0, $\frac{1}{n}$). Therefore, whether the observation residuals are consistent with the iid noise can be tested by examining the sample correlations of the residuals. The null hypothesis of iid noise will be rejected if more than two or three out of 40 fall outside the bounds $\pm 1.96/\sqrt{n}$ or if one falls far outside the bounds (Brockwell and Davis, 2002).

2.5.2 Tests for Randomness of the Residuals

A popular test, formulated by Ljung and Box (1978), called Ljung-Box test, is commonly used to check whether the residuals of a fitted ARIMA model are observed values of independent and identically distributed random variables. It is referred to as a portmanteau test, since it is based on the autocorrelation plot and tests the overall independence based on a few lags. The Ljung-Box test is as follows.

$H_0:$ The sequence data are iid

$H_a:$ The sequence data are not iid

with the test statistic:

$$\hat{Q}(\hat{r}) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{r}_k^2 ,$$

where $\hat{r}_k = \sum_{l=k+1}^{n} \hat{a}_l \hat{a}_{l-k} \Big/ \sum_{l=1}^{n} \hat{a}_l^2$, the estimated autocorrelation at lag k ,

n = sample size,

m = number of lags being tested

As a rule of thumb, the sample ACF and PACF are good estimates of the ACF and PACF of a stationary process for lags up to about a third of the sample size (Brockwell and Davis, 2002).

After a model has been fitted to a series $z_1,..., z_n$, we got the residuals $\hat{a}_1,...,\hat{a}_n$. If no model is being fitted, then $\hat{a}_1,...,\hat{a}_n$ are the "mean corrected" vectors of $z_1,..., z_n$.

If the sample size n is large, the distribution of $\hat{Q}(\hat{r})$ is roughly χ^2_{m-p-q} under the null hypothesis, where $m - p - q$ is the degree of freedom of the chi-square distribution, and, $p + q$ is the number of parameters of the fitted model. The null hypothesis will be rejected at level α , if $\hat{Q} > \chi^2_{1-\alpha;m-p-q}$. Consequently, the sequence data are not independent, or their autocorrelations are significantly different from zero.

2.5.3 AIC, BIC and AICC Statistics

Another approach to model selection is the use of information criteria such as Akaike information criterion (AIC), or the Bayesian information criterion (BIC), which is a Bayesian modification of the AIC statistic. The bias-corrected version of the AIC statistic, the AICC statistic, introduced by Akaike in 1974, is employed in this thesis as information criterion to select appropriate models using the ITSM2000 package. Each information statistic is defined as the following,

$$AIC_{p,q} = N \log \hat{\sigma}_{\varepsilon}^2 + 2r$$

$$AICC_{p,q} = N \log \hat{\sigma}_{\varepsilon}^2 + 2rN/(N - r - 1)$$

$$BIC_{p,q} = N \log \hat{\sigma}_{\varepsilon}^2 + r \log N$$

where $\hat{\sigma}_{\varepsilon}^2$ is the maximum likelihood estimator of σ_{ε}^2, and $r = p + q + 1$ is the number of parameters estimated in the model, including a constant term. The second term in all three equations is a penalty for increasing r. Thus, minimizing the number of parameters is one of the ways to minimize the values of these criteria. The best model should be the model that has the fewest parameters yet still sufficiently describes the data. A small value of AICC shows a good model. Nonetheless, it should be used only as rough guide.

2.6 Forecasting

The appropriate ARIMA model obtained will be used to predict future values of the time series from the past values. The forecasting function given below will be chosen to have, as follows, has the minimum mean square error.

$$z_t = f(z_1, \dots, z_{t-1}) + a_t,$$

where $f(z_1, \dots, z_{t-1})$ is a function of the past values of the series and determined by the past value of data. The second part a_t, noise part, is a sequence of independent and I dentically distributed (iid) variables as mentioned before. Predictions will be achieved by forecasting the residuals and then inverting the transformations adopted to arrive at forecasts of the original series.

CHAPTER 3

BANK DATA

Commercial bank data were compiled from the Chicago Federal Reserve database (www.chicagofed.org). The report of Condition and Income data includes information from individual commercial banks and savings associations that are regulated by the Federal Reserve System, the Comptroller of the Currency, and the Federal Deposit Insurance Corporation (FDIC). The data are reported and published on a quarterly basis. The numbers of bank failures in the United States during 1989:Q1 to 2010:Q4 are obtained from the FDIC failed bank list. Based on this list, 1821 banks were reported to fail over the 88 quarters (Figure 3.1).

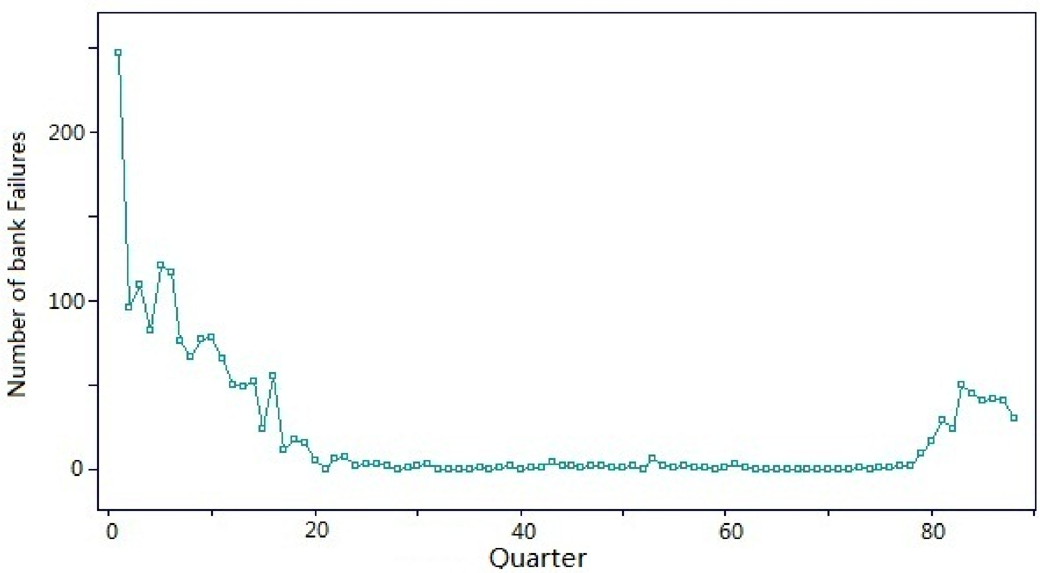

Figure 3.1 Plot of the Number of Bank Failures from 1989:Q1 to 2010:Q4

The FDIC (www.fdic.gov) reports bank failures on a weekly basis, typically on a Friday afternoon to avoid a run on bank assets. Bank failures in this thesis are drawn

from the FDIC bank failure reports, which list failed banks by name, location, charter type, total assets, and other characteristics. Consistent with the solvent bank data, however, we count the number of bank failures on a quarterly basis. In this study, individual banks that failed during 1989:Q1- 2010:Q4 are divided into three groups by total assets level.

3.1 CPI Adjustment

In economics, the nominal level of prices of goods and services changes over a period of time. When the price level rises, each unit currency buys fewer goods and services. The purchasing power of money --- the real value in the internal medium of exchange and unit of account in the economy changed over time. The Consumer Price Index (CPI) is used to bridge nominal values to real values. The total assets of banks reported are measured by nominal price. To make the total assets in different time periods comparable, the total assets of banks are converted to the real values which are based on:

$$Total\ Assets^{*} = \frac{CPI_b}{CPI_i} \times Total\ Assets_i,$$

where $Total\ assets_i$ is the nominal total assets of a failed bank at time i (the month a failure was reported); CPI_i is CPI at the *i*th month that bank failed; CPI_b is the CPI for the base month (taken as September 2010 in this thesis). $Total\ Assets^{*}$ is the total assets deflated by the CPI.

Monthly CPI data are obtained from the Federal Reserve Bank of St. Louis Federal Reserve Economic Data (FRED) (http://research.stlouisfed.org/fred2/).

3.2 Bank Classification

The data on bank failures will be divided into three groups, based on the adjusted total assets held by the banks at the time they failed, as follows: Group 1, banks with

assets under \$300 million; Group 2, banks with assets between \$300 million and \$1 billion; Group 3, banks with more than \$1 billion in assets. Quarterly numbers of bank failures for each group are retrieved from the original Failed Bank List are summarized in Table 3.1. Plots of the time series on the original failures are illustrated as Figure3.2. Plots of the time series on the original failures are illustrated as Figure3.2

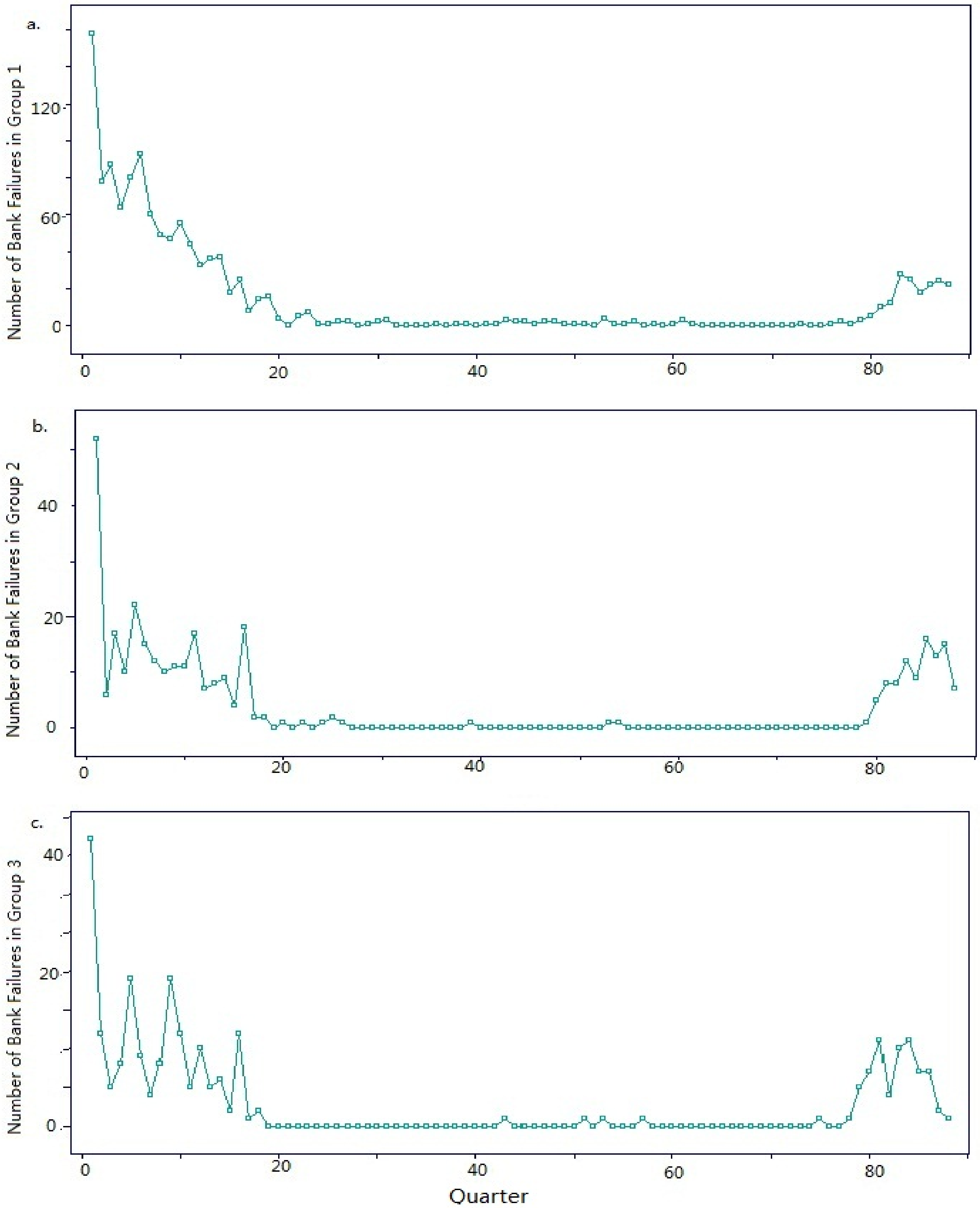

Figure 3.2 Plots of Numbers of Bank Failures from 1989:Q1 to 2010:Q4: **a.** Group 1; **b.** Group 2; **c.** Group 3

CHAPTER 4

EMPIRICAL RECURRENCE RATE

4.1 ERR-Plots

Figure 4.1 shows the Empirical Recurrence Rate plot (ERR-plot) for each group from 1989:Q1 to the present time 2010:Q4 with time step =1 quarter.

4.2 Data Splitting

Cross-validation is the statistical practice of splitting a sample of data into two subsets so that the analysis is initially performed on one subset, while the other subset is retained for subsequent use in confirming and validating the initial analysis. The first subset is called training sample and is used to develop a model for prediction. The second part, called prediction set is used to evaluate reasonableness and predictive ability of the selected model. In this study, cross-validation is used as an additional guide for model selection.

We will use the ITSM2000 software (Brockwell and Davis, 2002) to model the ERR data with time-step h = 1 quarter. Recall that there are 88 data points for the entire time series. First, we split the data into: training sample and prediction set. In this case, our training sample is the original data set excluding the last 6 ERRs, which will form the prediction set (Figure 4.1). These six ERR values in the prediction set, representing the most recent 6 quarters of each bank group, will be compared with those of the six-step predictions produced by a candidate model. Of course, the size of a prediction set is quite flexible as long as it fits a common goal of model selection.

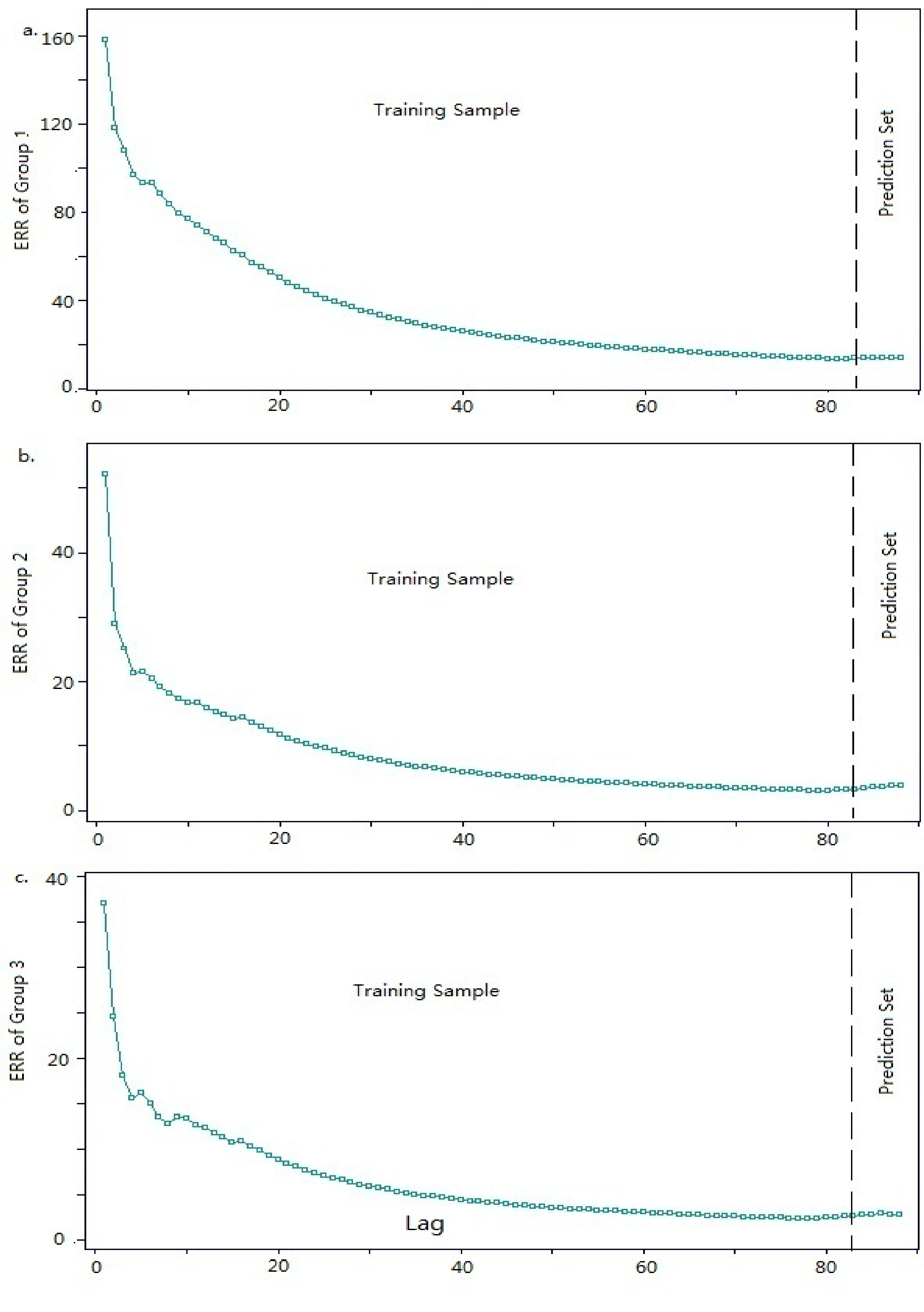

Figure 4.1 ERR Plots of Bank Failures through the Entire Time Period (Training Sample and Prediction Set): **a.** Group 1 (Assets Less than $300 Million); **b.** Group 2 (Assets between $300 Million and $1 Billion); **c.** Group 3 (Assets more than $1 Billion)

4.3 ARIMA Modeling for Group 2 ERRs

In this section, ARIMA modeling and computational techniques are presented to fit the ERRs of the training sample of Group 2 (Figure 4.1b) and to predict its future number of failures, which will then be compared to the prediction set. The plot of the sample ACF (Figure 4.2 b) show that the sample ACF is slowly decaying. It indicates non-stationary behavior and seasonality. Thus differencing is applied. Since the data has evident nonconstant variance, we use the Box-Cox transformation to stabilize the variability. After applying the Box-Cox transformation with $\lambda = 1.5$, we see the trend still exists (Figure 4.3). Initially we take the differencing operator ∇ on the training sample at lag 2. Figure 4.4 tells us the resulting series is almost stationary.

We then subtract the sample mean from each observation of the differenced series to generate a stationary zero-mean time series (Figure 4.4). The sample ACF and PACF suggest and lead to an AR(5) model. This leads to the following estimated model:

ARMA Model:

$X_t = 1.909\ X_{t-1} - 0.1431\ X_{t-2} - 1.430\ X_{t-3} + .5489\ X_{t-4} + .1113\ X_{t-5} + Z_t$

WN Variance = .120997E+03

Standard Error of AR Coefficients

0.000240 0.000053 0.000044 0.001002 0.000668

Note that X_t represents zero-mean stationary time series of ERR, and the error term Z_t represents a white noise process.

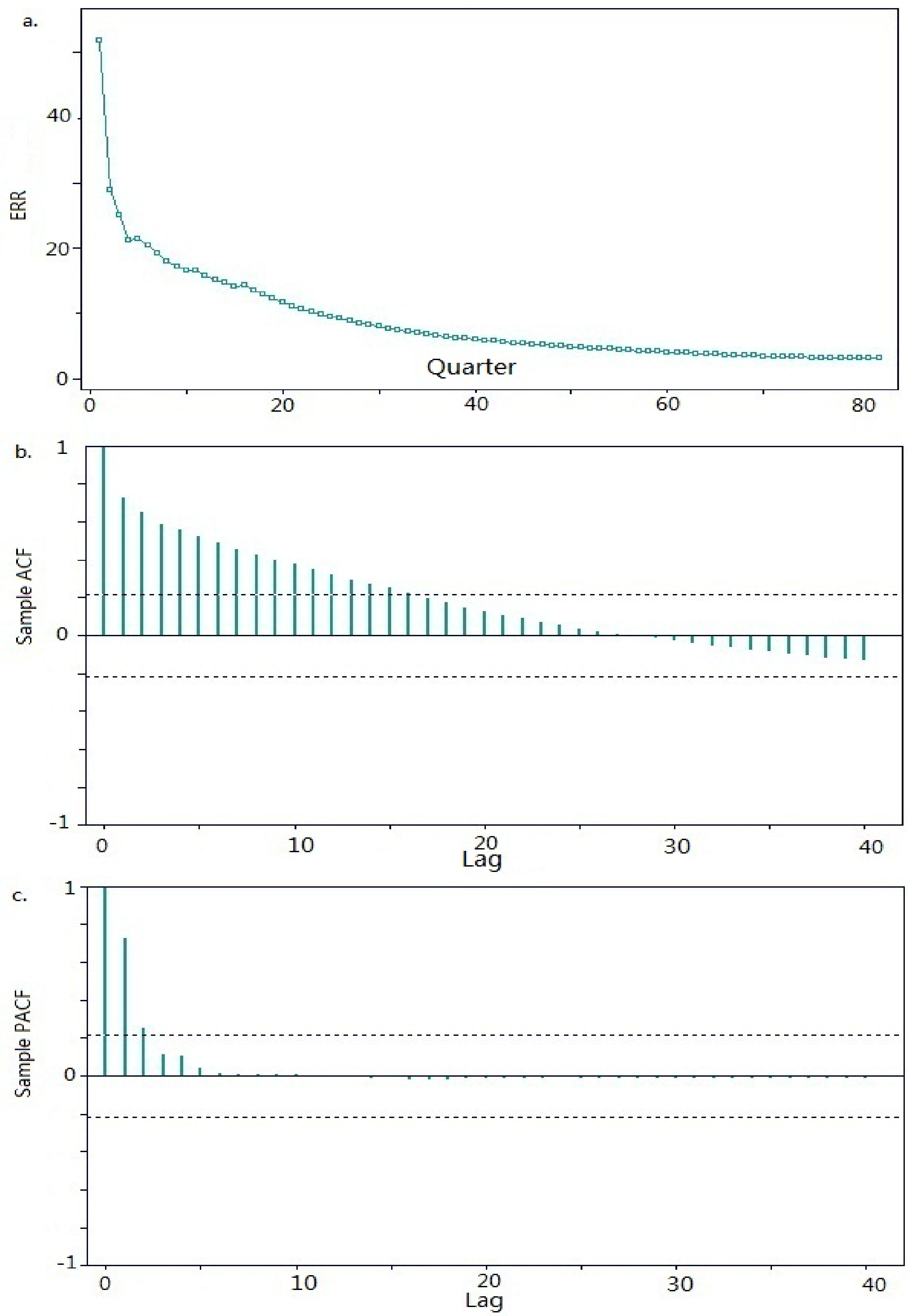

Figure 4.2. **a**, ERR-plot of Training Sample (Group 2); **b**, Sample ACF; **c**, Sample PACF.

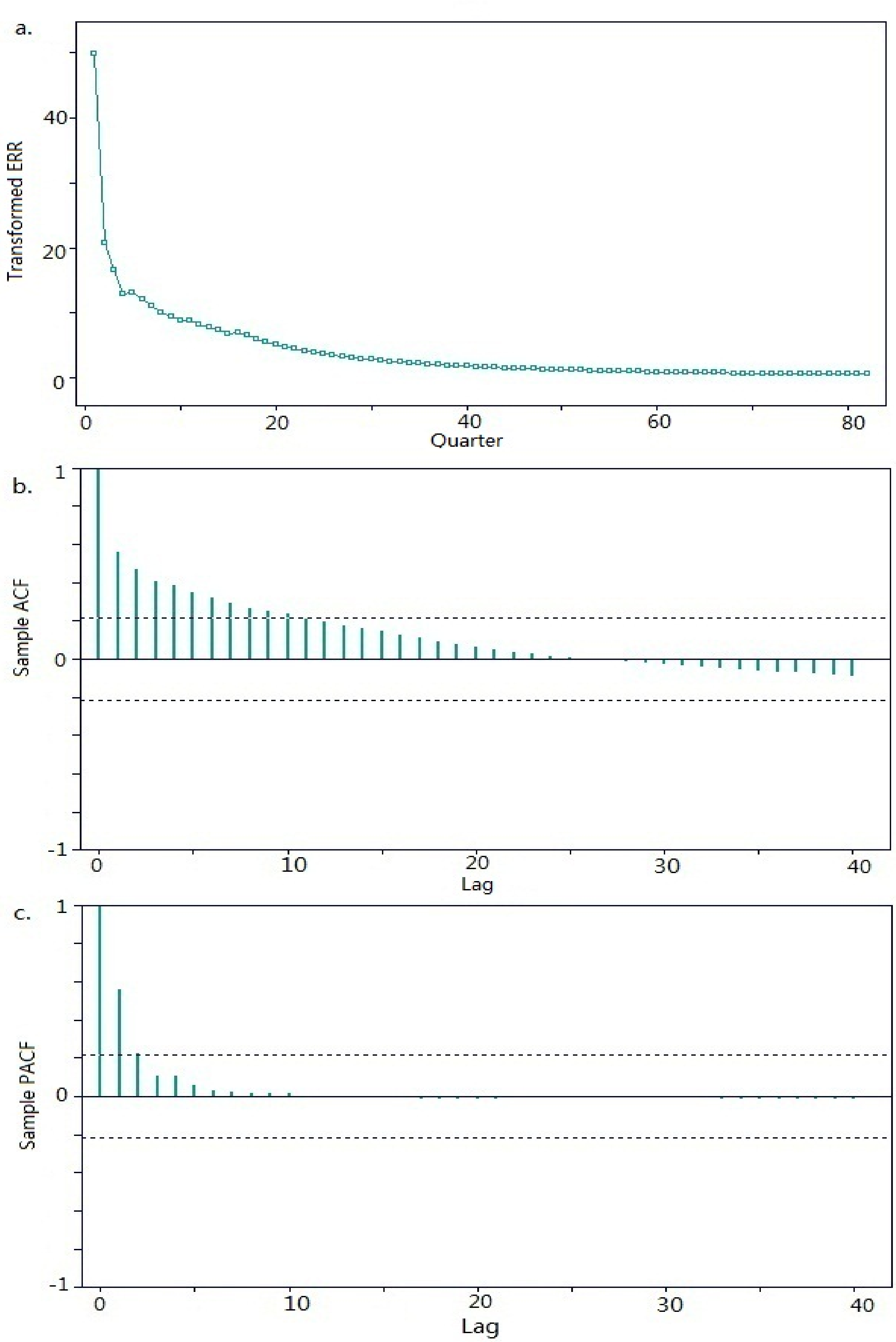

Figure 4.3. **a**, Group 2 Time-plot after Box-Cox Transformation with λ=1.5 ; **b**, Sample ACF; **c**, Sample PACF.

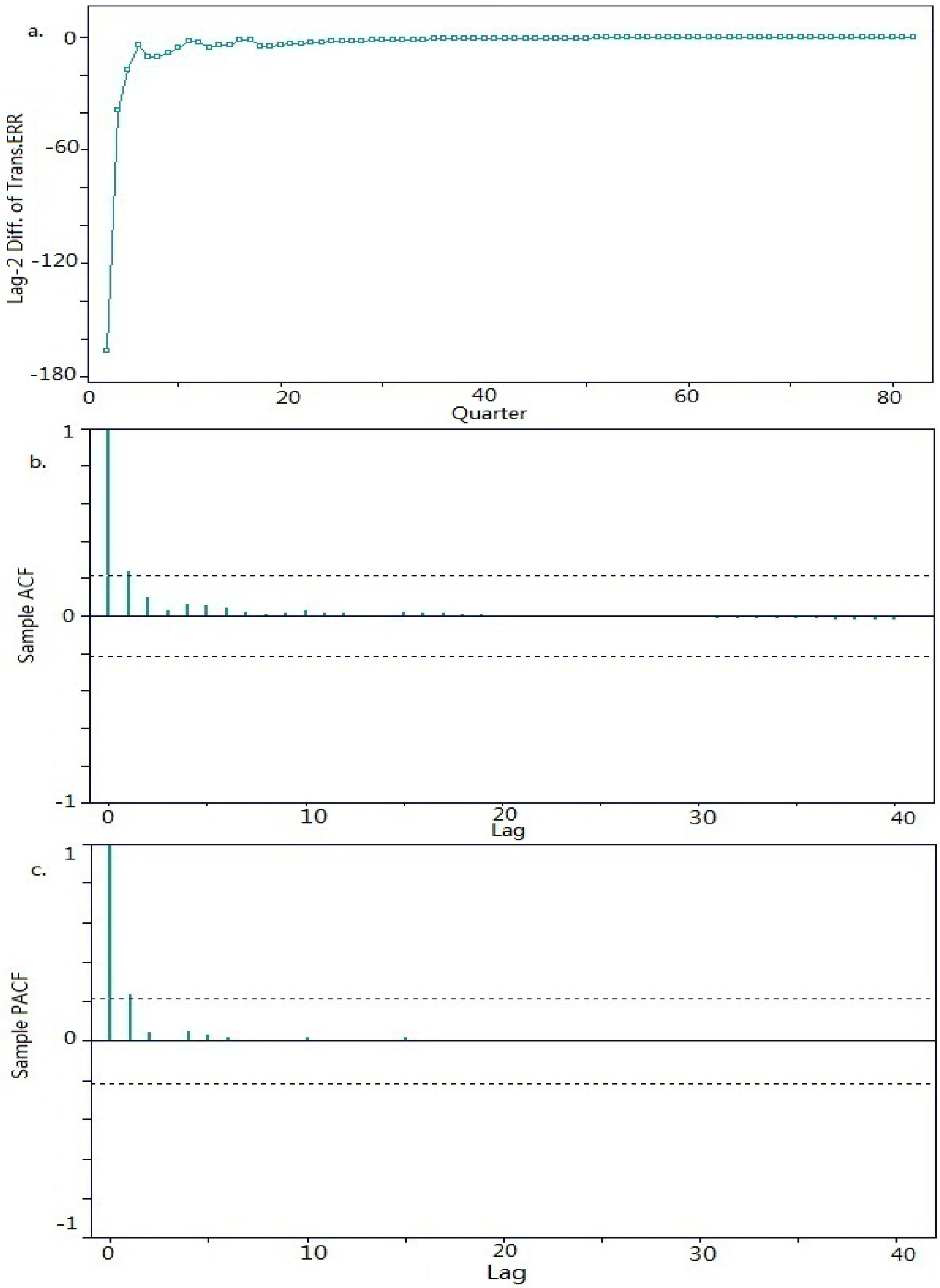

Figure 4.4. **a**, Group 2 Time-plot after Differencing at Lag 2; **b**, Sample ACF; **c**, Sample PACF.

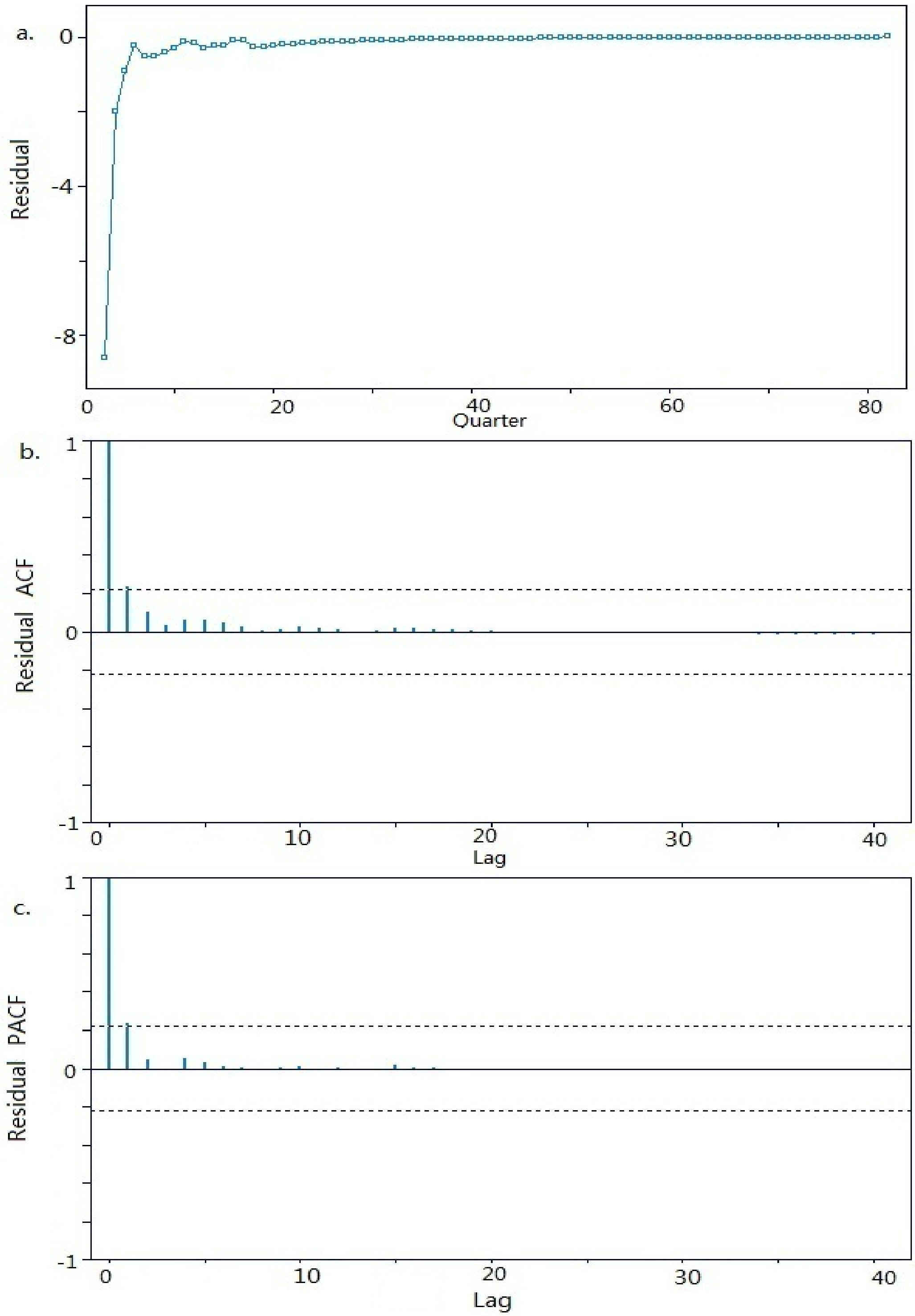

Figure 4.5. Diagnostics for the AR(5) Model. **a**, Residual plot; **b**, Residual ACF; **c**, Residual PACF.

Figure 4.5 is a set of diagnostic plots produced by ITSM2000 package, which show

the ACF and PACF of residuals of training sample. The AICC statistic is 637.718. And the Ljung-Box test is not significant (p-value = 0.95705) indicating that the residuals are approximately white noise.

Table 4.1 compares the numerical values of the observed ERRs to predicted ERRs and observed counts to predicted counts numbers. The predicted counts are derived from the predicted ERRs. The observed bank failure numbers and the predictions are compared in Figure 4.6.

Table 4.1 Numerical Values of Observed ERRs, Observed Counts in the Prediction Set, Predicted ERRs (Using AR(5)) and Corresponding Predict Counts for the Prediction Set, and the Predicted ERRs Using the AR(5) with their Counterparts (the Corresponding Values Derived from the Predicted ERRs)

Time	ERR		Counts	
	Observed	Predicted	Observed	Predicted
2009:Q3	3.325301	3.33014	12	12.40164 rounded to 12
2009:Q4	3.392857	3.39551	9	8.82122 rounded to 9
2010:Q1	3.541176	3.51463	16	13.52071 rounded to 14
2010:Q2	3.651163	3.5556	13	7.03805 rounded to 7
2010:Q3	3.781609	3.64738	15	11.54046 rounded to 12
2010:Q4	3.818182	3.63738	7	2.76738 rounded to 3

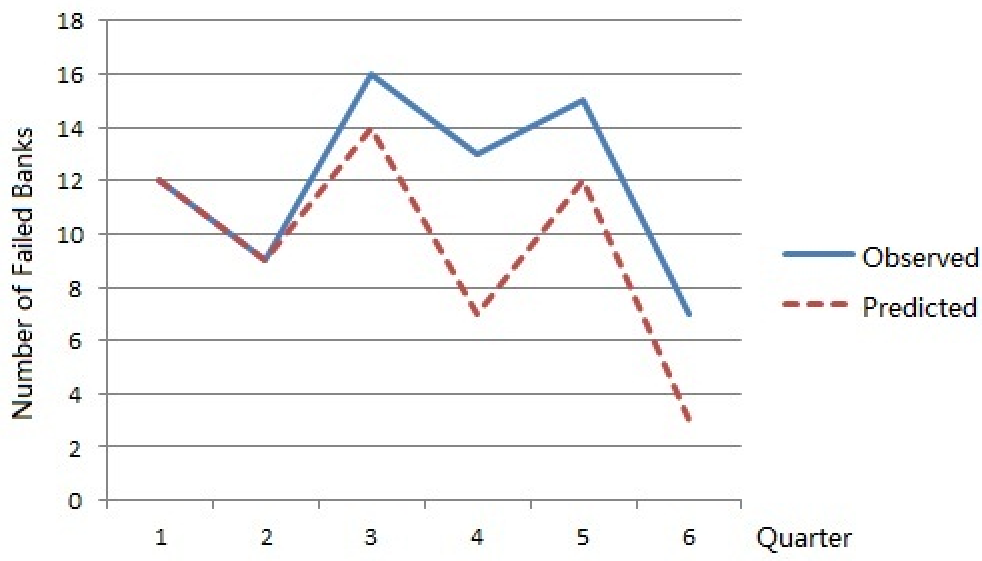

Figure 4.6 Comparison of Observed Number of Bank Failures with the Forecasts in the Presiction Set for Group 2, 2009:Q3-2010:Q4

CHAPTER 5

EMPIRICAL RECURRENCE RATES RATIO

5.1 Methodology

5.1.1 The Conditional Test

Let X_1 and X_2 be independent observations from Poisson (λ_1) and Poisson (λ_2) distributions respectively. Then, the joint distribution of X_1 and X_2 is given by:

$$\mathrm{f}(x_1,x_2) = \left[\frac{{\lambda_1}^{x_1} e^{-\lambda_1}}{x_1!}\right]\left[\frac{{\lambda_2}^{x_2} e^{-\lambda_2}}{x_2!}\right] = \frac{{\lambda_1}^{x_1}{\lambda_2}^{x_2}}{x_1!x_2!} e^{-(\lambda_1+\lambda_2)} \qquad X_1 = 0,1,2,\ldots$$

$$X_2 = 0,1,2,\ldots$$

Note that

$$X_1 + X_2 = S \sim Poisson(\lambda_1 + \lambda_2).$$

The well-known method of testing the difference between two Poisson means is the conditional test (Przyborowski and Wilenski, 1940). It is based on the fact that the conditional distribution of X_1 given $X_1 + X_2 = S$ is binomial, whose success probability is a function of the ratio $\frac{\lambda_2}{\lambda_1} = \rho$.

The proof goes as follows. Considering the conditional distribution, X_1 given S = s > 0. The probability mass function of the conditional distribution of X_1 given S = s is given by:

$$f(x_1|S = s) = \frac{P(X_1 = x_1, X_1 + X_2 = s)}{P(X_1 + X_2 = s)}$$

$$= \frac{e^{-\lambda_1}\frac{{\lambda_1}^{x_1}}{x_1!} \cdot e^{-\lambda_2}\frac{{\lambda_2}^{s-x_1}}{(s-x_1)!}}{e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^s}{s!}}$$

$$= \binom{s}{x_1}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{s-x_1}$$

$$= \binom{s}{x_1}\left(\frac{1}{1+\rho}\right)^{x_1}\left(\frac{\rho}{1+\rho}\right)^{s-x_1} \sim Binomial\ (s, \frac{1}{1+\rho})$$

Let $\frac{1}{1+\rho} = p$. Then, to test the equality of two Poisson means, is to test the following hypotheses:

$$H_0: \lambda_1 = \lambda_2 \text{ versus } H_1: \lambda_1 \neq \lambda_2$$

which is equivalent to

$$H_0: \rho = 1 \text{ versus } H_1: \rho \neq 1.$$

which is equivalent to

$$H_0: p = \tfrac{1}{2} \text{ versus } H_1: p \neq \tfrac{1}{2},$$

It can be generalized as follows:

$$H_0: p \geq p_0 \text{ versus } H_1: p < p_0,$$

where $0 < p_0 < 1$. And it is equivalent to

$$H_0: \rho \leq \rho_0 \text{ versus } H_1: \rho > \rho_0,$$

where $\rho_0 > 0$, and $\rho_0 = \frac{1-p_0}{p_0}$.

When $X_1 = k$ is observed, the conditional test (C-test) rejects H_0 , if

$$p\text{-}value = P(X_1 \leq k|S = s) = \sum_{i=0}^{k}\binom{S}{i} {p_0}^i (1-p_0)^{s-i} \leq \alpha,$$

where α is the level of significance. Of course, normal approximation can be implemented for the above binomial test for large *s*.

5.1.2 Conditional Tests for Bank Failures

In this thesis, we divide the banks into three groups based on the levels of total assets of the banks. For each bank group, we assume that the number of bank failures follows a

homogeneous Poisson process with failure rate λ. According to the classification criterion described in Chapter 3, Group 1 represents banks with assets under $300 million; Group 2 is banks with assets between $300 million and $1 billion; and banks in Group 3 have assets more than $1 billion. Let λ_i be the failure rate of *i*th group of banks, $i = 1,2,3.$ Also, let

$$\rho_{ij} = \frac{\lambda_j}{\lambda_i} \text{ and } p_{ij} = \frac{1}{1+\rho_{ij}},\ 1 \le i < j \le 3.$$

Then a hypothesis for bank failure rates comparison between any two groups i and *j* can be presented as follows:

$$H_0:\ \rho_{ij} \le \rho_{ij}^0 \text{ versus } H_1:\ \rho_{ij} > \rho_{ij}^0,$$

where $\rho_{ij}^0 > 0$, is a known reference ratio calculated from solvent bank database, which will described later .The corresponding C-test is then

$$H_0: p_{ij} \ge p_{ij}^0 \text{ versus } H_1: p_{ij} < p_{ij}^0,$$

where $0 < p_{ij}^0 < 1$ and $p_{ij}^0 = \frac{1}{1+\rho_{ij}^0}$.

The reference ratio ρ_{ij}^0, for each (*i*, *j*) pair, is calculated by taking the average of all the quarterly solvent commercial bank group ratios through the entire observation period. Consequently, if the failure rate ratio (ρ_{ij}) is tested significantly higher than the historical population ratio (ρ_{ij}^0), the *j*th group yields a disproportionally higher failure rate than the *i*th group.

For example, in comparing Group 1 and Group 2, the reference value, ρ_{12}^0, calculated from the solvent bank data base is 0.183689 and the corresponding p_{12}^0 is 0.844816. The total numbers of bank failures during the entire time period are 1238, and 336 for Group 1 and Group 2, respectively. Based on the C-test,

$$p\text{-}value = \mathrm{P}(X_1 \leq 1238 \mid S = 1574)$$

$$= \sum_{k=0}^{1238} \binom{1574}{k} (0.844816)^{k} (1 - 0.844816)^{(1574-k)} = 5.9482\text{E-}10$$

The null hypothesis is rejected, indicating that Group 1 has contributed less than 84.48% of the total failures, and it is statistically significant. In other words, compared with Group 1, banks in Group 2 are more likely to fail during the observation period. Recall that Group 1 includes banks with total assets below 300 million dollars, while Group 2 has total assets between 300 million dollars and 1 billion dollars. Therefore, the result of the above C-test implies that smaller banks have significantly higher survival rate during the observation period. Additionally, all pairwise comparisons reinforce the above conclusion. Table 5.1 lists the results. It seems that the statement: “Too Big to Fail.” is not supported by our data analysis during this particular observation period.

Table 5.1 Conditional Tests for Pairwise Comparisons

	Group (1, 2)	Group (2, 3)	Group (1, 3)
Total number of failures (X_i, X_j)	(1238, 336)	(336, 247)	(1238, 247)
Total number of both group (s)	1574	583	1485
Solvent bank ratio (ρ_{ij}^{0})	0.183689	0.522247	0.093755
Solvent bank probability (p_{ij}^{0})	0.844816	0.656924	0.914282
p-value	5.9482E-10	3.21769E-05	1.7062E-23

5.1.3 Empirical Recurrence Rates Ratio

The C-test examines the relationship of means of two homogeneous Poisson processes, which have constant expected values. Motivated by the ideas of the C-test and the Empirical Recurrence Rate developed by Ho (2008), we produce an Empirical Recurrence Rates Ratio (ERRR) time series for the bank failure rates ratio as follows:

Let $t_1, t_2, \dots, t_n$ be the n-ordered bank failure times during an observation period (t_0, t_0+Nh) from the past to the present. The ERRR is then defined as follows:

$$d_l = \frac{\sum_{j=1}^{l} X_{1j}}{\sum_{j=1}^{l}(X_{1j}+X_{2j})}, \ l = 1,2,\dots,\mathrm{N}.$$

X_{ij}= number of failures in group i in time (t_0, t_0+jh]

where $i = 1, 2$ and $j = 1, 2, \dots, \mathrm{N}$. Then a discrete time series $\{d_l\}$ is generated sequentially as $t_0 + h$, $t_0 + 2h$,…, $t_0 + lh$,…, $t_0 + Nh$ (= the present time). h presents the time step.

Both the ERR and ERRR offer the possibility of developing a model, monitoring and predicting bank failure rate ratios. Moreover, if both of the targeted processes are homogeneous Poisson processes, then the ERRR is the maximum likelihood estimator (MLE) of p, and the MLE of ρ can be obtained by the invariance property of the MLE.

5.2 ARIMA Modeling: All Groups

5.2.1 Training Sample Modeling: λ_2/λ_1

Along the same line of argument as for ERR, we apply the ARIMA class of models to handle our ERRR time series because it is a process that evolves over time. The modeling process is the same as that detailed in Chapter 4. The following analysis uses the ERRR time series (Figure 5.1) generated from Group 1 (=X_1) and Group 2 (=X_2).

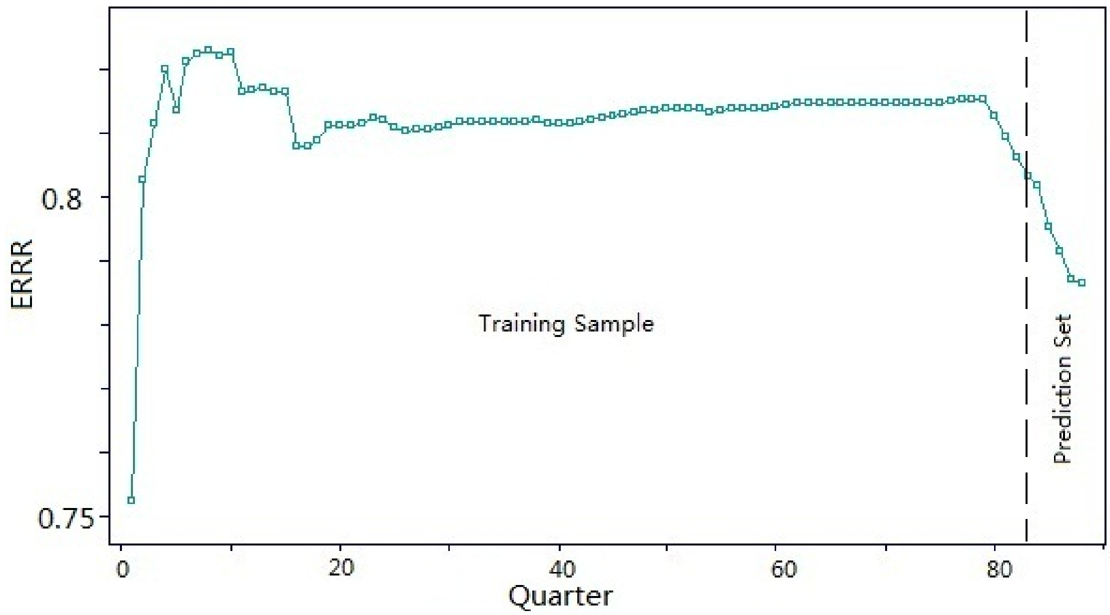

Figure 5.1 ERRR-plot for Group 1 versus Group 2 from 1989:Q1 to 2010:Q4

The plots of the training sample (first 82 quarters) and its sample ACF and PACF in Figure 5.2 show nonstationarity and periodicity. Therefore, the Box-Cox transformation, and differencing will be employed to remove the trend and seasonality. Since the plot (Figure 5.2) shows nonconstant variance, we consider the Box-Cox transformation to stabilize the variability. After the $\lambda = 1.5$ Box-Cox transformation, we see the trend still exists (Figure 5.3). We then take the differencing operator ∇ on the training sample at lag 3. Figure 5.4 tells us the series has not reached stationary yet. So we do further differencing at lag 1.

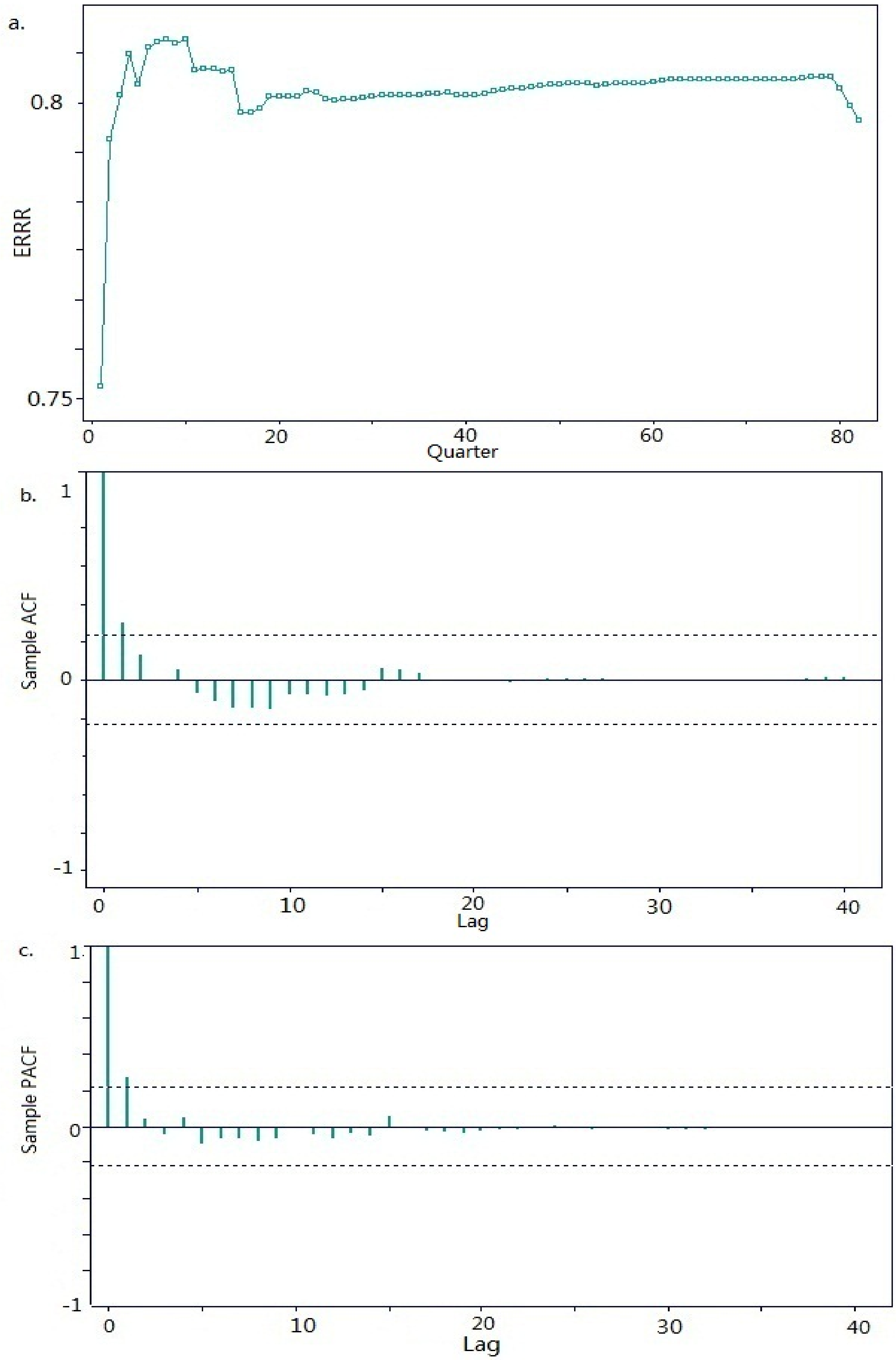

Figure 5.2 a. Time-plot **b**. Sample ACF, **c.** Sample PACF of Training Sample with the ERRRs from Group 1 versus Group 2

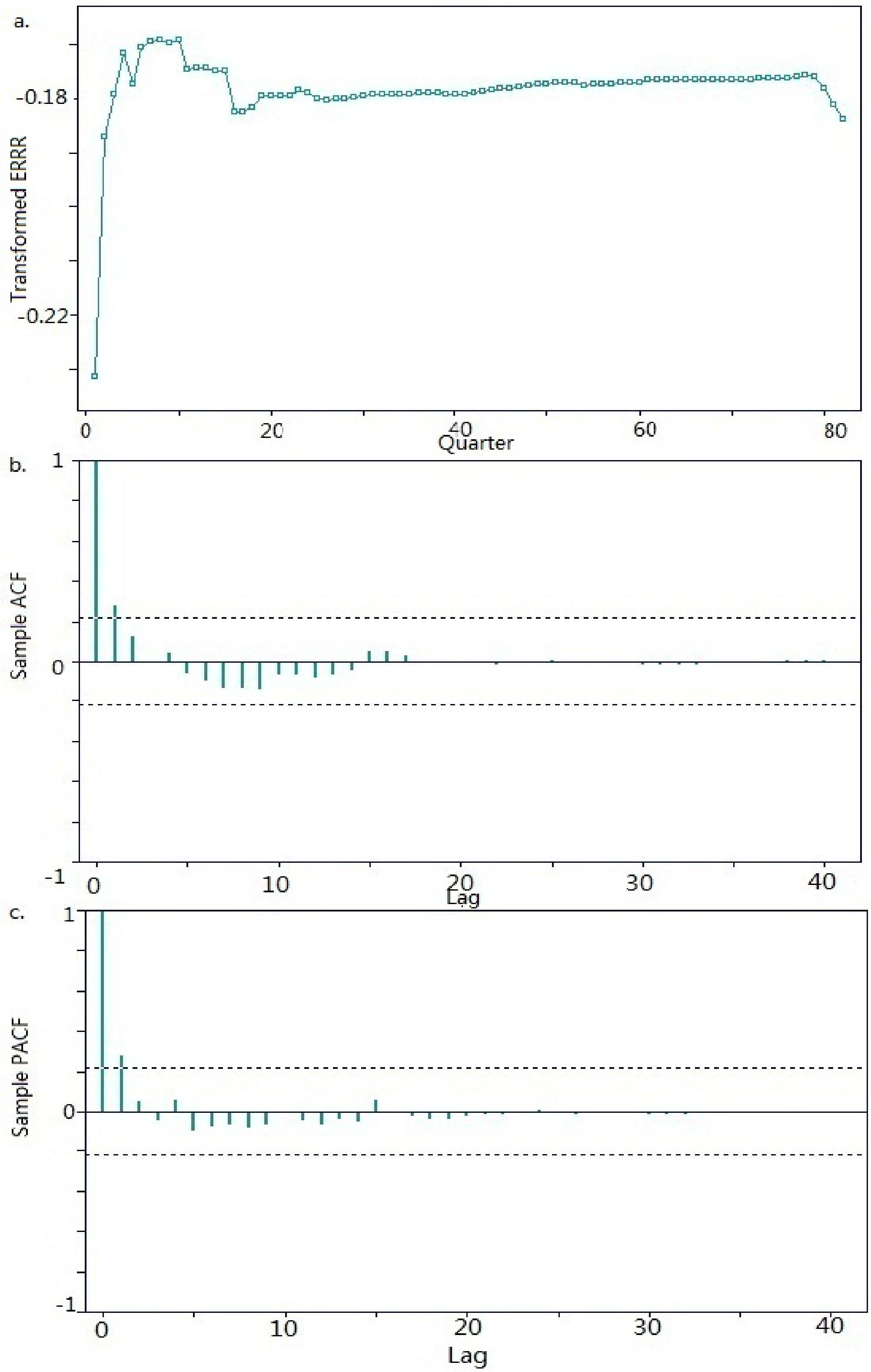

Figure 5.3. **a**, Time-plot after Box-Cox Transformation with λ=1.5 ; **b**, Sample ACF; **c**, Sample PACF for the ERRR of Group1 versus Group2.

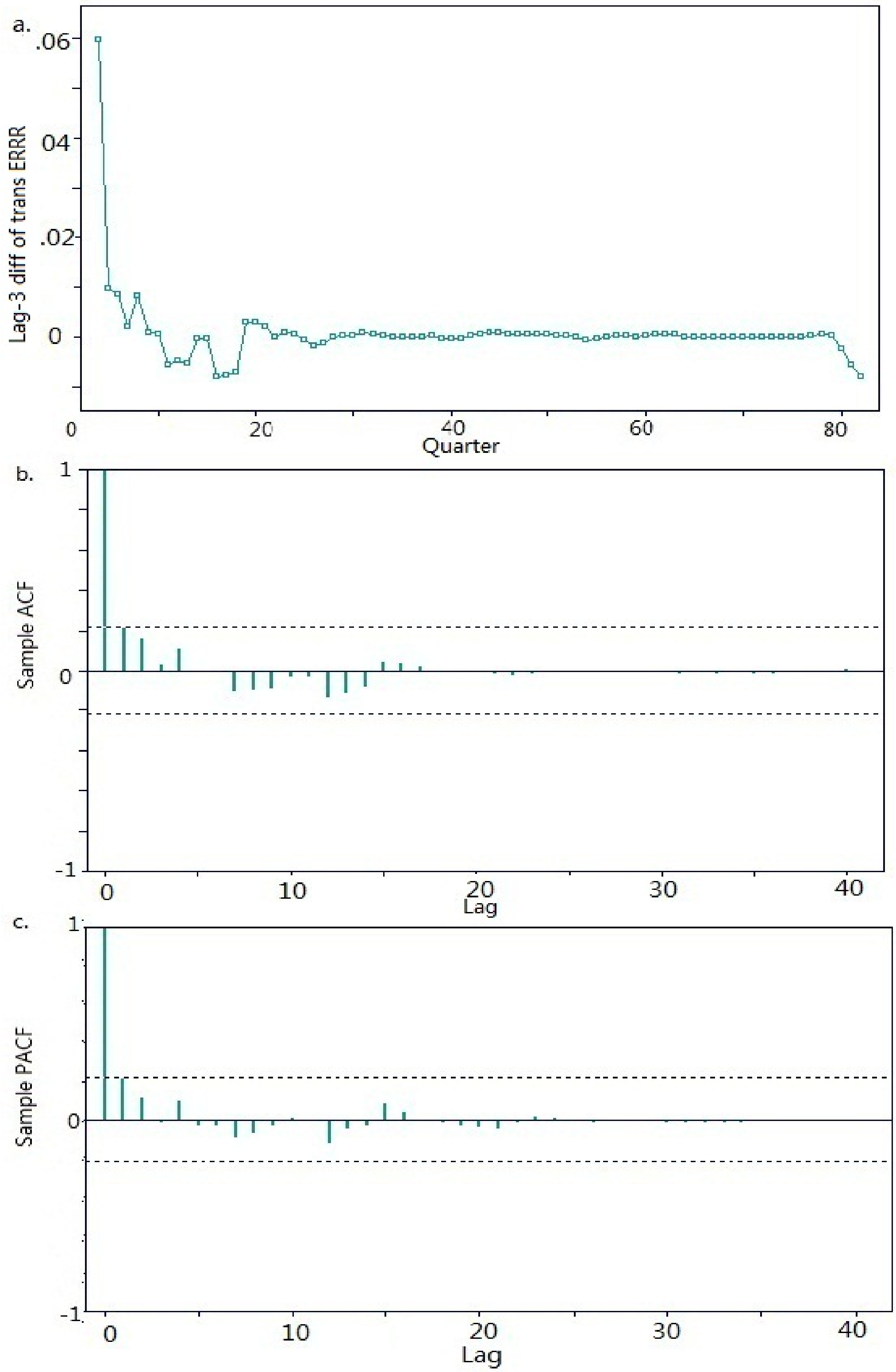

Figure 5.4. **a**, Time-plot after Differencing at Lag 3; **b**, Sample ACF; **c**, Sample PACF.

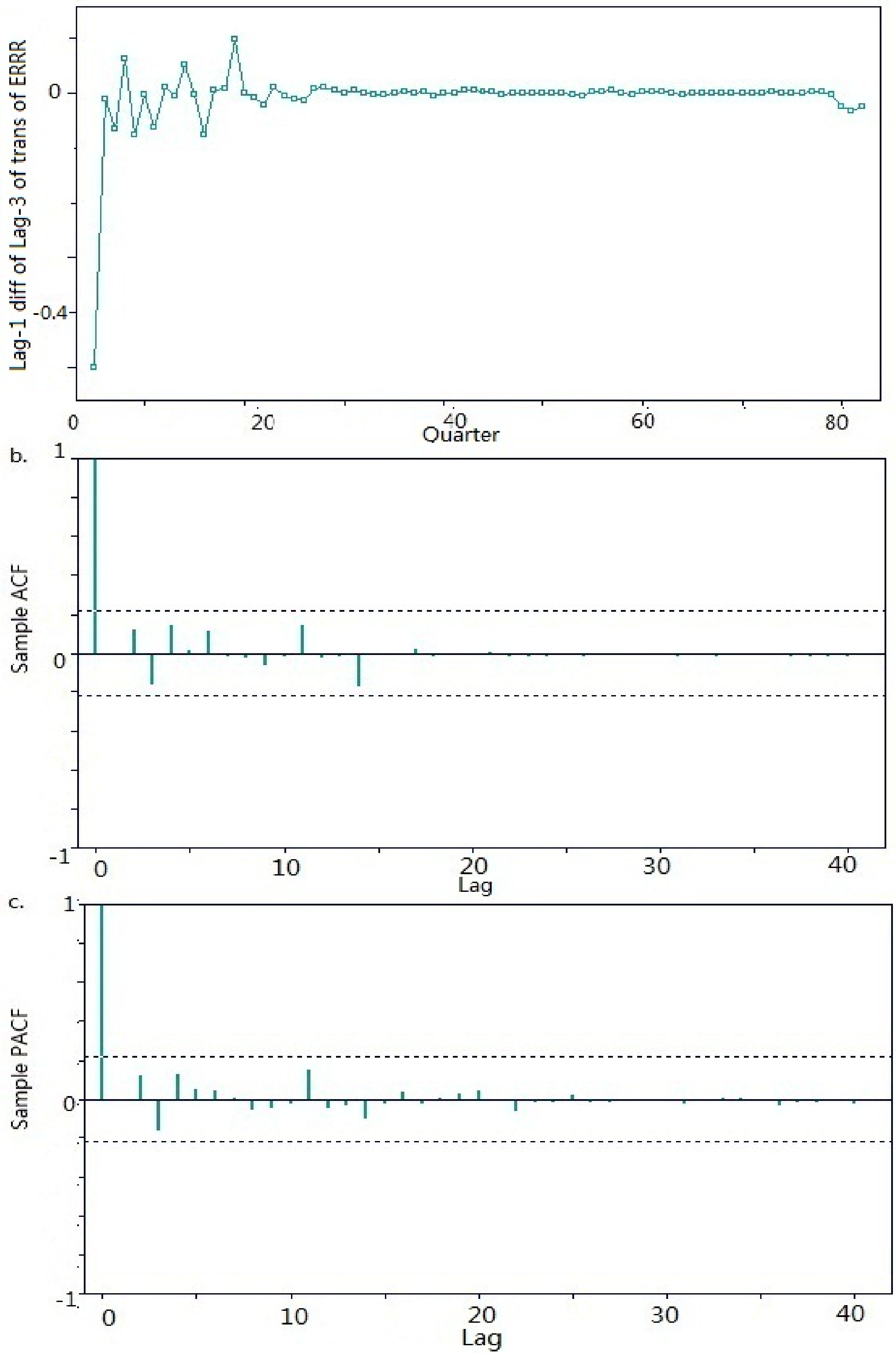

Figure 5.5. **a**, Time-plot after differenced at Lag 1of Lag 3 Transformed ERRR; **b**, Sample ACF; **c**, Sample PACF

We then subtract the sample mean from each observation of the differenced series to generate a stationary zero-mean time series (Figure 5.5). The sample ACF and PACF suggest and indicate an AR(3) model. Therefore, our estimated model is:

ARMA Model:
$X_t = .3829\ X_{t-1} + .5415\ X_{t-2} - .7467\ X_{t-3} + Z_t$
WN Variance = .000027
Standard Error of AR Coefficients
.210673 .189058 .170469

Note that X_t represents a twice-differenced stationary mean-corrected time series and the error term Z_t represents a white noise process. The AICC statistic is -586.602. Also, the Ljung-Box test is not significant with p-value= 0.45713, indicating that the residuals are approximately white noise. The plots of sample ACF/PACF of the residuals are shown in Figure 5.6.

We also compare the predicted ERRRs with the actual ERRRs in the prediction set. Figure 5.7 indicate that the model fit relatively well. Table 5.2 shows the numerical comparison among these two sets of ERRR.

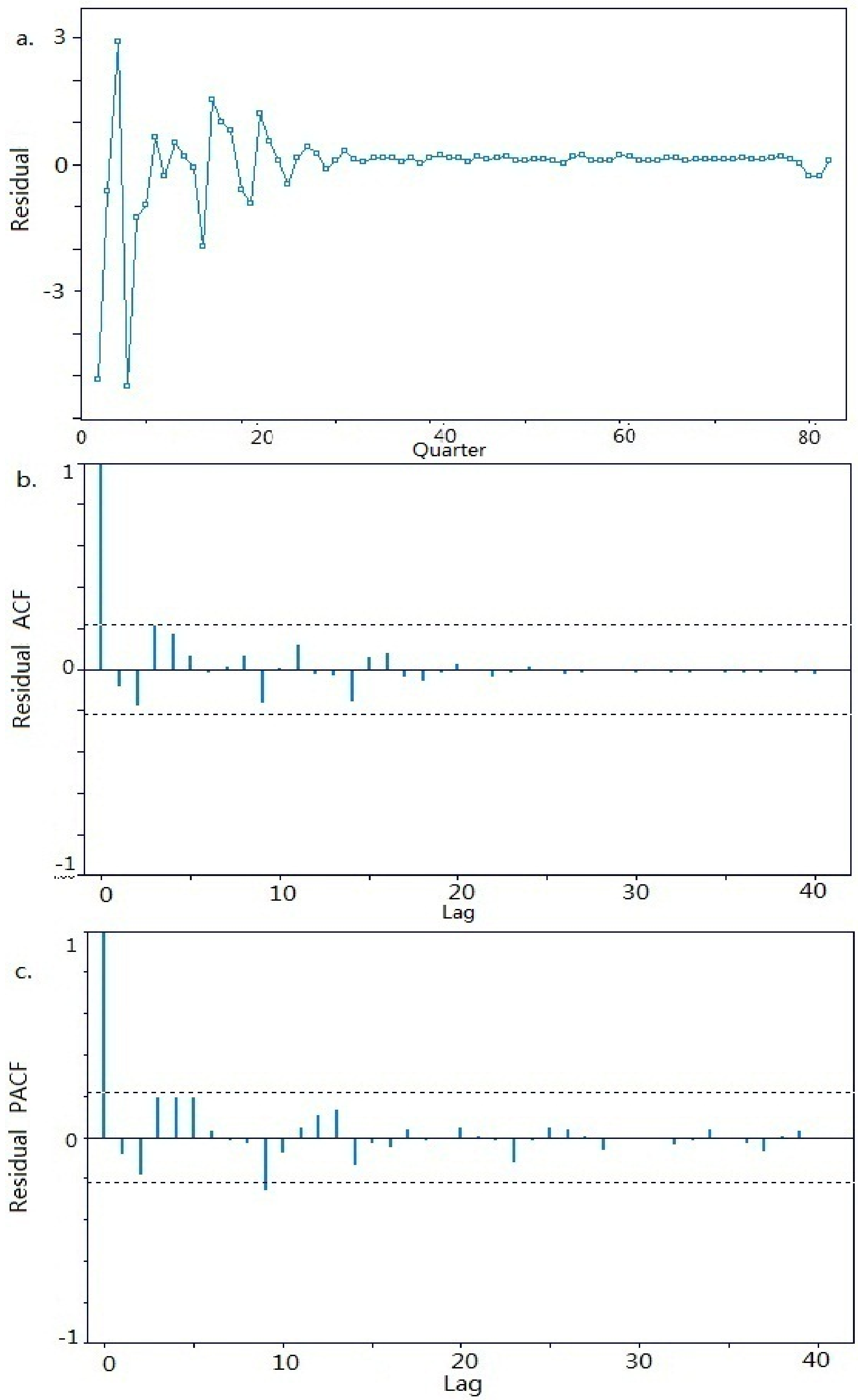

Figure 5.6. Diagnostics for the AR(3) Model. **a**, Residual plot; **b**, Residual ACF; **c**, Residual PACF.

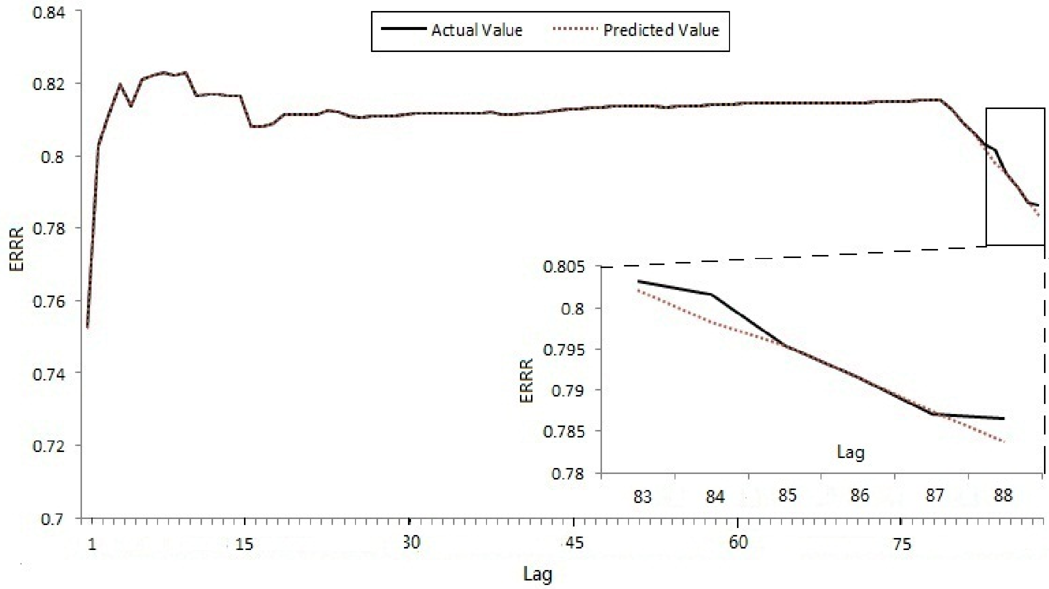

Figure 5.7 The Complete Data (Training Sample and Prediction Set) with Six Forecasts Appended to the Training Sample for Model Validation; Inset: Comparison of Six Forecasted ERRRs with the Prediction Set

Table 5.2 Numerical Comparison between the ERRRs (Predicted versus Observed)

Time	**Observed ERRR**	**Predicted ERRR**
2009:Q3	0.803278689	0.80206
2009:Q4	0.801670146	0.79822
2010:Q1	0.795377294	0.79533
2010:Q2	0.791500664	0.79152
2010:Q3	0.787055016	0.78741
2010:Q4	0.786531131	0.7837

5.2.2 Full Data Forecasting: λ_2/λ_1

We next extend the ARIMA modeling to the full data set of ERRR values. As in the previous case, we still take the Box-Cox transformation at λ=1.5, and difference at lag 3 and lag 1. The fitted model is also an AR(3) as follows.

$$X_t = .3447\ X_{t-1} + .5046\ X_{t-2} - .7525\ X_{t-3} + Z_t$$

WN Variance = .000026

Standard Error of AR Coefficients

.204349 .184433 .164831

The AICC statistic is -636.969, and the Ljung - Box statistic of residuals is not significant, as *p*-value = .53327. The plots of the residuals and their sample ACF and PACF are shown in Figure 5.8. Table 5.3 shows the 8 predicted values of the model, for the time period 2011:Q1 to 2012:Q4. The corresponding forecasted failure ratios (λ_2/λ_1) are: 0.28, 0.28, 0.28, 0.29, 0.30, 0.31, 0.32, and 0.33 (Table 5.3). The overall trend of the failure rate ratio $\rho_{12} = \frac{\lambda_2}{\lambda_1}$ is increasing with a mean of 0.30, which is larger than the reference population ratio ($\rho_{12}^0 = 0.183689$) (Figure 5.9). In other words, Group 2 consistently contributes more than its fair share of failures relative to Group 1 during the forecasted period.

5.2.3 Comparisons: All Groups

We extend our data analysis to the following two pairs: Group 2 versus Group 3 and Group 1 versus Group 3. Table 5.3 summarizes the results. Figure 5.10 depicts the temporal trends. All the results point to the same directions: smaller banks have a significantly and disproportionally higher survival rate than banks with larger total assets.

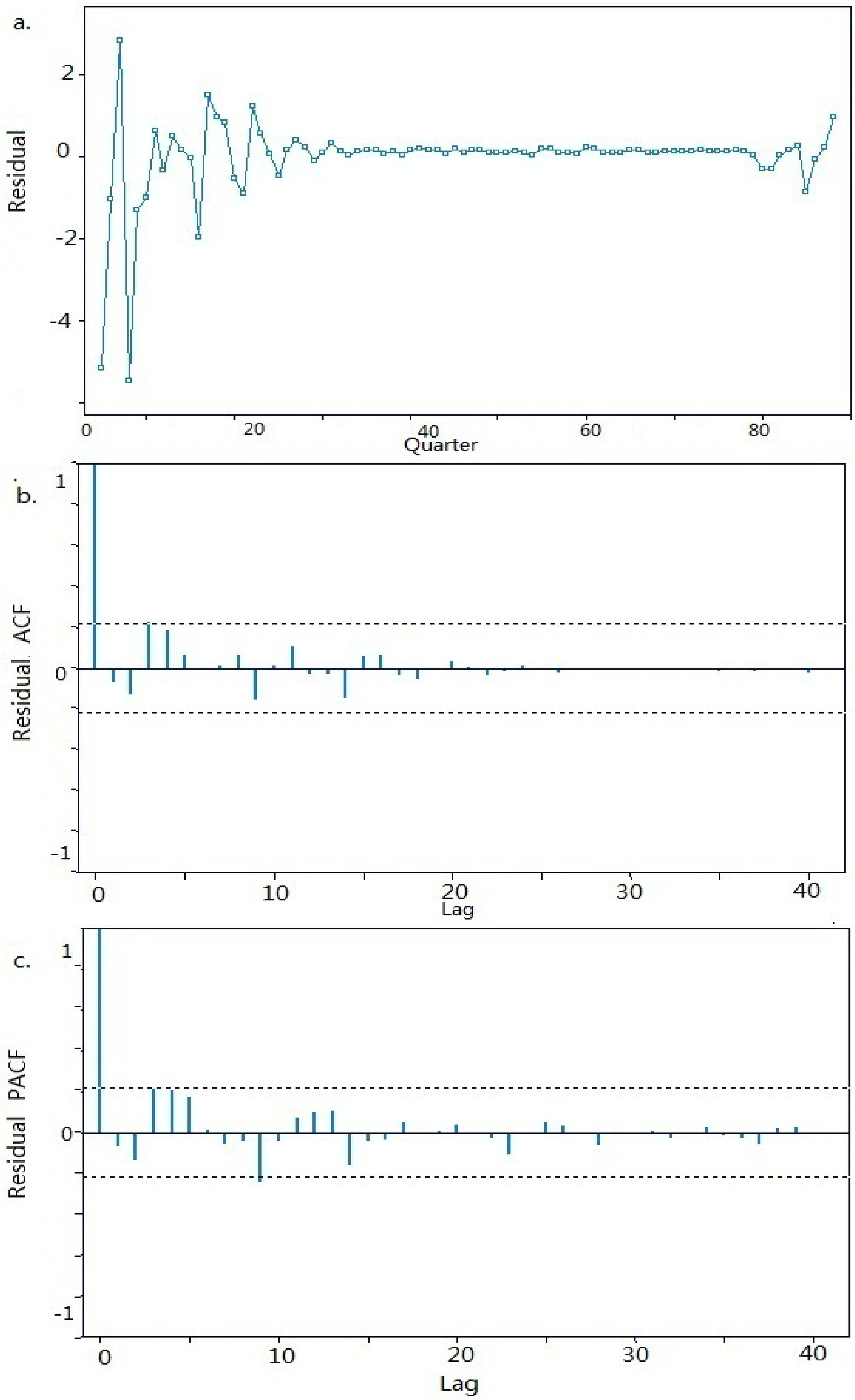

Figure 5.8. Diagnostics for the AR(3) Model for the Full Data. **a**, Residual Plot; **b**,

Residual ACF; **c**, Residual PACF.

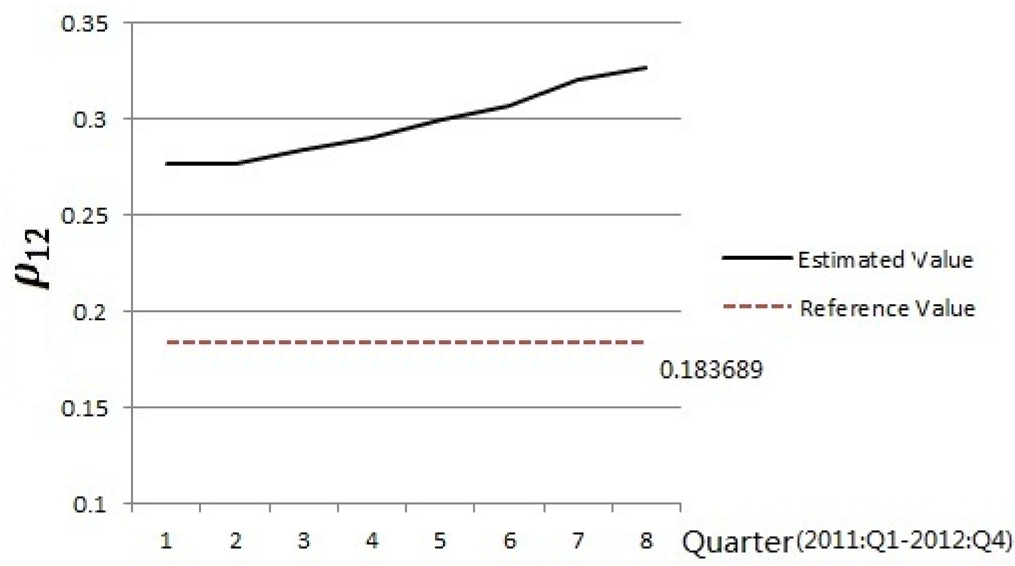

Figure 5.9 Comparisons of Predicted Values and Reference Value of Bank Ratio of Group1 versus Group 2

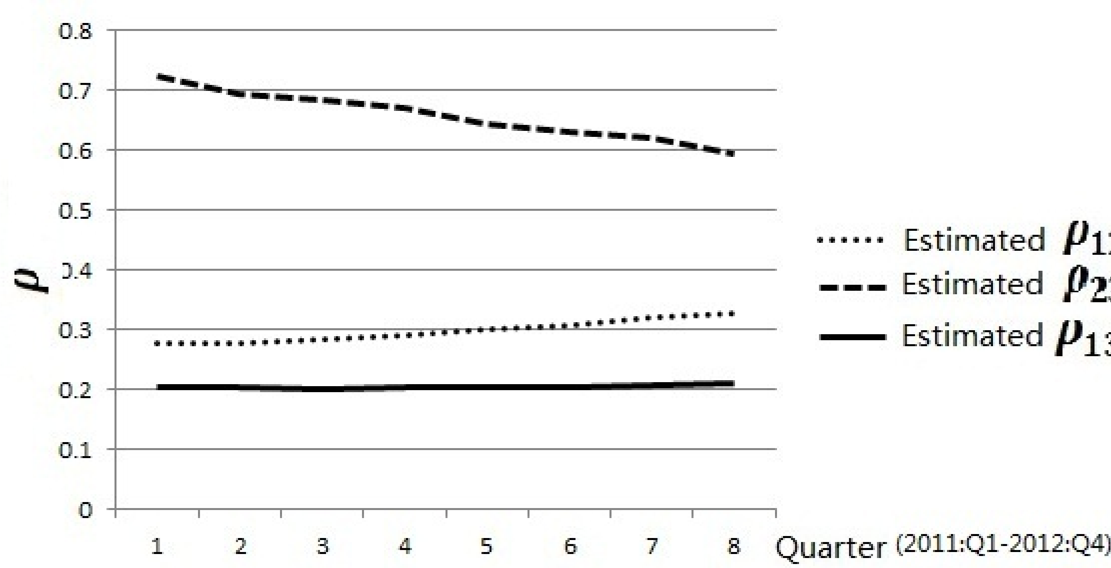

Figure 5.10 The Predicted Values of all Pairwise ERRRs during 2011:Q1 to 2012:Q4

Table 5.3 Numerical Values of the Predicted ERRRs of 2011:Q1 to 2012:Q4 of Group (1, 2), Group (2, 3) and Group (1, 3)

Time	Predicted ERRR (1,2)	Estimated ρ_{12} $\rho_{12}^{0} = 0.18$	Predicted ERRR(2,3)	Estimated ρ_{23} $\rho_{23}^{0} = 0.52$	Predicted ERRR(1,3)	Estimated ρ_{13} $\rho_{13}^{0} = 0.09$
2011:Q1	0.78301	0.2771229	0.58041	0.72292	0.83144	0.202733
2011:Q2	0.78293	0.27725339	0.58981	0.695461	0.83153	0.202602
2011:Q3	0.77889	0.28387834	0.59422	0.682878	0.83306	0.200394
2011:Q4	0.77525	0.28990648	0.59849	0.670872	0.83063	0.203905
2012:Q1	0.7692	0.300052	0.60802	0.644683	0.8301	0.204674
2012:Q2	0.76484	0.307463	0.61262	0.632333	0.83092	0.203485
2012:Q3	0.75728	0.32051553	0.6171	0.620483	0.8279	0.207875
2012:Q4	0.75337	0.32736902	0.62673	0.595583	0.82681	0.209468

CHAPTER 6

CONCLUSION

Coupled with the conditional test (Przyborowski and Wilenski, 1940), the empirical recurrence rates ratio, extended from the empirical recurrence rate (Ho, 2008), allows us to apply the well-known ARIMA modeling techniques to compare and forecast bank failures in the USA based on the most recent 22 years of financial data. The ERR and ERRR not only smooth and reduce the volatility of a financial system modeled by a stochastic process, but operate as a linking bridge between a classical time series and a point process. In this thesis, all the results of the statistical data analyses point to the same direction: Smaller banks have a significantly and disproportionally higher survival rate than banks with larger total assets. In other words, it seems that the statement: “Too big to fail.” is not supported by the most recent financial data.

APPENDIX

DATA

Table 1A: Quarterly Bank Failures Data from 1989:Q1 to 2010:Q4

Time	Group1	Group 2	Group3
1989:Q1	158	52	37
1989:Q2	78	6	12
1989:Q3	87	17	5
1989:Q4	64	10	8
1990:Q1	80	22	19
1990:Q2	93	15	9
1990:Q3	60	12	4
1990:Q4	49	10	8
1991:Q1	47	11	19
1991:Q2	55	11	12
1991:Q3	44	17	5
1991:Q4	33	7	10
1992:Q1	36	8	5
1992:Q2	37	9	6
1992:Q3	18	4	2
1992:Q4	25	18	12
1993:Q1	8	2	1
1993:Q2	14	2	2
1993:Q3	16	0	0

1993:Q4	4	1	0
1994:Q1	0	0	0
1994:Q2	5	1	0
1994:Q3	7	0	0
1994:Q4	1	1	0
1995:Q1	1	2	0
1995:Q2	2	1	0
1995:Q3	2	0	0
1995:Q4	0	0	0
1996:Q1	1	0	0
1996:Q2	2	0	0
1996:Q3	3	0	0
1996:Q4	0	0	0
1997:Q1	0	0	0
1997:Q2	0	0	0
1997:Q3	0	0	0
1997:Q4	1	0	0
1998:Q1	0	0	0
1998:Q2	1	0	0
1998:Q3	1	1	0
1998:Q4	0	0	0
1999:Q1	1	0	0
1999:Q2	1	0	0

1999:Q3	3	0	1
1999:Q4	2	0	0
2000:Q1	2	0	0
2000:Q2	1	0	0
2000:Q3	2	0	0
2000:Q4	2	0	0
2001:Q1	1	0	0
2001:Q2	1	0	0
2001:Q3	1	0	1
2001:Q4	0	0	0
2002:Q1	4	1	1
2002:Q2	1	1	0
2002:Q3	1	0	0
2002:Q4	2	0	0
2003:Q1	0	0	1
2003:Q2	1	0	0
2003:Q3	0	0	0
2003:Q4	1	0	0
2004:Q1	3	0	0
2004:Q2	1	0	0
2004:Q3	0	0	0
2004:Q4	0	0	0
2005:Q1	0	0	0

2005:Q2	0	0	0
2005:Q3	0	0	0
2005:Q4	0	0	0
2006:Q1	0	0	0
2006:Q2	0	0	0
2006:Q3	0	0	0
2006:Q4	0	0	0
2007:Q1	1	0	0
2007:Q2	0	0	0
2007:Q3	0	0	1
2007:Q4	1	0	0
2008:Q1	2	0	0
2008:Q2	1	0	1
2008:Q3	3	1	5
2008:Q4	5	5	7
2009:Q1	10	8	11
2009:Q2	12	8	4
2009:Q3	28	12	10
2009:Q4	25	9	11
2010:Q1	18	16	7
2010:Q2	22	13	7
2010:Q3	24	15	2
2010:Q4	22	7	1

Table 2A: The ERR Data of Bank Failures during 1989:Q1 to 2010:Q4

Time	Group1	Group 2	Group3
1989:Q1	158	52	37
1989:Q2	118	29	24.5
1989:Q3	107.6667	25	18
1989:Q4	96.75	21.25	15.5
1990:Q1	93.4	21.4	16.2
1990:Q2	93.33333	20.33333	15
1990:Q3	88.57143	19.14286	13.42857
1990:Q4	83.625	18	12.75
1991:Q1	79.55556	17.22222	13.44444
1991:Q2	77.1	16.6	13.3
1991:Q3	74.09091	16.63636	12.54545
1991:Q4	70.66667	15.83333	12.33333
1992:Q1	68	15.23077	11.76923
1992:Q2	65.78571	14.78571	11.35714
1992:Q3	62.6	14.06667	10.73333
1992:Q4	60.25	14.3125	10.8125
1993:Q1	57.17647	13.58824	10.23529
1993:Q2	54.77778	12.94444	9.777778
1993:Q3	52.73684	12.26316	9.263158
1993:Q4	50.3	11.7	8.8
1994:Q1	47.90476	11.14286	8.380952
1994:Q2	45.95455	10.68182	8
1994:Q3	44.26087	10.21739	7.652174
1994:Q4	42.45833	9.833333	7.333333
1995:Q1	40.8	9.52	7.04
1995:Q2	39.30769	9.192308	6.769231
1995:Q3	37.92593	8.851852	6.518519
1995:Q4	36.57143	8.535714	6.285714

1996:Q1	35.34483	8.241379	6.068966
1996:Q2	34.23333	7.966667	5.866667
1996:Q3	33.22581	7.709677	5.677419
1996:Q4	32.1875	7.46875	5.5
1997:Q1	31.21212	7.242424	5.333333
1997:Q2	30.29412	7.029412	5.176471
1997:Q3	29.42857	6.828571	5.028571
1997:Q4	28.63889	6.638889	4.888889
1998:Q1	27.86486	6.459459	4.756757
1998:Q2	27.15789	6.289474	4.631579
1998:Q3	26.48718	6.153846	4.512821
1998:Q4	25.825	6	4.4
1999:Q1	25.21951	5.853659	4.292683
1999:Q2	24.64286	5.714286	4.190476
1999:Q3	24.13953	5.581395	4.116279
1999:Q4	23.63636	5.454545	4.022727
2000:Q1	23.15556	5.333333	3.933333
2000:Q2	22.67391	5.217391	3.847826
2000:Q3	22.23404	5.106383	3.765957
2000:Q4	21.8125	5	3.6875
2001:Q1	21.38776	4.897959	3.612245
2001:Q2	20.98	4.8	3.54
2001:Q3	20.58824	4.705882	3.490196
2001:Q4	20.19231	4.615385	3.423077
2002:Q1	19.88679	4.54717	3.377358
2002:Q2	19.53704	4.481481	3.314815
2002:Q3	19.2	4.4	3.254545
2002:Q4	18.89286	4.321429	3.196429
2003:Q1	18.5614	4.245614	3.157895
2003:Q2	18.25862	4.172414	3.103448

2003:Q3	17.94915	4.101695	3.050847
2003:Q4	17.66667	4.033333	3
2004:Q1	17.42623	3.967213	2.95082
2004:Q2	17.16129	3.903226	2.903226
2004:Q3	16.88889	3.84127	2.857143
2004:Q4	16.625	3.78125	2.8125
2005:Q1	16.36923	3.723077	2.769231
2005:Q2	16.12121	3.666667	2.727273
2005:Q3	15.8806	3.61194	2.686567
2005:Q4	15.64706	3.558824	2.647059
2006:Q1	15.42029	3.507246	2.608696
2006:Q2	15.2	3.457143	2.571429
2006:Q3	14.98592	3.408451	2.535211
2006:Q4	14.77778	3.361111	2.5
2007:Q1	14.58904	3.315068	2.465753
2007:Q2	14.39189	3.27027	2.432432
2007:Q3	14.2	3.226667	2.413333
2007:Q4	14.02632	3.184211	2.381579
2008:Q1	13.87013	3.142857	2.350649
2008:Q2	13.70513	3.102564	2.333333
2008:Q3	13.56962	3.075949	2.367089
2008:Q4	13.4625	3.1	2.425
2009:Q1	13.41975	3.160494	2.530864
2009:Q2	13.40244	3.219512	2.54878
2009:Q3	13.57831	3.325301	2.638554
2009:Q4	13.71429	3.392857	2.738095
2010:Q1	13.76471	3.541176	2.788235
2010:Q2	13.86047	3.651163	2.837209
2010:Q3	13.97701	3.781609	2.827586
2010:Q4	14.06818	3.818182	2.806818

Table 3A: The ERRR Data of Bank Failures during 1989:Q1 to 2010:Q4

Time	Group1:2	Group 2:3	Group1:3
1989:Q1	0.752381	0.58427	0.810256
1989:Q2	0.802721	0.542056	0.82807
1989:Q3	0.811558	0.581395	0.856764
1989:Q4	0.819915	0.578231	0.861915
1990:Q1	0.813589	0.569149	0.85219
1990:Q2	0.821114	0.575472	0.861538
1990:Q3	0.822281	0.587719	0.868347
1990:Q4	0.822878	0.585366	0.867704
1991:Q1	0.822044	0.561594	0.855436
1991:Q2	0.822839	0.555184	0.852876
1991:Q3	0.816633	0.570093	0.855194
1991:Q4	0.816956	0.56213	0.851406
1992:Q1	0.817006	0.564103	0.852459
1992:Q2	0.816489	0.565574	0.852778
1992:Q3	0.816522	0.567204	0.853636
1992:Q4	0.808047	0.569652	0.847845
1993:Q1	0.80798	0.57037	0.848168
1993:Q2	0.80886	0.569682	0.848537
1993:Q3	0.811336	0.569682	0.850594
1993:Q4	0.81129	0.570732	0.8511
1994:Q1	0.81129	0.570732	0.8511
1994:Q2	0.811396	0.571776	0.851727
1994:Q3	0.81245	0.571776	0.852596
1994:Q4	0.811952	0.572816	0.85272
1995:Q1	0.810811	0.574879	0.852843
1995:Q2	0.810468	0.575904	0.853088
1995:Q3	0.810768	0.575904	0.853333
1995:Q4	0.810768	0.575904	0.853333

1996:Q1	0.810918	0.575904	0.853455
1996:Q2	0.811216	0.575904	0.853699
1996:Q3	0.811663	0.575904	0.854063
1996:Q4	0.811663	0.575904	0.854063
1997:Q1	0.811663	0.575904	0.854063
1997:Q2	0.811663	0.575904	0.854063
1997:Q3	0.811663	0.575904	0.854063
1997:Q4	0.811811	0.575904	0.854184
1998:Q1	0.811811	0.575904	0.854184
1998:Q2	0.811959	0.575904	0.854305
1998:Q3	0.811469	0.576923	0.854425
1998:Q4	0.811469	0.576923	0.854425
1999:Q1	0.811617	0.576923	0.854545
1999:Q2	0.811765	0.576923	0.854666
1999:Q3	0.812207	0.57554	0.854321
1999:Q4	0.8125	0.57554	0.85456
2000:Q1	0.812793	0.57554	0.854799
2000:Q2	0.812938	0.57554	0.854918
2000:Q3	0.81323	0.57554	0.855155
2000:Q4	0.81352	0.57554	0.855392
2001:Q1	0.813665	0.57554	0.85551
2001:Q2	0.813809	0.57554	0.855628
2001:Q3	0.813953	0.574163	0.855049
2001:Q4	0.813953	0.574163	0.855049
2002:Q1	0.8139	0.57381	0.854826
2002:Q2	0.813416	0.574822	0.854943
2002:Q3	0.813559	0.574822	0.855061
2002:Q4	0.813846	0.574822	0.855295
2003:Q1	0.813846	0.57346	0.854604
2003:Q2	0.813989	0.57346	0.854722

2003:Q3	0.813989	0.57346	0.854722
2003:Q4	0.814132	0.57346	0.854839
2004:Q1	0.814559	0.57346	0.855189
2004:Q2	0.814701	0.57346	0.855305
2004:Q3	0.814701	0.57346	0.855305
2004:Q4	0.814701	0.57346	0.855305
2005:Q1	0.814701	0.57346	0.855305
2005:Q2	0.814701	0.57346	0.855305
2005:Q3	0.814701	0.57346	0.855305
2005:Q4	0.814701	0.57346	0.855305
2006:Q1	0.814701	0.57346	0.855305
2006:Q2	0.814701	0.57346	0.855305
2006:Q3	0.814701	0.57346	0.855305
2006:Q4	0.814701	0.57346	0.855305
2007:Q1	0.814843	0.57346	0.855422
2007:Q2	0.814843	0.57346	0.855422
2007:Q3	0.814843	0.572104	0.854735
2007:Q4	0.814985	0.572104	0.854852
2008:Q1	0.815267	0.572104	0.855084
2008:Q2	0.815408	0.570755	0.854516
2008:Q3	0.815209	0.565116	0.851469
2008:Q4	0.81283	0.561086	0.847364
2009:Q1	0.809382	0.555315	0.841331
2009:Q2	0.80631	0.55814	0.840214
2009:Q3	0.803279	0.557576	0.837296
2009:Q4	0.80167	0.553398	0.833575
2010:Q1	0.795377	0.55948	0.831557
2010:Q2	0.791501	0.562724	0.830084
2010:Q3	0.787055	0.572174	0.831737
2010:Q4	0.786531	0.576329	0.83367

Table 3A: The Number of Solvent Bank and the Pairwise Ratios during 1989:Q1 to 2010:Q4

Time	Group 1	Group 2	Group 3	G2/G1	G3/G2	G3/G1
1989:Q1	11922	1410	792	0.118268747	0.561702128	0.066431807
1989:Q2	11855	1425	792	0.120202446	0.555789474	0.066807254
1989:Q3	11711	1434	801	0.12244898	0.558577406	0.068397233
1989:Q4	11583	1453	809	0.125442459	0.556779078	0.069843737
1990:Q1	11508	1431	791	0.124348279	0.552760307	0.068734793
1990:Q2	11417	1419	795	0.124288342	0.5602537	0.069633003
1990:Q3	11354	1383	800	0.121807293	0.578452639	0.07045975
1990:Q4	11285	1406	784	0.124590164	0.557610242	0.069472751
1991:Q1	11195	1390	789	0.124162573	0.567625899	0.070477892
1991:Q2	11108	1374	804	0.123694634	0.585152838	0.072380266
1991:Q3	11012	1389	795	0.126135125	0.572354212	0.07219397
1991:Q4	10864	1395	791	0.128405744	0.56702509	0.072809278
1992:Q1	10770	1367	798	0.126926648	0.583760059	0.074094708
1992:Q2	10660	1375	790	0.128986867	0.574545455	0.074108818
1992:Q3	10570	1382	783	0.130747398	0.566570188	0.074077578
1992:Q4	10478	1388	780	0.132468028	0.561959654	0.074441687
1993:Q1	10415	1356	765	0.130196831	0.564159292	0.073451752
1993:Q2	10299	1358	763	0.131857462	0.56185567	0.074084863
1993:Q3	10181	1354	772	0.13299283	0.570162482	0.075827522

1993:Q4	10060	1377	759	0.136878728	0.551198257	0.075447316
1994:Q1	9929	1369	763	0.13787894	0.557341125	0.076845604
1994:Q2	9808	1354	769	0.138050571	0.567946824	0.078405383
1994:Q3	9682	1359	758	0.140363561	0.557763061	0.07828961
1994:Q4	9530	1355	773	0.142182581	0.570479705	0.081112277
1995:Q1	9359	1316	763	0.140613313	0.579787234	0.081525804
1995:Q2	9271	1313	772	0.141244	0.587966489	0.083270413
1995:Q3	9117	1336	781	0.146539432	0.584580838	0.085664144
1995:Q4	8989	1329	793	0.147847369	0.59668924	0.088218934
1996:Q1	8900	1305	786	0.146629213	0.602298851	0.088314607
1996:Q2	8792	1266	764	0.14399454	0.603475513	0.086897179
1996:Q3	8684	1274	753	0.146706587	0.591051805	0.086711193
1996:Q4	8621	1272	757	0.147546688	0.595125786	0.087808839
1997:Q1	8531	1271	758	0.148986051	0.596380803	0.088852421
1997:Q2	8423	1263	727	0.149946575	0.575613618	0.086311291
1997:Q3	8341	1268	697	0.152020141	0.549684543	0.083563122
1997:Q4	8243	1278	696	0.155040641	0.544600939	0.084435278
1998:Q1	8121	1271	696	0.156507819	0.547600315	0.085703731
1998:Q2	8060	1279	688	0.158684864	0.53792025	0.085359801
1998:Q3	7986	1269	683	0.15890308	0.53821907	0.085524668
1998:Q4	7846	1267	677	0.161483559	0.53433307	0.086286006
1999:Q1	7782	1241	680	0.159470573	0.547945205	0.087381136
1999:Q2	7738	1248	666	0.161281985	0.533653846	0.086068752

1999:Q3	7676	1241	659	0.161672746	0.531023368	0.085852006
1999:Q4	7621	1238	662	0.162445873	0.534733441	0.086865241
2000:Q1	7576	1230	638	0.162354805	0.518699187	0.084213305
2000:Q2	7511	1240	642	0.1650912	0.517741935	0.085474637
2000:Q3	7398	1246	637	0.168423898	0.511235955	0.086104353
2000:Q4	7303	1267	639	0.173490346	0.504340963	0.087498288
2001:Q1	7223	1274	631	0.176381005	0.495290424	0.087359823
2001:Q2	7175	1285	645	0.179094077	0.501945525	0.08989547
2001:Q3	7125	1303	644	0.182877193	0.494244052	0.090385965
2001:Q4	7023	1321	654	0.188096255	0.495079485	0.093122597
2002:Q1	6963	1305	637	0.187419216	0.488122605	0.091483556
2002:Q2	6899	1312	638	0.190172489	0.486280488	0.092477171
2002:Q3	6827	1333	649	0.195254138	0.486871718	0.095063718
2002:Q4	6748	1343	660	0.199021932	0.491437081	0.097806758
2003:Q1	6707	1357	658	0.202325928	0.484893147	0.098106456
2003:Q2	6630	1378	677	0.207843137	0.491291727	0.102111614
2003:Q3	6584	1395	682	0.211877278	0.488888889	0.103584447
2003:Q4	6538	1397	674	0.213673906	0.482462419	0.10308963
2004:Q1	6493	1377	677	0.212074542	0.491648511	0.104266133
2004:Q2	6447	1396	674	0.216534822	0.482808023	0.104544749
2004:Q3	6386	1408	677	0.220482305	0.480823864	0.106013154
2004:Q4	6331	1414	686	0.223345443	0.485148515	0.10835571
2005:Q1	6281	1432	675	0.227989174	0.471368715	0.107466964

2005:Q2	6196	1466	672	0.236604261	0.458390177	0.108457069
2005:Q3	6193	1466	666	0.236718876	0.454297408	0.107540772
2005:Q4	6122	1496	683	0.244364587	0.456550802	0.111564848
2006:Q1	6070	1503	683	0.247611203	0.454424484	0.112520593
2006:Q2	6132	1503	692	0.245107632	0.460412508	0.11285062
2006:Q3	6102	1499	694	0.245657162	0.462975317	0.113733202
2006:Q4	6020	1518	700	0.252159468	0.46113307	0.11627907
2007:Q1	5988	1522	699	0.254175017	0.459264126	0.116733467
2007:Q2	5977	1496	704	0.250292789	0.470588235	0.117784842
2007:Q3	5944	1480	699	0.248990579	0.472297297	0.117597577
2007:Q4	5918	1480	699	0.250084488	0.472297297	0.118114228
2008:Q1	5890	1464	695	0.248556876	0.474726776	0.117996604
2008:Q2	5864	1461	679	0.24914734	0.464750171	0.115791269
2008:Q3	5801	1464	678	0.252370281	0.463114754	0.116876401
2008:Q4	5654	1506	712	0.266360099	0.472775564	0.125928546
2009:Q1	5577	1509	719	0.270575578	0.476474486	0.12892236
2009:Q2	5551	1495	714	0.269320843	0.477591973	0.128625473
2009:Q3	5470	1499	701	0.274040219	0.467645097	0.128153565
2009:Q4	5427	1495	690	0.275474479	0.461538462	0.127142067
2010:Q1	5339	1488	695	0.278703877	0.467069892	0.13017419
2010:Q2	5280	1459	682	0.276325758	0.467443454	0.129166667
2010:Q3	5231	1451	682	0.277384821	0.470020675	0.130376601
2010:Q4	5198	1401	682	0.269526741	0.486795146	0.131204309

BIBLIOGRAPHY

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

Brockwell P, and Davis R. (2002) *Introduction to Time Series and Forecasting,* 2nd editon Springer, New York.

Goldfeld, S., and Chandler, L. (1981) *The Economics of Money and Banking*, 8th edition,.

Ho, C.-H. (2008) *Empirical Recurrent Rate Time Series for Volcanism: Application to Avachinsky* Volcano, Russia, Volcanol Geotherm Res, **173**, 15-25.

Cryer J D. and Chan KS. (2008) *Time Series Analysis With Applications in R*. 2nd Edition. Springer, New York.

Krishnamoorthy, K. and Thomson, J. (2004) *A More Powerful Test for Comparing Two Poisson Means.* Journal of Statistical Planning and Inference, 119, 23-35.

Miller, S. and Jeon, Y. (2002) *Explaining U.S. Commercial Bank Births, Deaths, and Marriages*, the Eastern Economic Association meetings in Boston.

Mishkin, F. (2010) Money*, Banking and Financial Market*, 9th edition,

Shumway R H. and Stoffer D S.. (2006) *Time Series Analysis and Its Applications with R Examples.* Springer, New York.

Cowpertwait P S.P. and Metcalfe V.. (2009) *Introductory Time Series with R*. Springer, New York.

Przyborowski, J. Wilenski, H. (1940) *Homogeneity of Results in Testing Samples from Poisson Series*, Biometrika, 31, 313-323.

Smith, R. and Sidel, R. (2010) *Banks Keep Failing, No End in Sight*, Wall Street Journal. September 27, 2010

VITA

Graduate College
University of Nevada, Las Vegas

Fangjin Cui

Degree:
Bachelor of Engineering in Polymer Science and Engineering, 2003
Beijing University of Chemical Technology

Publication:
New Type of Piezo-damping Epoxy-matrix Composites with Multi-walled Carbon Nanotubes and Lead Zirconate Titanate, Materials Letters, Volume 62, Issue 23, 2008, 3859 - 3861.

Thesis Title: ARIMA Models for Bank Failures: Prediction and Comparison

Thesis Examination Committee:
Chairperson, Chih-Hsiang Ho, Ph. D.
Committee Member, Kaushik Ghosh, Ph. D.
Committee Member, Amei Amei, Ph. D.
Committee Member, Stephen M. Miller, Ph. D. Economics
Graduate Faculty Representative, Nasser Daneshvary, Ph. D. Economics