Journal of Finance and Bank Management
June 2016, Vol. 4, No. 1, pp. 15-33
ISSN: 2333-6064 (Print), 2333-6072 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/jfbm.v4n1a2
URL: https://doi.org/10.15640/jfbm.v4n1a2

Modeling Interaction between Bank Failure and Size

Chih-Hsiang Ho[1], Guancun Zhong[2], Fangjin Cui,[3] & Moinak Bhaduri[4]

Abstract

Finance is the industry that rules the world, and banks, by and large, represent a great percentage of that industry. After the financial crisis, public cries for a more stable financial system grew louder. It is quite likely that "too-big-to-manage" institutions may be reined in or even split up into smaller entities. Is the shift in banking a happy event? In this paper, we will characterize the relationship between the size of assets and bank failure from the beginning of the savings and loan crisis to the end of the Great Recession, bridged by a period called the Great Moderation.

Keywords: ARFIMA models; Conditional test; Empirical recurrence rate; Non-parametric methods.

JEL Classification Codes: E5, C1, C4, G2, G3.

1. Introduction

The 2007–2012 global recessions, sometimes referred to as the Great Recession, is a marked global economic decline that began in December 2007 and took a particularly sharp downward turn in September 2008. According to the U.S. National Bureau of Economic Research, the recession began in December 2007 and ended in June 2009. There is an intense debate about the causes and the vulnerabilities that amplified the crisis. In the face of financial crisis, the U.S. government provided cash and guarantees to financial institutions whose failure, it feared, might bring down the whole system, primarily around a handful of big players: A.I.G., Citigroup, Bank of America, and so on. In the 1980s, when the government rescued Continental Illinois Bank, Stewart B. McKinney, a Connecticut congressman, declared that the government had created a new class of banks - those too big to fail. The phrase returned and stuck (Dash, 2009). If the crisis had a single lesson, it is that the too-big-to-fail problem must be solved (Turner, 2010).

As public cries for a stable financial system grew louder, it is quite likely that banks will eventually be divided into smaller, more manageable entities and forced to hold more capital, moving the industry away from global laissez-faire business as usual and to a more traditional banking model (Foroohar, 2012). It is a convincing argument designed to make future bailouts unnecessary and restore market discipline, but there are skeptics (e.g., Krugman, 2010). Moreover, in the case of banks, how big is too big to fail? Moreover, how would you measure it? Nearly a century ago, the jurist Louis Brandeis railed against what he called the "curse of bigness." "Size, we are told, is not a crime," Brandeis wrote. "But size may, at least, become noxious by reason of the means through which it is attained or the uses to which it is put." Policy makers nowadays argue that the interconnectedness of modern finance, as much as the size of the players, is the real issue (Dash, 2009).

[1]Professor, Department of Mathematical Sciences, University of Nevada, Las Vegas. E-Mail:chho@unlv.nevada.edu
[2] Computer Scientist, BMM North America, Inc.
[3] CEO, ZXMM American LLC.
[4] Ph.D. Student, Department of Mathematical Sciences, University of Nevada, Las Vegas.

A bank fails when it can no longer cover its obligations (liabilities) with its assets and must file for bankruptcy. Washington Mutual (WaMu) failed on September 25, 2008, and reportedly had over $30 billion in assets at the time of the failure. Two years after WaMu failed, the number of bank failures significantly increased compared to the previous six years, during which period only around 40 banks failed. In retrospect, the number of bank failures has increased dramatically over the last 25 years; nearly 3000 occurred between 1985 and 2010. 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 (Levin and Coburn, 2011).The precarious situation has been nicely depicted by Admati and Hellwig (2013) among many other researchers, who argue how the $8 trillion debt incurred by the top five U.S. banks during the crisis could have been even worse, had accounting rules prevalent in Europe been in vogue. On the other hand, a help from the government might prove hugely detrimental to the large national economic interests – chilling evidence of which is provided by countries such as Ireland and Spain. The situation is aggravated by the advent of an extremely complex and interconnected financial structure, where failures of big banks have the potential to trigger a large scale global economic meltdown. This is in stark contrast to the pre-2007 era when a country's banking system was largely disjoint from any others' and hence the effect of a failure was largely localized.

Since the 1960s, several economists have conducted studies on forecasting bankruptcy and firm failures. Altman (1968), perhaps the first researcher to examine firm failures, used discriminate analysis to predict the failures of firms from different industries. Bank failure prediction models, also called early warning systems, were developed using statistical methodologies and numerical econometric models. Martin (1977), the first researcher to consider bank failures, employed logistic regression to predict bank and firm failures, and found that the relevance of conventional measures of bank soundness differs over the business cycles. Crowley and Loviscek (1990) considered several different model specifications and concluded that logit and probit models outperform linear probability and discriminate analysis models, using annual data of small bank failures in 1984. Lane *et al*. (1986) applied the Cox (1972) proportional hazard model to the analyses of bank failures using the corporate bankruptcy data over the period 1962 – 1992 in the U.S. Shumway (2001) demonstrated that the hazard model outperforms the traditional bankruptcy models (Altman, 1968; Zmijewski, 1984), and that a new hazard model combining both accounting and market variables can substantially improve the accuracy in predicting corporate bankruptcy.

Early warning systems can predict the likelihood and timing of bank failures, and researchers continue to improve the accuracy and predictive power of these models. Consequently, bank failure studies educate and inform federal regulators and policy makers, who can then provide strategies to increase the stability for the banking industry. For instance, should the too-big-to-manage institutions be reined in or divided, allowing smaller entities to focus on what they do best? Will such a shift in banking in the U.S. be the right course? Perhaps, there is a magic "size" that keeps banks solvent. The difficulties in modeling are not unknown. For example, it is an article of faith in much applied work that disturbance terms are IID - Independent and Identically Distributed - across observations. Why would that be so for banks?

However, there is an alternative that should always be kept in mind. Disturbances are DDD - Dependent and Differently Distributed - across subjects (Freedman, 2009). In any of the bank failure prediction models, the disturbance terms could easily be DDD. In this paper, the intention is to minimize our model assumptions and maximize the utilities of our modeling output in quantifying the relationship, if any, between bank assets size and the likelihood of failure. Due both to simplicity and to their greater robustness, non-parametric methods are seen by some statisticians as leaving less room for improper use and misunderstanding.

This study commences with an origination of the data set. In Section 2, raw data of bank failures are surrogated by an empirical recurrence rate time series designed to fingerprint the temporal pattern of each data set. More graphical and exploratory data analyses are presented in Section 3 using the empirical recurrence rates ratio plots, proposed for pair-wise comparisons. In Section 4, we outline the modeling strategy, present the techniques, and highlight the analytical findings. The final section concludes our work.

2. Preliminary

Savings and Loans were specialized banks that used low-interest, but federally-insured, deposits in savings accounts to fund mortgages. The savings and loan crisis of the 1980s and 1990s (commonly dubbed the S & L crisis) was the failure of about 747 out of the 3,234 savings and loan associations in the U. S. In addition to the failures specific to the S & L crisis, many other banks failed in the near future. Between 1980 and 1994, more than 1,600 banks insured by the Federal Deposit Insurance Corporation (FDIC) were closed or received FDIC financial assistance. Characteristics of bank failures may differ over business cycles, which justifies the choice of the following observation period for this study: 1980:Q1 to 2011:Q4, from the beginning of the S&L crisis to the end of the Great Recession, bridged by a period of unprecedented market calm that economists called the Great Moderation.

2.1 Count Data

The numbers of solvent and failed banks in the U. S. during 1980:Q1 to 2011:Q4 are obtained from the FDIC website (www.fdic.gov), which provides information on banks by name, location, charter type, total assets, and other characteristics. We tallied the number of bank failures on a quarterly basis, and documented a total of 3,212 failed banks during the entire observation period of 128 quarters. Meanwhile, there were 7,300 banks that remained solvent at the end of 2011:Q4. Without loss of generality, we decided to use the "total assets" of the banks to analyze the association between the likelihood of failure and the bank size. Of course, the following classification method is flexible enough to accommodate any plausible alternatives for quantifying the bank size using expert opinions. 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 amounts of goods and services.

The purchasing power of money – the real value in the internal medium of exchange and unit of account in the economy, changes over time. The Consumer Price Index (CPI) is used to connect nominal values to real values. The total assets of banks are reported in terms of nominal price. To make the total assets in different time periods comparable, the total assets of banks are converted to the real values using (e.g., see Mankiw, 2002): Adjusted total assets = (CPI_b/CPI_i) x $(\text{Total assets})_i$ (1) where the Total assets is the nominal total asset at time i (the month a failure was reported); the CPI_i is the CPI at the it month that bank failed; and the CPI_b is the CPI for the base month (taken as January 2011 in this study). Subsequently, the five-number-summary of the adjusted total assets for all the 10,512 banks in the data set is (in millions):

(Min, Q_1, Q_2, Q_3, Max) = (3.0, 67.4, 150, 360, and 1,810,000). This classification scheme for a two-way contingency table makes a good start for straightforward group comparisons. We set the banks with adjusted assets (in millions) lower than the first quartile (Q_1 = 67.4) as our "small" banks group (also known as G_1). Banks with adjusted assets between Q_1 and Q_2 are referred to as "medium" banks group (a. k. a. G_2). Likewise, banks with adjusted assets between Q_2 and Q_3 are referred to as "large" banks group (a. k. a. G_3), while the remaining banks are referred to as "grand" banks group (a. k. a. G_4). Counts of bank failures by "status" (coded as "Yes" for failed and "No" for solvent) and "group" (1 through 4) are summarized in Table 1. Note that we deliberately set an equal number of banks for each bank size on the first day (January 1, 1980) of the observation period for the data analysis of bank failures. This balanced classification is not necessary if the groups are well defined, say, banks in the U.S. versus that in Japan. However, in this case, the ratio of bank population of these countries needs to be estimated to justify the direction of the difference for all of the comparisons to be developed in this paper.

Table 1: Counts of banks by status and group during 1980:Q1 - 2011:Q4

		Group				
		1 [3.0, 67.4]*	2 (67.4, 150]	3 (150, 360]	4 (360, 1810000]	Totals
Status	Yes	1172	663	590	787	3212
	No	1456	1965	2038	1841	7300
	Totals	2628	2628	2628	2628	10512

* In millions

2.2 Empirical Recurrence Rate

We first present the quarterly bank failures with two conventional time series plots: Figure I represent the total counts and Figure II tracks failures for each group. These time series plots show that bank failures occur in spurts or clusters, reflecting the performance of the banking industry during the S&L crisis and the Great Recession. It also poses a challenge to a serious time series modeler, because there are a lot of zero quarterly failure counts during a portion of the Great Moderation, a period of unprecedented market calm. A glimpse of Figure II appears messy, but it is clear that there are more failures for small banks during the period of the S & L crisis. In addition, during the Great Recession period, the situation reversed.

Consequently, bank size would be a significant predictor if one uses a simple logistic regression to model the probability of bank failure based on the data either from the S & L crisis or from the Great Recession. However, the model built from one would fail horribly, if it were used to predict the other.

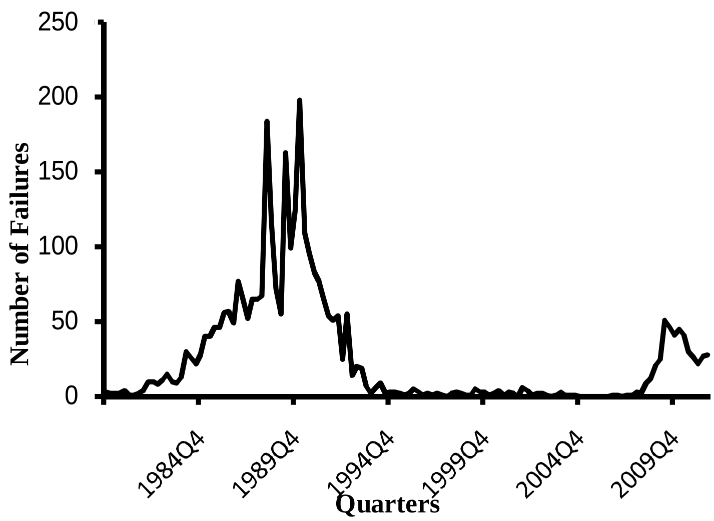

Figure I. Number of bank failures from 1980:Q1 to 2011:Q4.

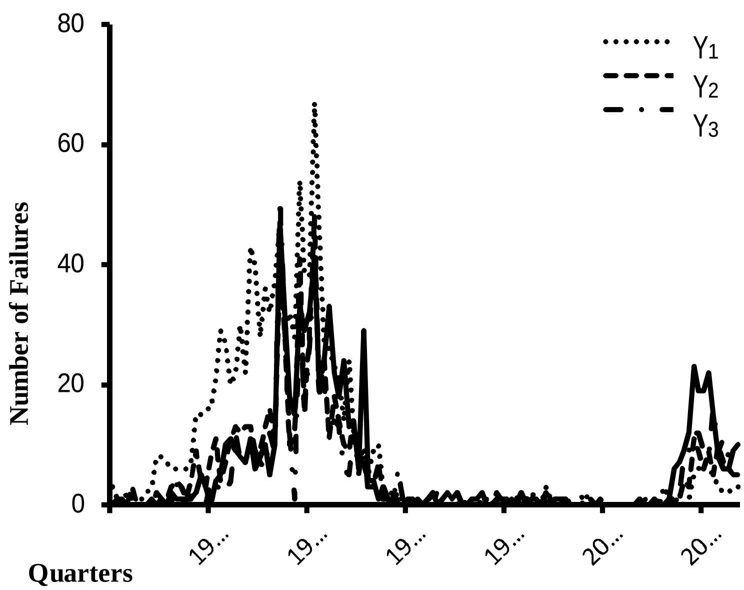

Figure II. Numbers of bank failures from 1980:Q1 to 2011:Q4 by groups.

Next, a smoothing technique, referred to as the "empirical recurrence rate" or "ERR" for the transformed data, is considered. The technique has been introduced by Ho (2008) and popularized by Tan, Bhaduri and Ho (2014), Ho and Bhaduri (2015) and Ho (2016) in studies related to sandstorm, earthquake and mining accidents modeling.

Success in these areas encourages us to implement analogous techniques to the financial situation too, especially due to experts' widely held beliefs that with the banks' excessive dependence on deposits, fragility in the banking system is a rueful and unavoidable consequence. These events are thus intrinsically similar to unforeseen natural calamities such as earthquakes or hurricanes.

Let $t_1, \dots, t_n$ be the time of the n ordered bank failures during an observation period $(0,T)$ from the first occurrence to the last. Then a discrete ERR time series $\{z_l\}$ is generated sequentially at equidistant time intervals $h, 2h, ..., lh, ..., Nh\ (=T)$. If 0 is adopted as the time-origin and h as the time-step, then we regard z_l as the observation at time $t = lh$. Specifically, the smoothed series is defined as follows:

$z_l = n_l / lh =$ Total number of bank failures in $(0, lh)/lh$, (2)

Where $l = 1, 2, ..., N$. Note that z_ℓ evolves over time by updating the maximum likelihood estimator (MLE) of the mean unit rate, λ, at each time step, if the underlying process observed in $(0, \ell h)$ is a homogeneous Poisson process (HPP). The ERR curve should be approximately flat for a typical HPP, where the correlation function for z_j and z_{j+k} is

$$\rho_{j,j+k} = [\,j/(\,j+k)]^{1/2}, \text{ for } j = 1, 2, ..., n-1, \quad k = 0, 1, ..., n-j. \qquad (3)$$

Fixing *k*, the correlation converges to 1, indicating that the ERRs converge to a constant which is the recurrence rate of the underlying HPP. In general, higher correlation among the ERRs reflects a consistent pattern of the process and promotes better predictive ability for the fitted model. In order to accommodate the complexity of the data, we do not assume that the structure of the underlying process is fixed.

2.3 ERR Plot

The merit of the ERRs is first illustrated in Figure III, where we plot the quarterly bank failures during 1980:Q1 to 2011:Q4 (labeled as Y) and its corresponding ERRs (labeled as Z) in the same graph. It's clear that the raw data are inevitably volatile, with quarterly failures ranging from 0 to 198. In contrast, the ERR tends to smooth the data because of its nature as a cumulative function. It slowly grows with the S&L crisis, and peaks after the crisis. Although it declines, it appears never appears to capture the extremely low levels of bank failures during the Great Moderation. It barely registers the Great Recession at all. Interesting enough, in the ending part, the ERRs show a little rebound which precisely explains the events that occurred in the Great Recession. Because the number of bank failures was much lower than that of the S&L crisis, the rebound on ERRs is very limited. All of the above observations point to a curve that can retain the trend of the raw data, and is quite possible to fit.

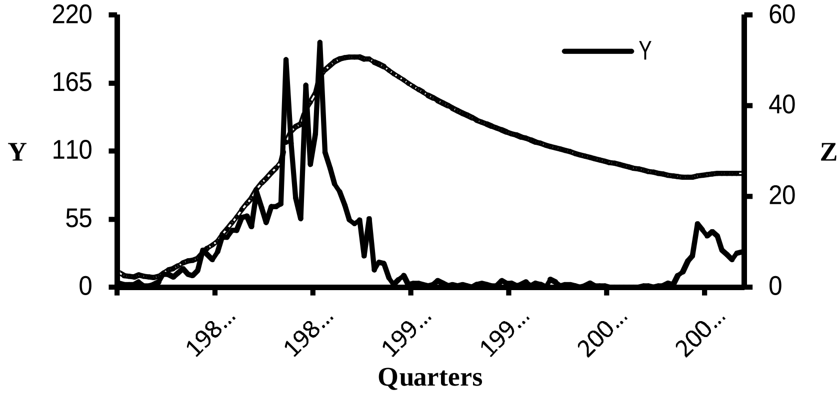

Figure III. Plots of bank failures (Y) and ERRs (Z).

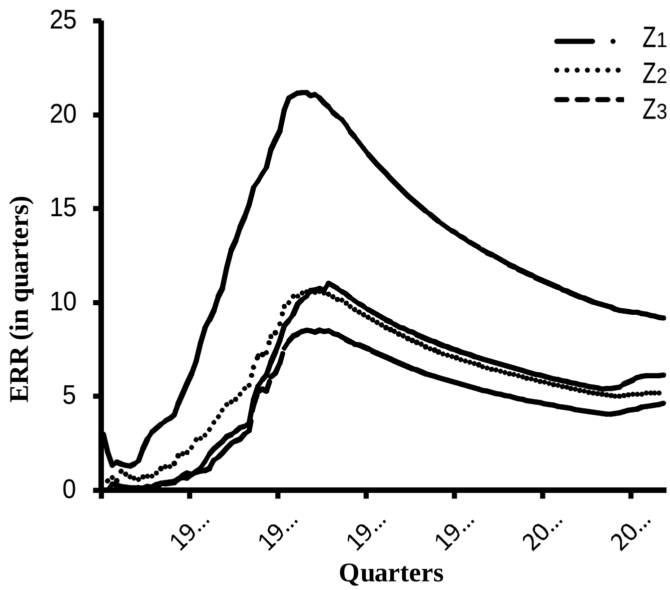

Figure IV. ERR plots by groups.

The magic of the ERR plot extends to Figure IV, plotting bank failures of the four groups G_1 - G_4 (labeled as Z_1, Z_2, Z_3 and Z_4, respectively). It reveals that the group of small banks (i.e., G_1) has the highest cumulative failure rate while large banks group (G_3) has the lowest among all four groups at any quarter throughout the entire observation period. If we take a closer look at G_2 and G_4, we see that medium banks group (G_2) has higher failure rate than the grand banks group (G_4) until 1991:Q3, a single change-point for G_2 vs. G_4. As in the sports of running, a winner may not be ahead all the time during a race. For instance, if we start an ERR fresh from 1996:Q1 to 2011:Q4 (Figure V), the ERR shows its appeal again. From Figure V, in the beginning, most of the ERRs are quite small, a reflection of the stable period after the S&L crisis. A closer look at the ending part of the graph shows that there are rapid increases for all groups from 2008, especially for the grand banks group (G_4). The dramatic rise is a reflection of the Great Recession, when banks in G_4 failed the most which is exactly what we noted earlier. Further developments for pair-wise comparisons follow.

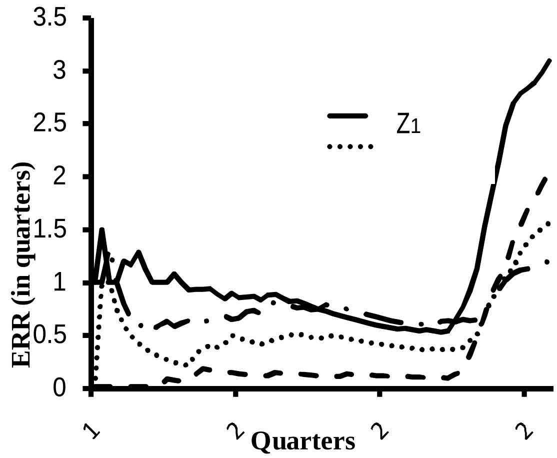

Figure V. ERR plots by groups for period, 1996:Q1 to 2011:Q4.

3. Modeling Fundamentals

In probability theory and statistics, the Poisson distribution is a discrete distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event. Applications of the Poisson distribution can be found in many fields focusing on certain random variables *N* that count, among other things, the number of discrete occurrences (sometimes called "arrivals") that take place during a time-interval of given length. The Poisson distribution arises in connection with a Poisson process, a stochastic process which counts the number of events and the time that these events occur in a given time interval.

The basic form of Poisson process is a continuous-time counting process $\{N(t), t \geq 0\}$ that possesses specific properties (e.g., see Beichelt, 2006). This process is characterized by a rate parameter λ, also known as "intensity," such that the number of events in time interval $(t, t + \tau]$ follows a Poisson distribution with associated parameter $\lambda\tau$. Just as a Poisson random variable is characterized by its scalar parameter λ, an HPP is characterized by its rate parameter λ, which is the expected number of "events" or "arrivals" that occur per unit time. In general, the rate parameter may change over time; such a process is called a non-homogeneous Poisson process (NHPP) or inhomogeneous Poisson process.

In this case, the generalized rate function is given as $\lambda(t)$, making the expected number of events between time *a* and time *b* : $\lambda_{a,b} = \int_a^b \lambda(t)dt$. Thus, the number of arrivals in the time interval $(a, b]$, given as $N(b) - N(a)$, follows a Poisson distribution with associated parameter $\lambda_{a,b}$. An HPP may be viewed as a special case when $\lambda(t) = \lambda$, a constant rate. An NHPP generalizes an HPP, and is often appropriate for modeling a series of events that occur over time in a non-stationary fashion. NHPPs have been used to model event occurrences in a variety of applications. A major difficulty with the NHPP is that it has infinitely many forms for the intensity function. An NHPP, referred to as a power law process, has a monotonic intensity, and has proved versatile in the reliability studies of repairable systems. In certain cases, even when the use of parametric methods is justified, non-parametric methods may be easier to use.

The parameter μ of a Poisson distribution requires careful definition. Typically, it needs to be described as a rate; for example, the average number of customers who buy a particular product out of every 100 customers who enter the store. For motor vehicle crashes, however, the rate parameter may be defined in many different ways. The time scale often is included in the definition as well, such as crashes per 100,000 kms per year, while the rate of tropical cyclones refers to the cyclone season from November to April in Northeastern Australia (Dobson and Barnett, 2008). Moreover, when dealing with data on frequencies, we count how many times something happened, but we have no way of knowing how often it did not happen (e.g. lightning strikes, bankruptcies, deaths, births).

This is in contrast to count data on proportions, where we know the number doing a particular thing, but also the number not doing that thing (e.g. the proportion dying, sex ratios at birth, and proportions of different groups responding to a questionnaire). In other words, more banks failing in the U.S than in Japan does not mean that American banks are more likely to fail unless the bank population sizes are equal for both countries. Thus, in Table 1, we set an equal number of banks for each bank size on the first day (January 1, 1980) of the observation period to ease our model development and interpretation.

3.1 Empirical Recurrence Rates Ratio

Let X_1 and X_2 be independent observations from Poisson (λ_1) and Poisson (λ_2) distributions, respectively. A well-known method of testing the difference of two Poisson means is the conditional test (Przyborowski and Wilenski, 1940), referred to as the C-test. It is based on the fact that the sum, $S = X_1 + X_2$ follows a Poisson distribution with rate parameter, $\lambda_1 + \lambda_2$, and the conditional distribution of X_1 given S = s is distributed as Binomial(s, p_{12}), where $p_{12} = \lambda_1 / (\lambda_1 + \lambda_2) = \omega_{12} / (1 + \omega_{12})$ with $\omega_{12} = \lambda_1 / \lambda_2$. Thus, for the C-test, testing $H_0 : \lambda_1 = \lambda_2$ vs. $H_a : \lambda_1 \neq \lambda_2$ is equivalent to testing $H_0 : \omega_{12} = 1$ vs. $H_a : \omega_{12} \neq 1$, which is also equivalent to testing $H_0 : p_{12} = 0.5$ vs. $H_a : p_{12} \neq 0.5$.

Motivated by the simplicity of the C-test and the smoothing power of the ERR, we propose the following "empirical recurrence rates ratio" time series, to be referred to as "ERRR" to measure the bank failure rates ratio between groups, paired as G_i and G_j.

First, we partition the observation period (0, *T*) into *N* equidistant time intervals $h, 2h, ..., lh, ..., Nh \ (= T)$ for a given time-step, *h*. The ERRR, $R_{ij,l}$, at time $t = lh$ is then generated as follows:

$R_{ij,l} = n_{il} / (n_{il} + n_{jl})$, for $n_{il} + n_{jl} > 0$ (4)

where:

n_{il} = Total number of bank failures for G_i in $(0, lh)$;

n_{jl} = Total number of bank failures for G_j in $(0, lh)$;

$1 \leq i < j \leq 4$ and $l = 1, 2, ..., N.$

If 0 is adopted as the time-origin, then we regard $R_{ij,l}$ as the observation at time, $t = lh$, discarding the burn-in period where $n_{il} + n_{jl} = 0$. Also, if both of the targeted processes are HHPs, then at every time-step, the ERRR updates the maximum likelihood estimator (MLE) of p_{ij}, which can be used to find the MLE of $\omega_{ij} (= \lambda_i / \lambda_j)$ using the invariance property of the MLE. The ERRs and ERRRs are unconventionally created to be cumulative to offset the potential of creating a time series with a lot of detrimental but seasonal zero values through a discretization process. For instance, point processes that characterize small recurrence rates or, in particular, exhibit seasonality with a lot of off-season zero counts such as sand-dust storms and hurricane data, are safeguarded as well. While sheltered by the versatility provided by the time-step parameter, data analysis with a counter intuitive time-step should be supported by a rigorous sensitivity analysis.

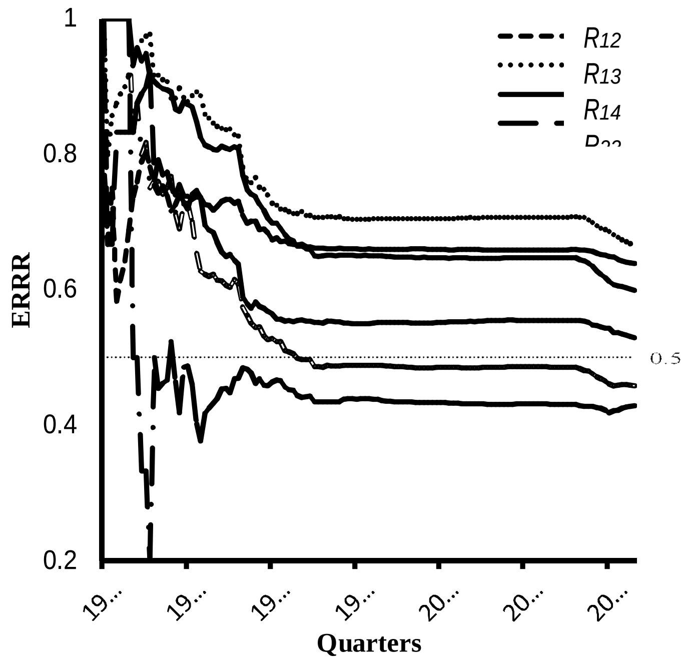

Figure VI. ERRR plots for pair wise comparisons.

3.2 ERRR Plot

Intuitively, every $ERRR_{ij}$ (= R_{ij}) is simply the (absolute) frequency of failures in G_i normalized by the total number of failures from both groups, a relative frequency cumulated at each time-point. Mathematically, $ERRR_{ij} = ERR_i/(ERR_i + ERR_j)$. Therefore, any information accrued from comparing any pair of the ERR curves in Figure IV, is reproduced and displayed by a single ERRR time series plot (ERRR plot). In addition, we set the larger banks group, G_j, as the baseline group in each pair-wise comparison. Consequently, $R_{ij} = 0.5$ means that there are same numbers of bank failures in both groups up to that time-point. If $R_{ij} < 0.5$, there are more bank failures in G_j than in G_i, while $R_{ij} > 0.5$ indicates that, so far, the smaller banks group fails more. Therefore, we add a reference line at $R_{ij} = 0.5$ in Figure VI, plotting the ERRR time series for all six pair-wise comparisons. These ERRR curves all start at lag-3, because at least one of them starts with two undefined ERRRs.

Evidenced by all the figures discussed so far, in comparing the number of bank failures collected at the finish line, the ranking is $G_1 > G_2 > G_4 > G_3$ in the period of the S&L crisis. In contrast, it is $G_4 > G_3 > G_2 > G_1$ during the Great Recession. Table 1 also confirms that the order is $G_1 > G_4 > G_2 > G_3$ for the entire observation period. In Figure VI, all of the smaller banks groups in each comparison, except G_3, surpassed its competitor early in the race, but there was a catch-up in production which accelerated during the Great Recession.

Unfortunately, only G_4 was able to cross the reference line in the first quarter of 1992, and stayed ahead of G_2 thereafter. Here, curve R_{34} deserves a bit more attention, because it says that the large banks group is consistently more robust than the grand banks, throughout the entire race. However, the validity of a constant reference line puzzled us given that the smallest banks group lost a large number of members early on which could change the population ratios and distorted the picture. Therefore, in Figure VII, we update and let each reference line move with its companion. For the reference lines that involve G_1, labeled with p_{1j} (j = 2, 3 and 4; in Figure VII), they did all dip below 0.5 and begin to stabilize at about one third of the way. There was a noticeable rebound by G_4 for p_{14} in the end. Line p_{24} grew suddenly above 0.5 in the end, while p_{34} edged up early on. All of the observed changes are minor. Hence, a constant reference line is acceptable and what has been ascertained thus far is valid. Descriptive statistics tell us about the data that we happen to have. For the remainder of the paper, we turn to prediction, which is more complicated than description. On the other hand, if we make a series of predictions and evaluate them against data, it may be possible to show that the ERRR/ERR time series is stable enough for model to be helpful.

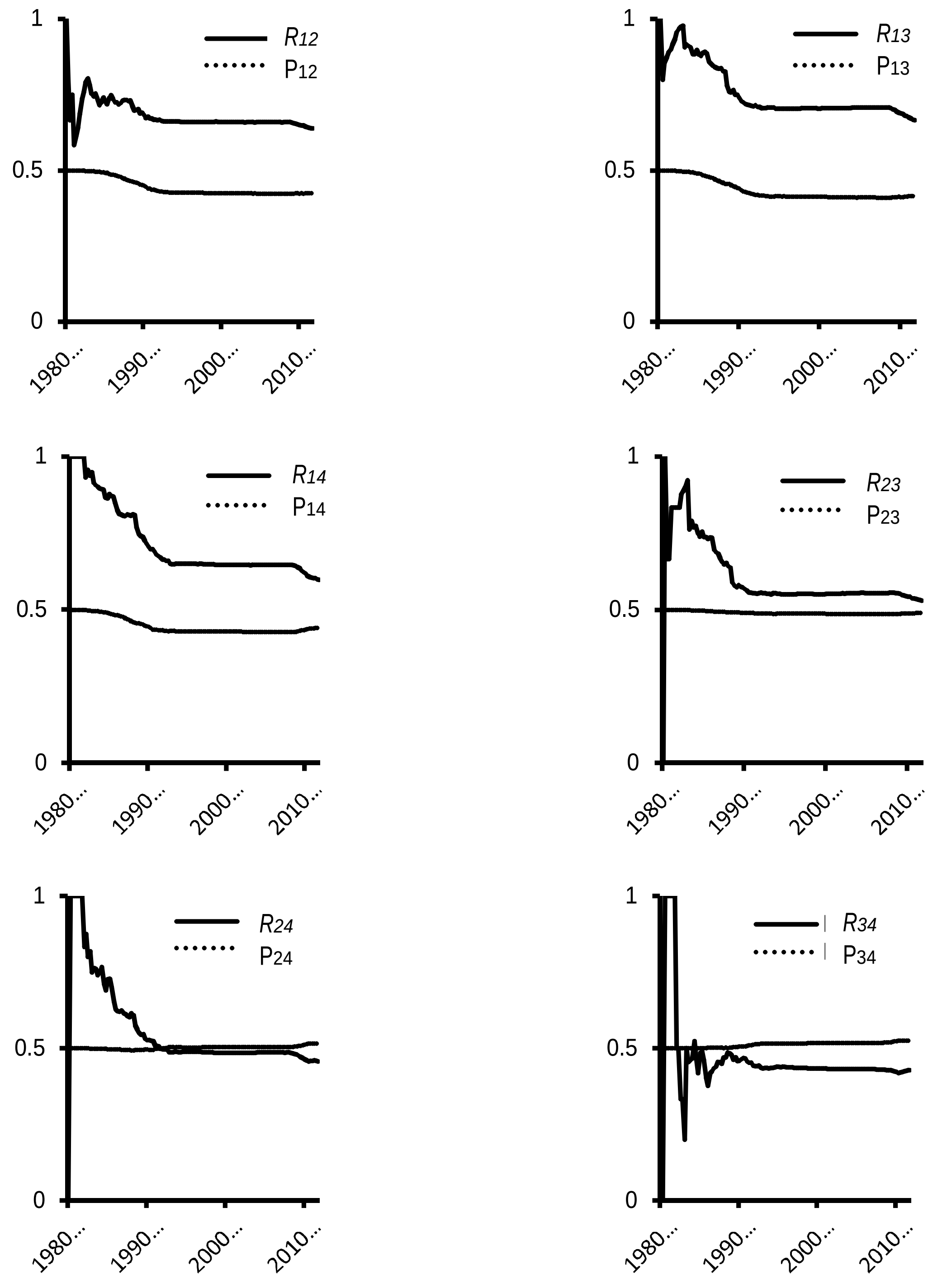

Figure VII. Plots of R_{ij} vs. p_{ij} (the reference line) for 1 = i < j = 4 during 1980:Q1 and 2011:Q4.

4. Modeling

The ERRR time series offers the possibility of predicting bank failure rates ratio using time series modeling tools. For instance, we start at time T , observe a sample $(R_1, \ldots, R_T)$ of an ERRR time series and want to predict a value $R_{T+k},\ k \geq 1$. This forecast is made at (forecast) origin T for lead-time (or forecast horizon) k .

In a regression situation, we let X denote the time index and R the response value, and then use the fitted regression model to obtain R_{T+k}. However, a regression model assumes that the error terms and consequently, the responses are uncorrelated and this is not a reasonable assumption for a process that evolves over time. On the other hand, the autoregressive integrated moving average (Box and Jenkins, 1976), referred to as "ARIMA" and other class of models have been found useful to represent the serially dependent relationship of many time series encountered in practice.

4.1 Strategy

A coherent modeling strategy increases the chance of building models that are useful approximations and can provide good predictions. At Stage 1, ARIMA modeling techniques are applied to a major portion of the ERRR time series, referred to as a "training sample." Three processes are distinguished in this stage: (i) model identification, (ii) parameter estimation and (iii) prediction and comparison of future values with a set of holdout-ERRRs, referred to as the "prediction set." The predictability of all of the candidate models can be assessed, and consequently, the pool of the selected models is narrowed down to produce the most useful model, say ARIMA(2, 1, 1), that fits the training sample best and adequately reproduces the prediction set. For Stage 2, we fit the final model, ARIMA (2, 1, 1), attained in Stage 1 to the entire data set. If the model passes all of the model diagnostic procedures, forecasts are produced.

If the desired results are not successful, a more advanced modeling technique maybe required, or the model does not exist. In Stage 1, there are many models that may fit the training sample, but some fail to produce forecasts that resemble the prediction set. Prediction and forecasting are necessary to allow plans to be made about possible developments. Unfortunately, accurate prediction and forecasting are very difficult in some areas. For example, mathematical models of stock market behavior are often unreliable in predicting future behavior. Consequently, stock investors may anticipate or predict a stock market boom, or fail to anticipate or predict a stock market crash. The validation step at Stage 1 not only forestalls bad full data predictions developed at Stage 2, but should also sustain our confidence in the predictability of the final model.

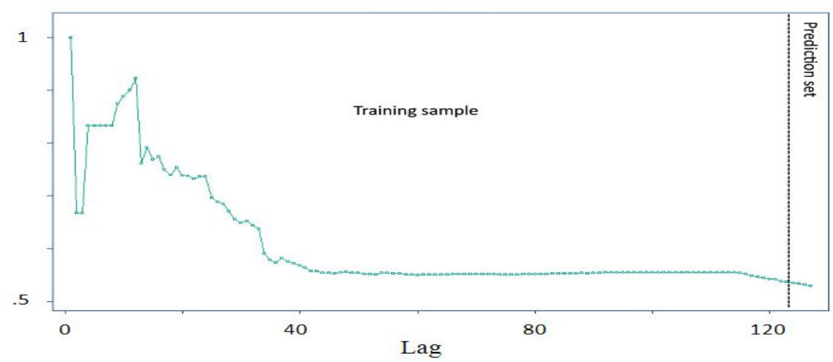

Figure VIII. ERRR plots (R_{23}) with training sample and prediction set.

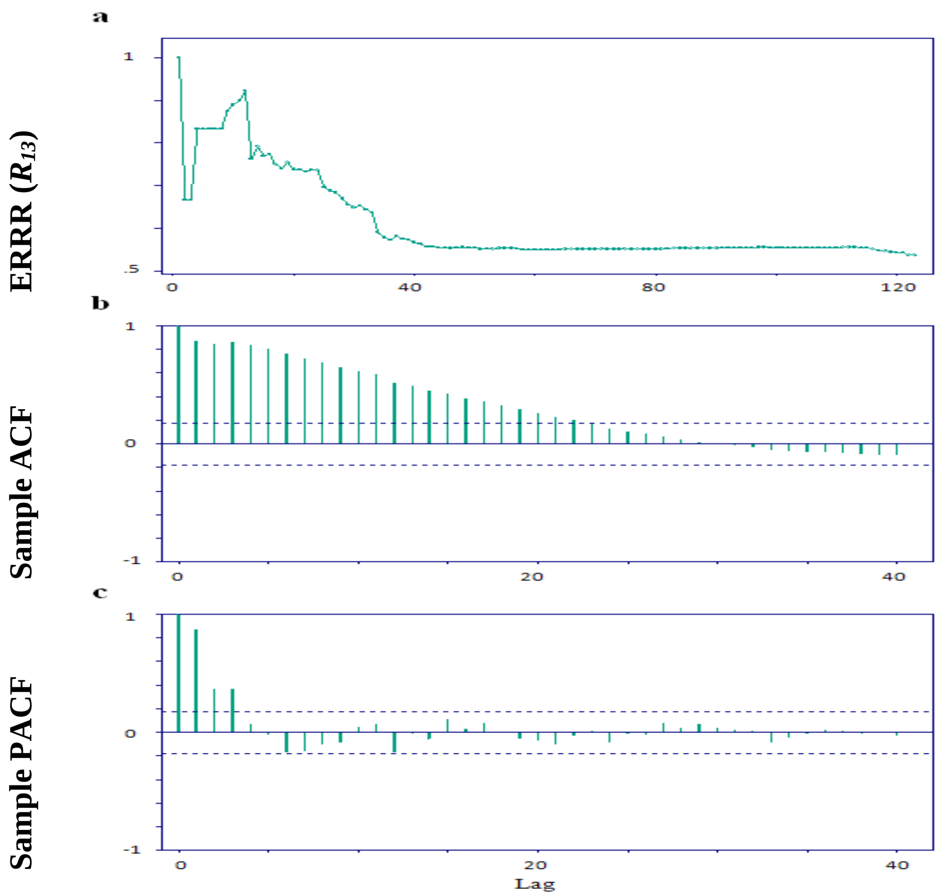

Figure IX. a. Time plot; b. ACF; c. PACF of R_{23} in the training sample.

4.2 ARIMA Modeling

We have five ERR and six ERRR time series to fit, among which, only the R_{23} series modeling procedures and output will be discussed in detail. First, we split the whole R_{23} time series into a training sample and a prediction set. The training sample is the full data set excluding the last four ERRRs, which will form the prediction set, as shown in Figure VIII. These four ERRR values in the prediction set, representing data from the most recent four quarters, will be compared with the four predictions produced by each candidate model. The length of a prediction set is quite flexible, and can be considered as one of the parameters for a model/data sensitivity analysis. Figure IX shows the corresponding time plot, sample ACF, and sample PACF of the training sample, which indicates non-stationary behavior. Applying the differencing operator ∇ to the ERRRs of the training sample, we find that a lag-1 differencing (same as a difference of order 1) operation is sufficient to produce a series with no apparent trend (Figure X).

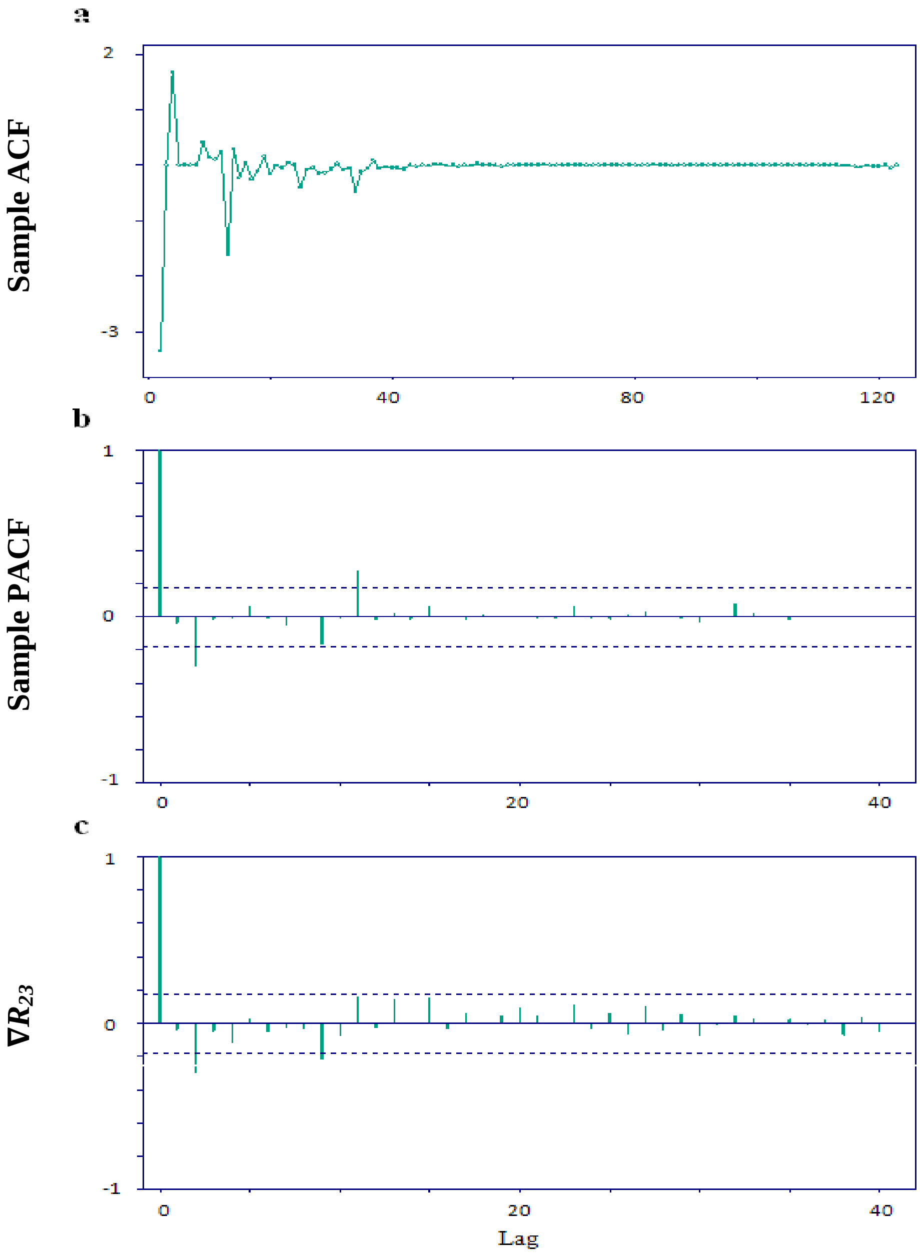

Figure X. a. Time plot; b. ACF; c. PACF of the lag-1 differenced R_{23} series in the training sample.

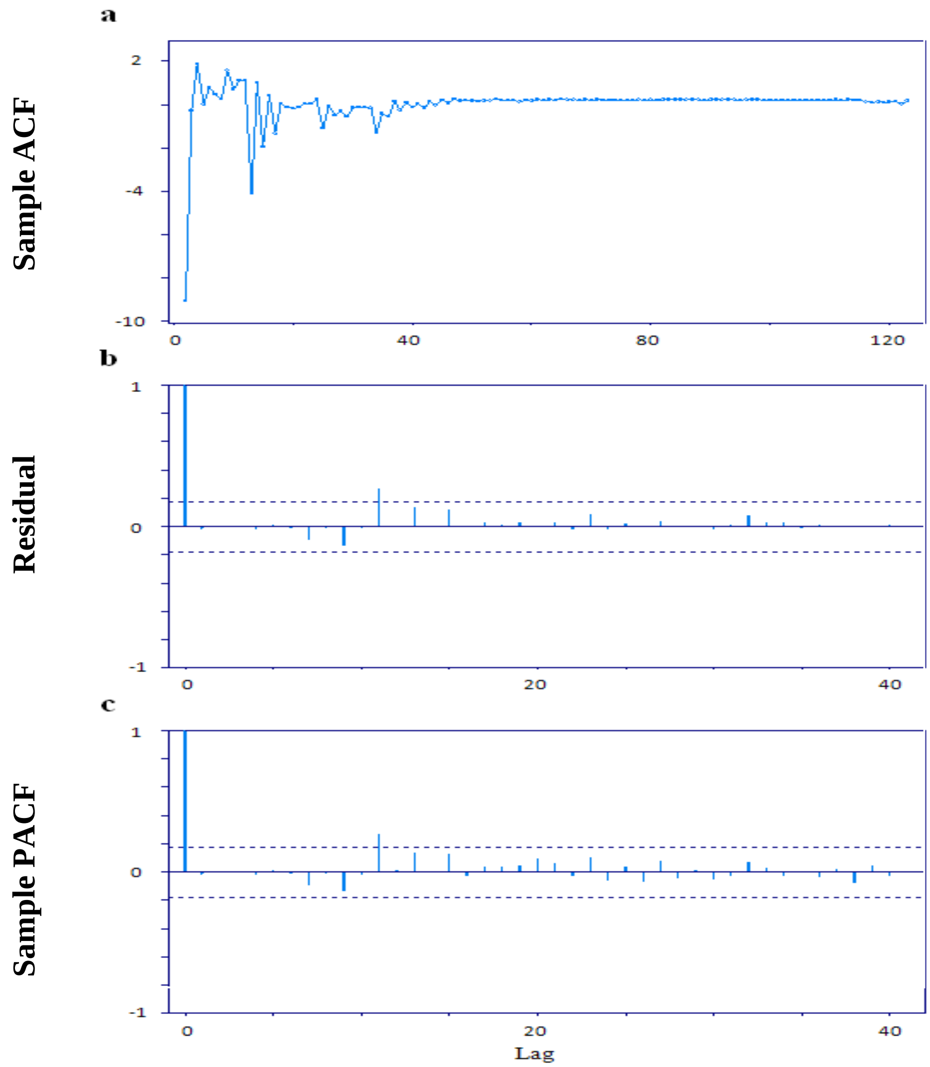

Figure XI. Diagnostics for ARIMA (4,2) fitted to the mean-corrected and lag-1 differenced R_{23} training sample. a. time plot; b. ACF; c. PACF of the residuals.

Next, using the ITSM package (Brockwell and Davis, 2002) and after traversing all possible models for a reasonable range of specified model parameters, we find that model ARMA(4, 2) appears adequate for the mean-corrected, lag-1 differenced ERRRs ∇R_{23}. It has the lowest AICC (= $-$ 463) and the residual plot (Figure XIa), and its ACF and PACF plots (Figure XIb and XIc) exhibit no significant spikes. The portmanteau goodness-of-fit test (Ljung and Box, 1978) is not significant (p-value = 0.45), indicating that the residuals are approximately white noise. The estimated (MLE) model for the mean-corrected, lag-1 differenced ERRRs (modeled as variable X) are:

$X_t = -.3176 X_{t-1} + .1690 X_{t-2} - .1078 X_{t-3} + .04347 X_{t-4} + Z_t + .2380 Z_{t-1} - .5709 Z_{t-2}$,(5) and the estimated variance for the white noise, Z, is 0.001159. A more parsimonious subset ARMA (4, 2) model may also be considered if desired, by deleting some or all of the non-significant model parameters. In addition, an attentive modeler might urge us to consider an ARFIMA model instead, because the sample ACF displayed in Figure IXb decays slowly, a classic characteristic of a long memory time series. In other words, there may be a problem of over-differencing of the original process when we use an integer difference parameter. Thus, this ARFIMA alternative is pursued.

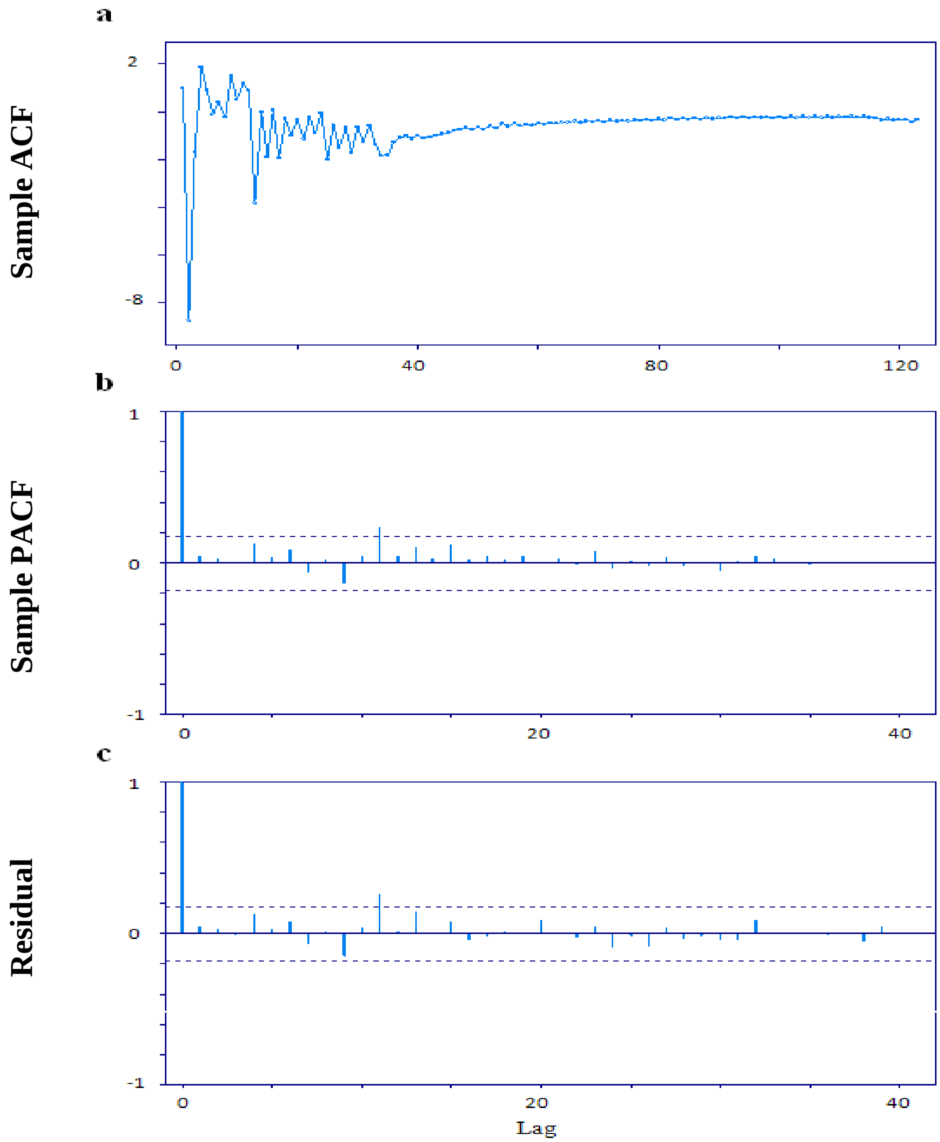

Figure XII. Diagnostics for ARFIMA (1, 0.498, 1) fitted to the mean-corrected R_{23} training sample. a. time plot; b. ACF; c. PACF of the residuals.

4.3 ARFIMA Modeling

The conventional ARIMA (p, d, q) process is often referred to as a short memory process. Autoregressive fractionally integrated moving average (ARFIMA) models are time series models that generalize ARIMA models by allowing non-integer values of the differencing parameter and are useful in modeling time series with long memory. The acronyms "ARFIMA" or "FARIMA" are often used, although it is also conventional to simply extend the "ARIMA (p, d, q)" notation by simply allowing the order of differencing, d, to take fractional values. Long memory is considered as an intermediate compromise between short memories ARMA type models and the fully integrated non-stationary processes (e.g., see Shumway and Stoffer, 2006). The model fitted in this paper will be written as, ARFIMA (p, d, q), if $0 < d < 0.5$. Now d becomes a parameter to be estimated along with the white noise variance.

We thus fit the fractionally differenced model to the mean-adjusted training sample of series R_{23}. The ITSM package is capable of a full analysis of ARFIMA modeling; hence, no additional programming would be necessary. A final value of d = 0.498 estimated by the ITSM leads to a set of diagnostic plots, shown in Figure XII.

We can compare roughly the performance of the fractional difference operator with the ARIMA model by examining the autocorrelation functions of the two residual series as shown in Figure XIb and XIIb. The ACFs of the two residual series are roughly comparable with the white noise model.

For the fitted ARFIMA (1, 0.498, 1) model, the AICC statistic is -469.82 and the Ljung-Box test is not significant (*p*-value = 0.47). The estimated (MLE) model for the mean-corrected training sample of R_{23} (modeled as variable X) is:

$$\nabla^{0.498}[X_t + 0.7287\, X_{t-1}] = Z_t + 1.0000\, Z_{t-1}, \quad (6)$$

And the estimated variance for the white noise, Z, is 0.001056. Note that the AICC statistic and the estimated white noise variance are slightly better than that of ARIMA (4, 1, 2). The predictability of both models will be assessed to conclude Stage 1 of the model building process.

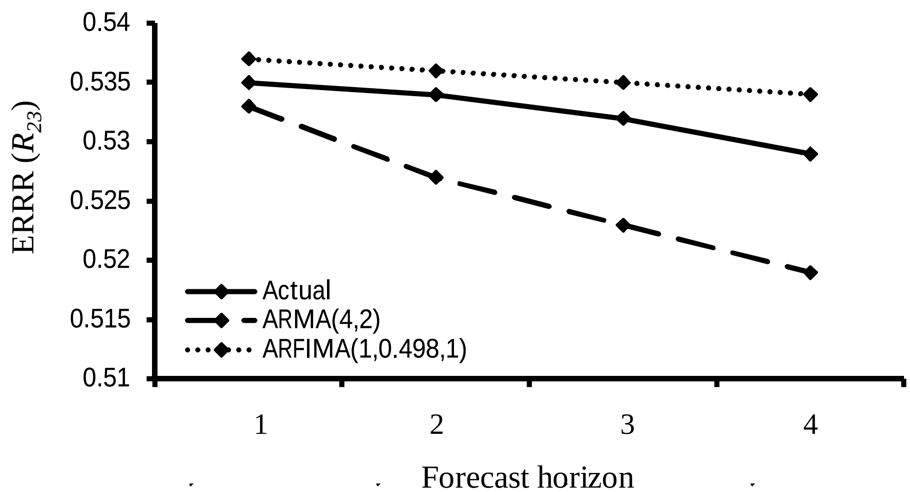

Figure XIII. Comparison of the forecasted ERRRs using ARMA (4, 2) and ARFIMA (1, 0.498, 1) models with the observed values in the R_{23} prediction set.

4.4 Training Sample Prediction

Four ERRRs, forecasted within this coherent methodology, are produced from each model and are displayed in Figure XIII. By a visual inspection, we conclude that both the selected models are not seriously biased and give an appropriate indication of the predictive ability of the model. The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the accuracy of a prediction model. The root-mean-square errors for a particular forecasting method are summarized across series by (e.g., see Armstrong and Collopy, 1992):

$$RMSE = \left(\sum_{i=1}^{4}(z_i - \hat{z}_i)/4\right)^{1/2}, \quad (7)$$

Where z_i is the actual value and $\hat{z}_i$ is its forecast. The result of the RMSEs ($RMSE_{ARIMA\,(4,\,1,\,2)} = 0.008 > RMSE_{ARFIMA(1,\,0.498,\,1)} = 0.003$) leads us to conclude objectively that ARFIMA (1, 0.498, 1) is the best model for Stage 1.

4.5 Full Data Prediction

For Stage 2, we fit the model, ARFIMA (1, 0.498, 1) declared in Stage 1, to the entire data set (training and prediction sets combined). In addition, as expected, the model, with coefficients slightly different from (6), passes all of the model diagnostic procedures. Finally, forecasts (95% prediction intervals) are produced for the coming year (2012:Q_1 - Q_4). The prediction intervals, in chronological order, are: (0.495, 0.561), (0.492, 0.560), (0.490, 0.558) and (0.476, 0.571).

As far as the significance of the difference in failure rates is concerned, the inclusion of 0.5 in all of the intervals furnished by the entire observed data implies that, statistically, it is predicted to be insignificant. And, simply put, even after 32 years, the race is still too close to call!

We have documented a complete round of analysis for R_{23}. The proposed modeling strategy and techniques are extended to the remaining ERRR time series. All of the fitted models along with a sequence of four prediction intervals for each curve are summarized in Table 2. Consistent with the notations previously developed, R_{ij} models the relative frequency of failures in G_i using G_j as the baseline banks group which is a larger banks group in terms of their adjusted total assets. A full-scale assessment of the modeling is then possible from Table 2. Analogous to what we've concluded for R_{23} (medium vs. large), the other pair that is predicted as insignificant is R_{34} (large vs. grand), even though it is consistently below the reference line (Figure 6) - another dead heat for the race!

Table 2: Full data prediction for time series R12, ...,and R34 for 2012:Q1 – Q4

Time Series	Transformed series	Model	2012 Forecast (95% Prediction Intervals.)			
			Q1	Q2	Q3	Q4
R12	∇∇R12	MA(2)	(0.582, 0.692)	(0.550, 0.723)	(0.502, 0.768)	(0.445, 0.823)
R13	∇R13	MA(2)	(0.623, 0.705)	(0.610, 0.716)	(0.600, 0.724)	(0.588, 0.732)
R14	∇∇R14	MA(4)	(0.578, 0.613)	(0.569, 0.617)	(0.560, 0.621)	(0.552, 0.624)
R23	R23	ARFIMA(1,0.498,1)	(0.495, 0.561)	(0.492, 0.560)	(0.490, 0.558)	(0.476, 0.571)
R24	∇∇R24	ARMA(4,4)	(0.423, 0.488)	(0.414, 0.492)	(0.407, 0.499)	(0.402, 0.499)
R34	∇R34	ARMA(1,5)	(0.316, 0.540)	(0.277, 0.580)	(0.246, 0.611)	(0.219, 0.637)

Armed with the records from the S&L crisis, the small banks group is still predicted as significantly higher than all of the other larger competitors. However, the difference becomes insignificant in 2012: Q4 when compared to the medium banks group. Excluding the smallest, the information accumulated from the aggregate behavior of the rest deserves some attention. Among the three groups, we already know that the differences for comparing medium vs. large and large vs. grand groups are insignificant. Surprisingly, in the end, the failures for the medium banks group are predicted to be significantly lower than the grand banks group as justified by all of the prediction intervals. This piece of information is educational and might provide additional justification towards what some experts have advocated for: a more traditional and smaller banking model.

5. Conclusions

In this paper, our intention is to characterize the relationship between the size of assets and bank failure. A historical record of banks passing through the S&L crisis, the Great Moderation and the Great Recession finally leads us to a series of predictions that conclude that in certain circumstances, certain things will happen. Policy makers nowadays argue that by breaking big financial institutions into units that are not too big to fail would make future bailouts unnecessary and restore market discipline.

With regards to the debate, we put our ERRR plots into action by broadcasting the bank failure collecting activities over three distinctive economic periods. Each episode is coupled with a model produced by consolidating two of the most powerful modeling tools for stochastic process and time series in the statistical literature to handle sets of transformation from counts of bank failures, perceived as events in a Poisson or Poisson-like process. The approaches are intuitive and the results, arising essentially from a graphical exploration to an implementation of a thoughtful modeling strategy, are compelling.

Firstly, our model predicts that the failures for the medium banks group are significantly lower than the grand banks group which is the largest group in our data set. The fact that the grand banks group is also consistently more vulnerable than the immediate group might have some practical significance. Secondly, recall that in the face of financial crisis, the U.S. government provided cash and guarantees to financial institutions whose failure, it feared, might bring down the whole system. This rescue was necessary, but it put taxpayers on the hook for potentially large losses. In addition, it also established a dangerous precedent: big financial institutions, we now know, will be bailed out in times of crisis. In addition, this, it is argued, will encourage even riskier behavior in the future, since executives at big banks will know that it is "heads they win, tails taxpayers lose" (Krugman, 2010). Adamati and Hellwig (2013) show how every possible eventuality that might stem from a banking crisis can be equally disastrous – be it simply the bank failing, a government aid, thereby letting it survive or allowing a bank to go on, even in the face of insolvency, thus rendering one firmly believe in the prevention of the nightmare. This work established the fact that, in the long run, there is a potential of a relatively higher vulnerability for the largest banks group when compared with other smaller groups. One may doubt that excessive risk-taking by large firms significantly contributed to the destiny. This is a claim worth examining - the most interesting and the most elusive.

References

Adamati, A. and Hellwig, M. (2013): The Bankers' New Clothes, Princeton University Press.

Altman, E. I. (1968): Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance,* 23: 589-609.

Altman, R. (2013): Why the economy could...POP!.*Time*, August 12.

Armstrong, J. S. and Collopy, F. (1992): Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting,* 8: 69-80.

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

Brockwell, P. J. and Davis, R. D.(2002):Introduction to Time Series and Forecasting. Second ed., Springer-Verlag, New York.

Crowley, F. D. and Loviscek, A. L. (1990): New directions in predicting bank failures: The case of small banks. *North American Review of Economics and Finance,* 1: 145-162.

Cui, F. (2011): ARIMA models for bank failures: Prediction and comparison. M.S. thesis, Department of Mathematical Sciences, University of Nevada, Las Vegas.

Dash, E. (1990): If it's too big to fail, is it too big to exist? New York Times, June 20.

Demsetz, R. S. and Strahen, P. E. (1977): Diversification, size and risk at bank holding companies. *Journal of Money, Credit and Banking,* 29: 300-313.

Emmons, W. R. Gilbert, R. A. and Yeager, T. J. (2004): Reducing the risk at community banks: Is it size or geographic diversification that matters?. *Journal of Financial Services Research,* 25: 259-281.

Estrella, A. Park, S. and Peristiani, S. (2000): Capital ratios as predictors of bank failure. *Economic policy review,* 6.

Foroohar, R. (2012): The economy's new rules: Go glocal. *Time*, August 20.

Furfine, C. H. (2003): Interbank exposures: Quantifying the risk of contagion. *Journal of Money, Credit and Banking,* 35: 111-128.

Ho, C.-H. (2008): Empirical recurrent rate time series for volcanism: application to Avachinsky volcano, Russia. *Journal of Volcanology and Geothermal Research,* 173: 15-25.

Ho, C.-H. (2016): Forecasting a point process with an ARIMA model. *Communications in Statistics –Theory and Methods*, 45:17, 5066-5076.

Ho, C.-H. and Bhaduri, M. (2015): On a novel approach to forecast sparse rare events: applications to Parkfield earthquake prediction. *Natural Hazards*, 78 (1): 669 – 679.

Koehn, M. and Santomero, A. M. (2012): Regulation of bank capital and portfolio risk. *Journal of Finance,* 35: 1235-1244.

Krause, A. and Giansante, S. (2012): Liquidity and solvency shocks in a network model of systemic risk: The impact of minimum capital and reserve requirement. 25th Australasian Finance and Banking conference.

Krugman, P. (2010): Financial reform 101. *The New York Times*, April 2.

Kuritzkes, A. and Schuermann, T. (2010): What we know, don't know and can't know about bank risks: A view from the trenches. In The Known, the Unknown, and the Unknowable in Financial Risk Management, edited by Diebold, Francis X., and Neil A. Doherty, and Richard J. Herring, pp. 103-144. Princeton University Pres.

Lane, W. R., Looney, S. W. and Wansley, J. W. (1986): An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance,* 10: 511-531.

Levin, C. and Coburn, T. (2011): Wall Street and the financial crisis: Anatomy of a financial collapse. Majority and Minority Staff Report, April 13.

Ljung, G. M. and Box, G. E. P. (1978): On a measure of lack of fit in time series models. *Biometrika,* 65: 297-303.

Mankiw, N. G. (2002): Macroeconomics. 5th ed., Worth, New York.

Martin, D. (1977): Early warning of bank failure: A logit regression approach. *Journal of Banking and Finance*, 1: 249-276.

Przyborowski, J. and Wilenski, H. (1940): Homogeneity of results in testing samples from Poisson series. *Biometrika,* 31: 313-323.

Shumway, T. (2001): Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business,* 74: 101-124.

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

Tan, S., Bhaduri, M. and Ho, C.-H. (2014): A staistical model for long-term forecasts of strong sand dust storms. *Journal of Geoscience and Environment Protection*, 2: 16-26.

Turner, A. (2010): Too much "too big to fail?. *Economist's View*, September 2.

Zhong, G. (2012): Analysis of Bank Failure and size of assets. M.S. thesis, Department of Mathematical Sciences, University of Nevada, Las Vegas.

Zmijewski, M. E. (1984): Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research,* 22: 59-82.