

Long-term Examination of Bank Crashes Using Panel Logistic Regression: Turkish Banks Failure Case

Birsen Eygi Erdogan

Correspondence: Marmara University, Arts and Sciences Faculty, Department of Statistics, Istanbul, Turkey. E-mail: birsene@marmara.edu.tr

Received: November 14, 2015 Accepted: February 26, 2016 Online Published: April 8, 2016

doi:10.5539/ijsp.v5n3p42

URL: <http://dx.doi.org/10.5539/ijsp.v5n3p42>

Abstract

Crises in the financial sector over the last two decades have shown the importance of early warning systems, especially for bank failures. This study aims to develop an early warning system for Turkish commercial bank failures using panel data from 2002 to 2012. The data was analyzed using pooled logistic regression versus random panel logistic regression. The dependent variable was the bank failure, defined as the return-on assets ratio. Factor analysis was used to construct independent variables of financial ratios. The meaningful factors were found as: Interest income and expenditures, Equity, Other income and expenditures, Balance sheet, Deposit, Due, Asset quality. When the focus is sensitivity, the best prediction performance was obtained using random-effect logistic regression.

Keywords: bank bankruptcy, panel data, random effect

1. Introduction

At the time of the Asian crisis, the Turkish bank system was damaged seriously. In 2007, a new economic crisis had started in United States that affected the whole world, and it was not known when it was going to end. Now, is not clear what banks will survive and continue and what banks will not.

There are several types of analysis that aim to discriminate between healthy and unhealthy banks. These studies were pioneered by Beaver (1966) for the univariate case and by Altman (1968) for the multivariate case. While Altman developed his model for manufacturing entities, Sinkey's (1975) model aimed to predict bank failure. For a fuller review of the literature, see Balcean and Ooghe (2006), Bellovary et al. (2007), Demyanyk and Hasan (2010), and Gepp and Kumar (2012).

There are some cross-sectional bank-failure studies that focus on the period before the Asian crisis, mainly between 1997 and 2001. These studies include the work done by Canbas et al. (2005), Celik and Karatepe (2007), Erdogan (2008), Boyacioglu et al. (2009), Erdogan (2012), and Inan and Erdogan (2013). There are very few panel data applications for bank bankruptcy related to the Turkish bank system. Büyükşalvarcı and Abdioglu (2013) undertook a panel data analysis to investigate the determinants of Turkish banks' capital-adequacy ratio. The results of their study show that loans, return on equity, and leverage have negative effects on the capital-adequacy ratio, while loan loss reserve and return on assets have positive effects.

Ilk et al. (2013) investigated bank failures in Turkey using longitudinal data structures. Using principal component analysis to avoid multicollinearity; they used three statistical models: logistic regression, generalized estimating equations, and marginalized transition models. All three models seemed to perform equally well according to the overall correct classification rate.

The objective of this study is to implement a panel data analysis using pooled logistic regression and random logistic regression on the financial ratios of Turkish commercial banks from 2000 to 2012. During the first period, that is the period before 2001, the Turkish Saving Deposit Insurance Fund (SDIF) took over all unsuccessful Turkish banks. Thus, all studies about Turkish commercial banks during this period defined bank failure as a bank that was taken over by the SDIF. During the second period, that is the period after 2001, there were not enough banks taken over by the SDIF to use the definition of failure as "being taken over by SDIF". Instead, this study uses a profit measure called Return on Assets (ROA) defined as the proportion of net profit after taxes over assets. This study is therefore distinguished from previous studies both in terms of time and failure definitions.

Section 1 reviews the literature. Section 2 introduces the theoretical aspects of the models used in the study. Section 3 discusses the data preparation operations, data sampling, and the selection of financial ratios. Section 4 discusses the

results. Finally, Section 5 provides conclusions.

2. Theoretical Background

2.1 Panel Data Structures

This study considers a panel data structure. The data was taken from the website of the Turkish Bank Association. With regression data, a cross-section of subjects for a specific time are collected. In contrast, with time series data, subjects are observed over time. Longitudinal/panel data include both regression and time series data. As with regression, a cross-section of subjects is collected and each subject is observed over time.

Panel data has benefits and drawbacks. Some advantages of using longitudinal data include studying dynamic relationships, studying heterogeneity, reducing the omitted-variable bias, and having more efficient estimators (Frees, 2004). However, in terms of drawbacks of panel data, a selection bias may occur if simple random sampling is not used to select observational units. And because the same subjects are followed over time, nonresponses typically increase over time.

2.2 Panel Data Model

In a panel data structure, observations are indexed with both i and t (Baltagi, 1995).

$$Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + \varepsilon_{it} \quad i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T \quad (1)$$

$$E(\varepsilon_{it}) = 0$$

$$\text{Var}(\varepsilon_{it}) = \sigma^2$$

$$E(\varepsilon_{it}, \varepsilon_{i't'}) = 0 \quad \text{if } i \neq i' \text{ or } t \neq t'$$

$$E(X_{jit}, \varepsilon_{it}) = 0 \quad \text{for all } j, i, t$$

Here i represents cross-section data such as households, individuals, firms, or countries, and t represent the time series. Some variables vary only across individuals or across time, and some vary with both individuals and time.

One key feature of the model is that we allow each individual i to have a distinct intercept β_{0i} . This intercept includes all aspects of unobserved heterogeneity that are fixed over the length of the panel.

A different panel data model can be given as below.

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3t} + \dots + \beta_k X_{kit} + v_i + \mu_{it} \quad (2)$$

$$i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T$$

In this model, there is a single intercept, β_0 , for all data points. However, the error term has been broken into two components, v_i and μ_{it} . When an individual i is drawn, one v_i is drawn that is fixed for that individual in all time periods. v_i includes all fixed, omitted variables.

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3t} + \dots + \beta_k X_{kit} + v_i + \mu_{it} \quad (3)$$

$$i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T$$

$$E(\mu_{it}) = 0 \quad \text{Var}(\mu_{it}) = \sigma_\mu^2$$

$$E(v_i) = 0 \quad \text{Var}(v_i) = \sigma_v^2$$

$$E(\mu_{it}, \mu_{i't'}) = 0 \quad \text{if } i \neq i' \text{ or } t \neq t'$$

$$E(v_i, v_{i'}) = 0 \quad \text{if } i \neq i'$$

$$E(\mu_{it}, v_{i'}) = 0 \quad \text{for all } i, i', t$$

$$E(x_{jit}, v_{it}) = 0 \quad \text{for all } j, i, t$$

$$\text{Either } E(X_{jit}, v_i) = 0 \quad \text{for all } j, i, t \text{ or } E(X_{jit}, v_i) \neq 0 \quad \text{for at least some } j, i, t$$

2.3 Panel Logistic Regression Analysis

In some economic studies, the dependent variable is discrete, indicating, for example, that an individual defaulted on a loan or was denied credit, or that a bank is successful or not. This dependent variable is usually represented by a binary-choice variable $y_{it} = 1$ if the event happens and 0 if it does not for individual i at time t (Baltagi 1995). The probability that the event happens for individual i at time t , p_{it} , is usually modeled as a function of some explanatory variables.

$$p_{it} = \Pr[y_{it} = 1] = E(y_{it}/x_{it}) = F(x'_{it}\beta) \quad (4)$$

Generally, a logistic cumulative distribution function is used that constrains $F(x'_{it}\beta)$ to between zero and one.

$$p_{it} = F(x'_{it}\beta) = \frac{e^{x'_{it}\beta}}{1+e^{x'_{it}\beta}} \quad (5)$$

For the panel data structure, when the dependent variable is binary, the first choice for constructing a discriminating model is the pooled logistic regression. With this method, the panel structure of the data is ignored.

A Fixed Effects (FE) estimator can be used with either the distinct intercepts or the error components if the heterogeneity term is correlated to explanatory variables. Here the basic idea is to estimate a separate intercept for each individual. The basic idea for the Fixed Effects estimator is to cancel out v_i by differencing observations for the same individual. It is possible to use all the observations for each individual if the individual's specific mean is subtracted from each observation.

Ordinary least square is consistent in the uncorrelated version of the error components. In the presence of serial correlation, it is necessary to correct for serial correlation between observations of the same individual. If the heterogeneity term is independent of the explanatory variables, the Random Effects (RE) method is used to overcome the serial correlation of panel data.

Logistic regression parameter estimators are usually found using the Maximum Likelihood Estimator (MLE) method. The maximum likelihood equation is constructed using the joint probability distribution and expresses the values of β in terms of known values for y .

Chamberlain (1980) suggests maximizing the conditional likelihood function to obtain the conditional logit estimates for β .

$$L_c = \prod_{i=1}^N \Pr(y_{i1}, \dots, y_{iT} / \sum_{t=1}^T y_{it}) \quad (6)$$

To test for fixed individual effects, one can perform a Hausmann-type test based on the difference between Chamberlain's conditional MLE and the usual logit MLE ignoring the individual effects.

The Hausmann test (Amini, 2012) states in null hypothesis that the difference in coefficients is not systematic, that is, random effect is appropriate. If the null hypothesis is rejected, it is accepted that the fixed model is appropriate and there is no need to consider random effect.

3. Implementation and Results

3.1 Datasets

The sample of banks and the data used to train and test the models were taken from the website of The Banks Association of Turkey (BAT) (<http://www.tbb.org.tr/english/>). This site serves bankers, researchers, and investors and offers banks' financial tables as well as an overall evaluation of the performance of the Turkish economy and the banking sector.

This study includes data from 22 banks with commercial banking activities from 2001 to 2011 (Appendix A). After the Asian crisis, new regulations were made in the Turkish bank system, and Turkish banks were mostly saved the fate of being taken over by SDIF. Although some of the banks are rumored to be unhealthy, SDIF did not take over any of them. It is therefore necessary to create a new definition for bank failure. Return on Assets (ROA) may be used as a proxy for profitability, and profitability can be used to describe failure.

The ROA provides information about how much profit is generated on average by each unit of assets. ROA is therefore an indicator of how efficiently a bank is being run.

$$ROA = \frac{\text{Net Profit After Taxes}}{\text{Assets}}$$

As a result, this study uses ROA to determine the health status of a bank; that is, ROA is used to determine dependent variable values. For this reason, this study examines the ROA ratios for the period from 1988 to 2000. The average ROA ratio for banks that were taken over by SDIF during this period is less than one for every year. Accordingly, if the ROA is less than one, the bank is classified as unsuccessful and coded as one. Otherwise, if the ROA is greater than one, the bank is classified as successful and coded as zero. The ROA is computed for each bank for each year.

Panel data from 2001 through 2011 were used for training, and the 2012 data set was used for testing. In this manner, the models, which are constructed from the previous year's panel data, are tested on the new data set, which is derived from the future and used for an ex-ante prediction.

3.2 Implementation

The variable set of this study includes the financial ratios of 22 Turkish commercial banks. Some of these ratios measure liquidity and some measure profitability, efficiency, solvency, and leverage. Appendix B contains a list of the financial ratios.

Because there are many ratios in financial tables that indicate the success levels of banks, and because some of the ratios are highly correlated with each other, this study uses a factor analysis to construct independent factors that affect the success status of the banks. Stata 12 was used as an analytical tool to perform the research in detail. Factor analysis was performed using IBM SPSS Statistics 21. Because there were some financial ratios with very high correlations, those with higher than 0.90 correlations were excluded before factor analysis was performed. Eighteen financial ratios were found appropriate for factor analysis.

Because the data is in panel form, to check whether classical factor analysis is appropriate for pooled data, the Levin-Lin-Chu unit root test was used as a stationarity test for each ratio before factor analysis. All ratios are stationary, that is, they do not contain unit roots.

For the factor analysis standard SPSS Dimension Reduction routine was used. According to the eigenvalue criteria, there were seven meaningful factors. Factors with an eigenvalue greater than one were treated as meaningful. The meaningful factors are: Interest income and expenditures, Equity, Other income and expenditures, Balance sheet, Deposit, Due, Asset quality. Then, the scores of these new factors were used as independent variables for panel logistic regression analysis.

As discussed in Section 2, with pooled panel logistic regression, the panel structure of data is ignored. When the panel structure is considered, there is a choice between fixed-effect logistic regression and random-effect logistic regression. To make this decision, a Hausmann test was performed. In the null hypothesis, a Hausmann test states that if the difference in coefficients is not systematic, random-effect regression is appropriate. The p-value ($\text{Prob} > \chi^2$) was found to be 0.6300, meaning do not reject the null hypothesis. This result means that the random-effect model is appropriate.

When the logistic regression routines were run, the small p-value, < 0.00001 , led to the conclusion that at least one of the regression coefficients in the model was not equal to zero, meaning that the models were meaningful.

To assess the models, each model was trained on both the training and test data sets. The models were evaluated using a confusion matrix.

A confusion matrix of a standard classification problem includes four counts: True Positive (TP) (TP=The model classifies a failed bank as failed), False Positive (FP) (FP=The model classifies a non-failed bank as failed), False Negative (FN) (FN=The model classifies a failed bank as non-failed), and True Negative (TN) (TN=The model classifies a non-failed bank as non-failed).

The confusion matrix (Kumar, 2005) thus represents model performance. Accuracy is defined as the number of correct predictions divided by the number of observations. Error, which is usually the most important metric for model performance, is defined as the number of mistakes divided by the number of observations. Sensitivity, another intuitive metric for model performance, is defined as TP divided by the sum of all positives. Because sensitivity determines how well a model finds positive instances, in some situations, like failed banks, sensitivity is more important than accuracy. Specificity indicates how well a model can find negative instances; it is defined as TN divided by the sum of all negatives. When evaluating the results of a classification procedure, the measure on this study focuses depends on the issue at hand.

4. Results

Because of the impact on the economy, it is more important to diagnose a bank that may be at risk for bankruptcy than to diagnose a healthy bank. Therefore, the focus of this study is sensitivity.

The overall accuracy, sensitivity and specificity metrics were obtained for all trained models and are shown in Table 1.

Table 1. The results for the train/test set.

		Training Data		Test Data	
		Pooled Panel LR	Random Effect LR	Pooled Panel LR	Random Effect LR
Sensitivity	Pr(1_1)	81%	96%	60%	100%
Specitivity	Pr(0_0)	85%	71%	100%	76%
Accuracy		91%	77%	91%	82%

With regard to sensitivity, it is interesting that when the rate for pooled panel logistic regression decreases from 81% to 60%, there is an increase in the random-effect logistic regression from 96% to 100%. Because this study attempted to make an ex-ante prediction using a model based on data from the period prior to the time of bank failure, it is reasonable to find a lower sensitivity rate for the test set. The result may be due to the chance.

5. Conclusions

In this study, pooled panel logistic regression and random-effect logistic regression were implemented to analyze bank financial ratios. Data sets belonging to Turkish commercial banks were used.

According to the results, both pooled panel logistic regression and random-effect logistic regression are capable of extracting useful information from financial data. But when the main goal is to find positive instances, that is, failed banks, it is more appropriate to focus on sensitivity than accuracy.

When the focus is sensitivity, the best prediction performance was obtained under consideration of the panel structure of the data, using random-effect logistic regression. Therefore, using a longitudinal data structure and implementing a random-effect logistic regression can give useful information about the failure status of banks and can be used as part of an early-warning system.

Acknowledgements

The research for this article was supported as a D-Type project (FEN-D-070814-0274) by Marmara University Scientific Research Projects Committee. Under the instruction of Birsen Eygi Erdogan, an early version of this study was presented on May 18, 2014 at the International Statistics Colloquia at Yildiz Technical University by Marmara University students Ahmet Anil Cicek, Batuhan Mengü, Dogan Dogukan Unal, and Ozan Hakki Okcu.

References

- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, c.XXIII, 589–609.
- Amini, S., Delgado, M. S., Henderson, D. J., & Parmeter, C. F. (2012). Fixed vs Random: The Hausman Test Four Decades Later, *Essays in Honor of Jerry Hausman.*, 479-513
- Balcaen, S., & Ooghe, H. (2006). 35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems. *The British Accounting Review*, 38(1), 63–93 <http://dx.doi.org/10.1016/j.bar.2005.09.001>.
- Baltagi, B. H. (1995). *Econometric Analysis of Panel Data*, New York, NY: John Wiley & Sons.
- Beaver, W. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 5, 71–111.
- Bellovary, J., Giacomin, D., & Akers, M. (2007). A Review of Bankruptcy Prediction Studies: 1930-Present. *Journal of Financial Education*, 33, 1–42.
- Boyacioglu, M. A., Kara, Y., & Baykan, O. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Application*, 36, 3355–3366.
- Büyüksalvarcı, A., & Abdioglu, H. (2013). Determinants of capital adequacy ratio in Turkish Banks: A panel data analysis. *African Journal of Business Management*, 5(27), 9 November, 2011 Available online at <http://www.academicjournals.org/AJBM> ©2011 Academic Journals, pp. 11199–11209. <http://dx.doi.org/10.5897/AJBM11.1957> ISSN 1993-8233.
- Canbas, S., Cabuk, A., & Kilic, S. B. (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structures. *European Journal of Operation Research*, 166, 528–546. <http://dx.doi.org/10.1016/j.ejor.2004.03.023>.
- Celik, A. E., & Karatepe, Y. (2007). Evaluating and forecasting banking crises through neural network models: An application for Turkish banking sector. *Expert Systems with Application*, 33, 809–815.
- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47, 225–238.
- Demyanyk, Y., & Hasan, I. (2010). Financial crises and bank failures: A review of prediction methods. *Omega - International Journal of Management Science*, 38(5), 315–324. <http://dx.doi.org/10.1016/j.omega.2009.09.007>.
- Eygi Erdogan, B. (2008). Bankruptcy prediction of Turkish commercial banks using financial ratios. *Applied Mathematical Sciences*, 60, 2973–2982.
- Eygi Erdogan, B. (2013). Prediction of bankruptcy using support vector machines: an application to bank bankruptcy.

Journal of Statistical Computation and Simulation, 83(8), 1543–1555.
<http://dx.doi.org/10.1080/00949655.2012.666550>.

- Frees, E. W. (2004). *Longitudinal and Panel Data Analysis and Applications in the Social Sciences*, Cambridge University Press.
- Gepp, A., & Kumar, K. (2012). Business failure prediction using statistical techniques: A review. In K. Kumar & A. Chaturvedi (Eds.), *Some Recent Developments in Statistical Theory and Applications*, Boca Raton, Florida, USA: Brown Walker Press, pp. 1–25.
- Ilk, O., Pekkurnaz, D., & Cinko, M. (2013). Modeling company failure: a longitudinal study of Turkish banks. *Optimization: A Journal of Mathematical Programming and Operations Research*, 63(12), 1837–1849. <http://dx.doi.org/10.1080/02331934.2013.855762>
- Inan, D., & Erdogan, B. (2013). Liu-Type Logistic Estimator. *Communications in Statistics - Simulation and Computation*, 42(7), 1578–1586. <http://dx.doi.org/10.1080/03610918.2012.667480>.
- Kumar, P. R. & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. *European Journal of Operational Research - EJOR*, 180(1), 1–28. <http://dx.doi.org/10.1016/j.ejor.2006.08.043>.
- Sinkey, Jr., J. (1975). A multivariate statistical analysis of the characteristics of problem banks. *Journal of Finance*, 30(1), pp. 21–36.
- Tan, P. N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*, Pearson Verbeek, M. (2012), *A Guide to Modern Econometrics 4th Edition*, Wiley.

Appendix A

The banks used in the analysis

Name of the Banks

Akbank t.a.ş
 Alternatifbank a.s.
 Anadolubank a.s.
 Arap Turk bank a.s.
 Burgan bank a.s.
 Citibank a.s.
 Denizbank a.s.
 Fibabanka a.s.
 Finans bank a.s.
 Hsbc bank a.s.
 Ing bank a.s.
 Sekerbank t.a.s.
 Tekstil bankası a.s.
 Turkish bank a.s.
 Turkland bank a.s.
 Turkiye cumhuriyeti ziraat bankasi a.s.
 Turkiye ekonomi bankasi a.s.
 Turkiye garanti bankasi a.s.
 Turkiye halk bankasi a.s.
 Turkiye is bankasi a.s.
 Turkiye vakiflar bankasi t.a.o
 Yapi kredi bankasi a.s.

Appendix B

The list of the financial ratios.

Capital Ratios (%)	
X1	Shareholders' Equity / (((Capital to be Employed to credit + market + operational risk)*12.5)*100
X2	Shareholders' Equity / Total Assets
X3	(Shareholders' Equity - Permanent Assets) / Total Assets
X4	Net On Balance Sheet Position / Total Shareholders' Equity
X5	Net On and Off Balance Sheet Position / Total Shareholders' Equity
Balance-Sheet Ratios (%)	
X6	TC Assets / Total Assets
X7	TC Liabilities / Total Liabilities
X8	FC Assets / FC Liabilities
X9	TC Deposits / Total Deposits
X10	TC Loans / Total Loans and Receivables
X11	Total Deposits / Total Assets
X12	Funds Borrowed / Total Assets
Assets Quality (%)	
X13	Financial Assets (net) / Total Assets
X14	Total Loans and Receivables / Total Assets
X15	Total Loans and Receivables / Total Deposits
X16	Loans Under Follow-up (gross) / Total Loans and Receivables
X17	Loans Under Follow-up (gross) / Total Loans and Receivables
X18	Specific Provisions / Loans Under Follow-up
X19	Permanent Assets / Total Assets
X20	Consumer Loans / Total Loans and Receivables
Liquidity (%)	
X21	Liquid Assets / Total Assets
X22	Liquid Assets / Short-term Liabilities
X23	TC Liquid Assets / Total Assets
Profitability (%)	
X24	Net Profit/Losses / Total Assets
X25	Net Profit/Losses / Total Shareholders' Equity
X26	Profit/Losses Before Taxes after Continuing Operations / Total Assets
Income-Expenditure Structure (%)	
X27	Net Interest Income After Specific Provisions / Total Assets
X28	Net Interest Income After Specific Provisions / Total Operating Income (Expenses)
X29	Non-interest Income (net) / Total Assets
X30	Other Operating Expenses / Total Assets
X31	Personnel Expenses / Other Operating Expenses
X32	Non-interest Income (net) / Other Operating Expenses

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).