The Vietnam banking industry and its retail banking: Artificial neural networks and statistical analysis

A thesis
submitted in partial fulfilment
of the requirements for the Degree of
Doctor of Philosophy in Finance

at
Lincoln University
by
Man Minh Ha

Lincoln University
2020

Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Finance .

Abstract

The Vietnam Banking Industry and its Retail Banking: Artificial Neural Network and Statistical Analysis

by

Man Minh Ha

The 2008 Global Financial Crisis (GFC) had adverse effects on Vietnamese banks. After the crisis, the banks faced many unexpected challenges, including recession, stagnation in credit growth rates and an increased number of non-performance loans (NPL). The Vietnamese government released the "Restructuring Financial Institutions 2011-2015" on 1st March 2012. This report detailed steps to enhance the banking industry's performance and ensure that it met international banking standards. The report recommended merger and acquisition (M&A) deals as one way to achieve its objectives to improve bank competitiveness and efficiency. The restructuring, via M&A deals, is expected to have an even greater impact on market concentration and competition in the Vietnamese banking industry. As Hoang, Phan, and Bandaralage (2016) note, monopolies may occur because of a decrease in the number of banks in the market.

Using unbalanced yearly data from 2008 to 2017, this is the first study to use the self-organisation map (SOM) technique to track financial trajectory patterns and categorise Vietnamese banks into super-class groups. The study identifies two super-class groups. Group 1 has only joint-stock banks; group 2 contains commercial state and joint-stock banks. Using the non-structural indicator the Lerner index to capture market power and the Data Enveloped Analysis technique to measure efficiency, this study shows there are significant differences in Lerner scores (which represent bank market power) of the two bank groups. Though there are notable significant (at the 1% level) differences between the two bank groups' market power, their cost-efficiency scores (which represent bank efficiency) are the same. The differences in the Lerner scores provide evidence of a group of strong banks that is isolated from other banks. This phenomenon is the result of market contraction through M&A activity. This study argues that this strong bank group has the potential to be monopolists and may damage the competitive banking environment.

Using the system generalised method of moments to overcome challenges associated with endogeneity and generating unbiased estimating parameters, this study also shows that bank size and capitalisation ratio have a positive relationship with market power. These findings suggest that large, well-capitalised banks enjoy greater market power because of their dominant position. Their greater capital reserves mean that they can take greater risks and lend to more borrowers. However, the negative relationship between bank size and cost efficiency indicates that big banks are less effective in distributing their costs than small banks, which leads to lower efficiency. This is because successful M&A activity may improve bank size quickly, but negatively impact on bank efficiency because of increased NPL values and operational costs. A higher NPL rate weakens both the efficiency and market power of Vietnamese banks. Bank executives in both groups 1 and 2 must focus on decreasing their NPL rate to ensure their efficiency and maintain their market power. The study also argues that Vietnamese banks gain efficiency during periods of rapid GDP growth or high inflation rate; in sum, banks raise their lending interest rates in response to a high inflation rate. The results also indicate that market power has trends to maintain over time. This is because the one-year lagged market power is positive and significantly impacts on bank market power. This finding indicates that the Vietnam banking industry is characterised by non-transparent information, networking lending relationships and limited regulations around monopolistic practices, which leads to bank monopolies and reduces competition and innovation.

Because of increased numbers of internet users and emerging e-commerce, retail banking is a profitable market for Vietnam domestic banks to ensure long-term development. However, less than 50% of bank executives are satisfied with their current credit internal rating systems; most believe that their credit scoring models (CSMs) are ‘poorly’ designed. This study proposes a more accurate combination method that enables banks to build a more efficient credit scoring model and hence make better credit-granting decisions. The component planes generated by the SOM technique provide a visual illustration of the classification process and thus can assist researchers in understanding the motivation process. The merger between Logit and Multilayer Perceptron outperforms the logit regression alone. This is because the overall classification accuracy is improved to 1.2% using the hybrid technique. Type II errors also decrease significantly, from 34.7% to 15.3% (a reduction of over 50 per cent), using the hybrid technique. This leads to a massive reduction in the misclassification cost (to 56%). Other evaluating discriminatory power criteria, the AUC, Gini and KS coefficients, provide robust evidence that the hybrid technique has greater classification power than logit regression alone.

Keywords: self-organisation map, financial trajectory patterns, the monopolists, multilayer perceptron, generalised method of moment, endogeneity

Acknowledgements

My sincere thanks and praise to the Avalokiteśvara Bodhisattva for giving me the opportunity to pursue my difficult PhD studies and for guiding me to discover the truth.

I would like to thank Prof. Christopher Gan for his continuous support, insightful comments and invaluable feedback throughout my PhD study. I also recognised the comments offered by my associate supervisors Dr Cuong Nguyen and Assoc. Prof. Patricia Anthony. Thank you for your guidance that enabled me to successfully complete my thesis. I also would like to thank Prof. Sandhya Samarasinghe for teaching me the COM607 Neural Networks Applications. My sincere appreciation to Rebecca Kambuta, I am thankful for your kindness and help in editing my writing.

Special thanks and sincere appreciation to my beloved mother Mrs. My, my mother-in-law Mrs. Khanh, and Aunty Mrs. Mai. Without their support and encouragement, I might not have been able to complete my PhD study. My sincere gratitude also goes to my brothers, sisters and friends for their support and motivation.

Finally, I wish to express my sincere gratitude to my wife Tram and my children Chau, Ngan and Quan. Thanks to you all for your patience and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AC	Accuracy rate
ANN	Artificial neural networks
AUC	Area under ROC curve
CAP	Capitalisation ratio
CE	Cost efficiency
CSM	Credit scoring model
DEA	Data enveloped analysis
GDP	Gross domestic product
GFC	Global Financial Crisis
Gini	Gini coefficient
GMM	Generalised method of moments
INF	Inflation rate
IV	Information value
K-S	Kolmogorov–Smirnov coefficient
M&A	Merger and acquisition
MLP	Multi-layer perceptron
NPL	Non-performance loans
PTE	Pure technical efficiency
ROC	Receiver operating character curve
SFA	Stochastic frontier analysis
SOM	Self-organisation maps
WOE	Weight of evident
WSS	Within group sum of square

Chapter 1

Introduction

1.1 Vietnam: Its Economy and Banking System

Vietnam's banking system plays a pivotal role in its economy and is the main pillar of the nation's financial system (Nguyen, Roca & Sharma, 2014). It has fundamentally changed over the 32 years since the introduction of the country's reformation programme known as Doi moi (1986 – 2018). From four state-owned commercial banks (SOCBs) and some small financial institutions in 1990, Vietnam's bank industry has undergone rapid expansion and diversification. In 2018, Vietnam's banks included four state-owned commercial bank (SOCBs) and 32 joint-stock banks (JSBs) (SBV, 2018; Tran, Ong & Weldon, 2015). In 2018, the total Vietnam bank assets reached 474.8 billion USD, more than the country's Gross Domestic Product (GDP) of 241.3 billion USD (SBV, 2018). The total bank assets were almost double the country's GDP (196.8%) even though they were 18% of GDP in 1996 (Nguyen et al., 2014). Vietnam's banks annually contribute 16% to 18% to the national GDP (Stewart, Matousek & Nguyen, 2016). Currently, the Vietnam banking system is a two-tiered system in which privately owned banks are permitted to operate alongside state-owned banks. Foreign banks are also allowed to set up wholly-owned operations in Vietnam (Ho & Baxter, 2011). There are nine foreign banks: HSBC, Standard Chartered Bank, ANZ, Shinhan Bank, Public Bank Bhd, Hong Leong Bank, Woori, UOB and CIMB. In 2016, the total registered capital of the foreign banks was almost 75% of the country's GDP (Ha, 2016). There are 50 foreign bank branches and 50 representative offices now operating in Vietnam (StoxPlus, 2016).

Since the 1990s, the Vietnamese economy has steadily progressed towards a market-based economy. However, there are still weaknesses in terms of macroeconomic stability and sustainable economic development (Nguyen, Tran & Tran, 2014). The Vietnamese economy is vulnerable to changes in the global economy. Fluctuations in oil prices and export demands reveal weaknesses in the country's economy. As oil prices increased in 2008, local transport costs also increased. This has led to an increased inflation rate that, in turn, affected the financial system. For instance, inflation increased by 8.3% in 2007 to 23.15% in 2008 (Worldbank, 2018a). To attract more customer deposits, the banks raised deposit rates dramatically from 7.5% in 2007 to 12.7% in 2008 (Worldbank, 2018b). This negatively affected investors and borrowers because the cost of finance (lending interest rate) also increased. The Vietnam Civil Code Law, clauses 1 and 2, article 476 declared that commercial banks can issue the lending interest rates by themselves. However, the lending interest rates must not exceed 150% of the basic interest rate announced by the State Bank (Nguyen, 2005). As a result, investors who invest in new projects face lower profits because of higher borrowing costs. Moreover, banks face declines in credit growth because of the cancellation or suspension of major projects.

The European Union, the United States, and Japan are the three primary export markets for Vietnamese companies. These three countries account for 60% of the total export value, approximately 37.6 million USD, over one-third of Vietnam's GDP in 2008 (approximately 99.13 million USD) (Le, 2009; WorldTrade, 2008). Fluctuating export demands negatively influence Vietnam's economy. The 2008 Global Financial Crisis had a negative effect on Vietnam; the country's export demand sharply declined. As a result, many Vietnamese export companies were unable to meet their loan repayments and announced bankruptcy. These situations contributed to an increase in bad debts and negatively affected the banking system. For example, 20 companies reported to the Ho Chi Minh City Export Processing and Industrial Zones Authority (HEPZA) that they were facing major financial problems and were unable to pay their workers and/or employ new staff. This led to a massive number of redundancies; 66,700 workers were laid off in 2008 and the national unemployment rate increased from 2.3% in 2007 to 4.65% in 2008 (Le, 2009). As these examples illustrate, fluctuating export demands impact indirectly on local Vietnam banks. In short, Vietnam banks need to recognise risks that may negatively affect their business. The next section discusses how the 2008 Global Financial Crisis affected the Vietnam banking industry, with a particular focus on retail banking.

1.2 The Impact of the 2008 Global Financial Crisis on the Vietnam Banking Industry

The 2008 Global Financial Crisis (GFC) harmed Vietnam's banks' performance. After the crisis, local banks faced many unexpected challenges. These factors included recession, credit growth rate stagnation, and high non-performancing loans (NPLs) (bad debts) (KPMG, 2013). The recession was the first major problem that Vietnam banks faced after the 2008 GFC. The GDP growth rate reduced sharply from 7.1% to 5.4% from 2007 to 2009 and remained under 6.8% till 2017 (Worldbank, 2018a). As an emerging market, the correlation between Vietnam's GDP growth and credit growth is noticeable (KPMG, 2013). Therefore, it is not surprising that the credit growth rate of Vietnam banks declined from 63% to 5% between 2008 and 2012 and remained at approximately 16% until 2017 (KPMG, 2013; Stewart et al., 2016; Tran, Tran & Ly, 2017).

Lending rates also affect the economy. During the 2008 GFC, Vietnam banks' lending interest rate increased from 11.2% to 15.7% in a single year (2007 to 2008) and reached 16.9% in 2011 (Worldbank, 2018b). Beck, Jakubik and Piloiu (2013) point out that a low GDP growth rate plus a higher lending interest rate, might expose bank asset quality to significant risk. This may explain why, after the 2008 GFC, the NPLs of Vietnam banks increased dramatically. In April 2013, NPLs accounted for 4.67% of total outstanding loans, equivalent to 6.5 billion USD (VND 137.13 trillion), despite low levels in 2005

(KPMG, 2013; Tran et al., 2015). By the end of 2016, the proportion of official NPLs on Vietnam banks' balance sheets had declined to 2.58% of the total lending, i.e., VND 131.7 trillion or 6 billion USD (Tran et al., 2017). However, the total amount owing was much higher - equalling 5.21% of total lending (approximately 266,149 billion VND) (Tran et al., 2017).

Restructuring Financial Institutions

The Vietnamese government released the "Restructuring Financial Institutions 2011-2015" programme on 1 March 2012 (Decision no.254/QD-TTg). The restructuring programme was designed to assist the banking system fundamentally and comprehensively bring its performance into line with international banking standards as an efficient system to overcome crisis or recession (Le, 2014; Nguyen et al., 2014; Nguyen, Nghiem, Roca, & Sharma, 2016a). The main features of the restructuring financial institutions programme are:

i. It allows foreign ownership of local banks with a maximum 20% share (Nguyen et al., 2016a). This strategy has provided local banks with opportunities to develop new banking products and services. It also enables local banks to improve their management techniques and adopt technology to bring their standards in line with international bank practice.

ii. It encourages all local banks to list shares on the Vietnamese stock exchange. The number of listed local banks increased from two to ten between 2006 and 2018.

iii. In bank capitalisation, the central bank determined that all commercial banks must hold at least 3.000 billion VND from 70 billion minimum bank capital. The programme also increased the minimum capital adequacy ratio (a ratio of total capital to risk-weighted assets) from 8% in 2005 to 9% in 2010.

iv. It encourages merger and acquisition (M&A) activities as a key strategy to improve the competitiveness and efficiency of Vietnam banks (Hoang, Phan, & Bandaralage, 2016; Le, 2014).

M&As are seen as an effective strategy to reduce turmoil in the banking industry (Hoang et al., 2016). M&As have been a key feature of the Vietnam banking industry since the 1990s. From1997 to 2005, there were 10 M&As. Since the 2008 GFC, M&As have increased significantly, both in size and economics. Eight M&A deals took place between 2011 and 2015 (4 years) compared to ten M&A deals from 1997 to 2005 (8 years) (Hoang et al., 2016). These M&A deals decreased the number of Vietnam banks from 43 to 34 between 2008 and 2015 (Hoang et al., 2016).

Restructuring via M&As is expected to have a significant impact on market concentration and competition levels within the Vietnam banking industry in several ways (Le, 2014). First, a reduced number of banks will increase market concentration (Fernández de Guevara, Maudos, & Pérez, 2005). Therefore, it is necessary to examine whether banks in the post-M&A period have increased their market share and gained greater market power (Angelini & Cetorelli, 2003; Fernández de Guevara et al., 2005).

Increased concentration in the Vietnam banking industry has come to the attention of researchers who have begun to measure banking competitiveness using the Lerner index (Coccorese, 2014; Nguyen, 2018; Nguyen & Nghiem, 2018; Nguyen, Nghiem, Roca, & Sharma, 2016b; Phan, Daly, & Akhter, 2016). Their results show that Lerner indexes range from 0.158 to 0.21 from 1995 to 2016. Le (2014) measured the market concentration of the Vietnam banking industry based on concentration ratio Herfindahl-Hirschman (HHI) index. The results show that the HHI index dropped from 0.13 in 2007 to 0.077 by 2012. This outcome suggests that the Vietnam banking industry has changed from a moderately concentrated to a less concentrated market (Le, 2014).

Second, Hoang et al. (2016) argues that a monopoly or a group of monopolies may occur as a result of M&As because of a decrease in the number of banks (or bank size is concentrated). They note that monopolies can harm the competitive environment of the banking industry (Hoang et al., 2016). The reason is that some banks may exist in an environment of low competition (or monopoly) whereas other banks may face high levels of competition. The first banking group can set higher prices for its financial products and so be profitable (Nguyen, 2018). In other words, these banks have higher market power than other banks and tend to become monopolists.

There are limited studies that investigate monopolist behaviour of the Vietnam banking industry. Nguyen and Nghiem (2018) is the first study that compares bank market power using the Lerner index between the Vietnam state-owned and joint-stock banks. Their results reveal no significant differences in bank market power between state-owned and joint-stock banks (Nguyen & Nghiem, 2018). This suggests that neither the Vietnam state-owned banks nor joint-stock banks are strong enough to become monopolists. In a competitive environment, larger banks survive better and tend to acquire smaller banks to maintain their dominant position as well as asset quality in the market (Tabak, Fazio, & Cajueiro, 2012; Wang, 2015). From 1997 to 2015, a total of 18 M&As occurred in the Vietnam banking industry (Hoang et al., 2016). Four of the 18 M&As were state-owned banks merging with joint-stock banks (Hoang et al., 2016). In short, a high market power bank group may exist. This group may include state-owned banks and joint-stock banks. However, no study has examined whether such a high market power bank group exists.

An artificial neural network (ANN) is a computational program that illustrates the relationships between inputs and outputs as a complexity model (Blanco, Pino-Mejías, Lara, & Rayo, 2013; Trinkle & Baldwin, 2016). The model consists of a set of units (nodes) that represent a set of neurons (Kiruthika & Dilsha, 2015). The neurons are interconnected with neighbours by weights that expose the strength of the connection. ANNs are classified mainly into competitive networks (unsupervised self-organisation map (SOM)); networks for nonlinear prediction and classification (multilayer perceptron (MLP) network); and networks for time series forecasting (Samarasinghe, 2006). The unsupervised machine learning technique, the SOM technique, is a type of neural network that was developed by Kohonen (1982). The SOM is used to group similar observations into clusters (Samarasinghe, 2006; Tsai & Chen, 2010). After training, the input data segregate into clusters with those with highest similarity and highest dissimilarity (Tsai & Chen, 2010).

Previous studies (see for example, Du Jardin & Séverin, 2011; 2012; Chen, 2012) use the SOM to chart trajectories that reflect dynamic changes in a company's finance. Du Jardin and Séverin (2011; 2012) investigated multiple French companies from 1995 to 2003 and Chen (2012) evaluated Taiwanese companies from 1999 to 2010. The SOM technique can be used to categorise companies into a super-class of healthy companies and bankrupt companies (Chen, 2012). Using the SOM technique to track financial trajectories can aid experts to assess companies' current financial situation and to observe financial development over time (Chen, Ribeiro, Vieira, & Chen, 2013). This study uses the SOM technique to track financial trajectories and categorise Vietnam banks into super-class groups. Applying the SOM technique is more suitable for categorising banks into groups because the technique does not distinguish between state-owned and joint stock commercial banks.

Third, M&As impact also on bank performance. Angelini & Cetorelli (2003) note that when banks are faced with an increase in market competitiveness, they may opt to consolidate (or M&A) to improve their performance. Therefore, it is necessary to examine the efficiency of the Vietnam banking industry in the post-M&A era. After analysing the competitiveness and efficiency of the Vietnam banking industry, it is important to investigate the impact of determinants on bank market power and efficiency. These determinants may be bank-specific characteristics (bank size, equity to asset ratio, non-performance loan ratio), macro-economic conditions (GDP growth rate and the inflation rate) (Abel & le Roux, 2017; Adjei-Frimpong, Gan, & Hu, 2016; Anwar, 2018; Erler, Gischer, & Herz, 2017). Numerous researchers have studied the effects of determinants on bank efficiency (Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016). However, there is a scarcity of studies that investigate the determinants of bank market power in Vietnam banks. Identifying such determinants could aid Vietnam banks to identify the critical factors to boost both their market power and efficiency.

In summary, after the 2008 GFC, credit growth rate stagnation and high NPLs have negatively impacted on Vietnam banks. The "Restructuring Financial Institutions 2011-2015" programme was implemented in March 2012 to restructure the banking industry using an M&A strategy (Le, 2014). M&As have impacted significantly on bank market power as well as efficiency. However, M&As may lead to the existence of a group of high market power banks (or monopolists) that may harm the competitiveness of the environment (Hoang et al., 2016). Therefore, identifying the existence of a group of high market power banks would enable policy-makers to maintain an efficient banking industry. This study uses the SOM technique to categorise Vietnam banks into super-class groups. This SOM technique, with an unsupervised algorithm, can categorise and identify any group(s) of high market power banks that may exist in Vietnam's banking industry. This study also investigates the determinants that impact on Vietnam bank market power and efficiency.

Our SOM results identifies two super-class groups of banks (1 and 2) based on unbalanced yearly data from 2008 to 2017. Using the non-structural indicator, the Lerner index, to capture market power and the Data Enveloped Analysis technique to measure bank efficiency, our result shows significant differences in Lerner scores (which represent bank market power) of the two groups of banks. Differences in the Lerner scores provide evidence of a group of strong banks that is isolated from other banks. This implies that this strong group of banks have the potential to be monopolists and impairs Vietnam's competitive banking environment.

Our results also show that banks maintain their market power over a longer period of time. This finding indicates that the Vietnam banking industry is characterised by non-transparent information, networking lending relationships and limited regulations exhibiting monopolistic practices: this ultimately leads to bank monopolies and reduces competition and innovation. Bank size and the capitalisation ratio have a positive relationship with market power. A higher NPL weakens both the efficiency and market power of Vietnamese banks. To reduce the NPL ratio, our study uses retail customer dataset. The next section will discuss in retail banking in Vietnam banks.

1.3 Retail Banking

1.3.1 The Vietnam Retail Banking Industry

As incomes increase, people are likely to spend money to enhance their living standard. The Vietnam retail banking market reported an average growth rate of 20% per annum from 2010 to 2016. This rate is expected to rise even more in the next several years (FirstData, 2016). The following factors explain why retail customers have recently become profitable to Vietnam banks:

- Only 27% of Vietnam's total population, or approximately 89.6 million citizens, have a personal bank account (Lin, Wu & Tran, 2015). Young people, who constitute 45% of the total population, are the retail customers who are more likely to adopt a modern lifestyle, have internet banking and have increased income and spending power (Tran et al., 2015). Banks need to focus on this ever-expanding group of retail customers.

- The number of internet users has increased throughout the country. Banks can benefit from the e-commerce revolution (Barbour-Lacey, 2014). Internet banking provides numerous benefits: (i) E-banking is a lower-cost distribution channel and has a significant impact on the banking market (Pham, Cao, Nguyen, & Tran, 2013); and (ii) E-banking can help banks expand their market share (Lin et al., 2015).

Increasing numbers of internet users and emerging e-commerce, provide new opportunities for Vietnam banks looking for long-term development in retail banking. This phenomenon also explains why some international commercial banks have started to operate in the retail sector. Aggressive foreign bank competitors such as HSBC, ANZ and Standard Chartered have quickly established and expanded their services in the Vietnam bank retail market (KPMG, 2013). Table 1.1 lists the proportion of retail loans in terms of total outstanding loan value of Vietnam's banks. The mean value has increased steadily from 30% to 45% between 2011 and 2017. This suggests that Vietnam banks are more aware of the profitability of retail customers and have begun to expand their services.

Table 1.1 The Proportion of Retail Loans in the Loan Portfolios of Vietnam Banks 2011 to 2017

Year	2011	2012	2013	2014	2015	2016	2017
Mean value	30%	34%	35%	40%	42%	44%	45%
Standard deviation	15%	18%	18%	22%	19%	16%	16%
Minimum	10%	5%	14%	12%	18%	15%	20%
Maximum	62%	78%	87%	82%	82%	78%	75%

Note: This table represents the proportion of retail loans in relation to the total outstanding loans of the 27 Vietnam banks.
Source: Author's summaries from annual Vietnam bank annual reports.

However, domestic Vietnam-owned banks seem to lag behind foreign banks in terms of technological systems. Vietnam's retail banks use credit rating systems such as the credit scoring model (CSM) to rate borrowers' creditworthiness (Dinh & Kleimeier, 2007; Tran & Ho, 2015). According to a KPMG survey (2013), Vietnam banks have used an internal credit rating system developed four or five years ago and applied the same risk system without considering different business sizes and objectives, such as market share and customer segment. Less than 50% of Vietnam banks are satisfied with their current credit internal rating system and 'poorly' designed CSMs (KPMG, 2013). This is a major factor contributing, in part, to the high level of non-performing loans in Vietnam. In short, the survey revealed that Vietnam banks need to realise their internal weaknesses; that their internal risk rating systems are 'poorly designed'; and improve their risk management policies and systems.

1.3.2 The Credit Scoring Model

The CSM is currently used by many banks worldwide and is becoming an important tool for measuring borrowers' creditworthiness (Chi & Hsu, 2012). For instance, 90% of lending institutions in the US use FICO scores to evaluate borrowers creditworthiness (FICO, 2015). The model illustrates the links between borrowers' characteristics and their behaviour and the risks they present to the lender (Schreiner, 2004). The CSM is an automatic system built with sophisticated statistical or intelligent techniques to evaluate applicants' creditworthiness. The primary objective of CSM is to assign applicants into two groups: 'non-default', who are likely to repay their loans, and 'default', who have a tendency to default on his/her financial obligations (Cubiles-De-La-Vega, Blanco-Oliver, Pino-Mejías, & Lara-Rubio, 2013). Hence, CSM is a form of classification analysis (Abdou, Pointon, & El-Masry, 2008).

Applying the efficiency CSM to assess the creditworthiness of borrowers, banks benefit in several ways. First, it reduces the time and cost associated with the credit evaluation. It allows faster credit decision-making and the recognition of immediate credit risks or losses (Cubiles-De-La-Vega et al., 2013; Kiruthika & Dilsha, 2015). Second, another approach to increase a bank's competitive advantage is by using a "risk-based pricing" strategy. After classifying applicants into 'good' and 'bad' borrower groups, banks can decide whether to offer applicants credit. Banks can choose to offer better services with a lower price to 'good' borrowers, charge high risk borrowers a higher interest rate and avoid providing credit to the highest risk or 'bad' borrowers, or implement the "risk-based pricing strategy" (Getter, 2006).

As a result of the expanding credit market, banks have become increasingly interested in CSM. Three technical approaches have been developed to increase the accuracy of CSM, include conventional statistics, advanced computing and combination techniques. First, researchers often use conventional techniques such as linear discriminant analysis, probit analysis, and logit regression to build a CSM. For example, Abdou et al. (2008) and West (2000) compared the classification power of logit regression, probit analysis and discriminant analysis and conclude that logit regression is the most accurate technique among traditional methods. Baklouti and Bouri (2014); Dinh and Kleimeier (2007) and Tran and Ho (2015) confirm that logit regression is an efficient tool for credit scoring. In short, prior scholars conclude that logit regression is the best conventional technique, which is broadly used to build CSM.

Second, researchers such as Abdou et al. (2008); Blanco, Pino-Mejías, Lara, and Rayo (2013); and Cubiles-De-La-Vega et al. (2013) apply statistical techniques such as logit regression and advanced computing ANN (that is an MLP network) to identify the default borrower group and measure the classification power of the methods. They all find that the MLP network outperforms conventional statistics in accuracy assessments of credit risk for banks. The MLP technique is classified as a supervised neural network (Samarasinghe, 2006). There are many types of supervised ANNs, such as probabilistic neural networks and MLP networks, but, in the finance discipline, MLP is commonly used; it is used 75% of the time compared with other types of ANN technique (Jagric & Jagric, 2011). The supervised machine learning algorithm is a learning strategy to adjust the interconnections (or weight) between given input data and desired output through training rules, so errors are minimised (Samarasinghe, 2006). However, a downside of ANN is that the technique cannot clearly interpret the relationship between input and output (Trinkle & Baldwin, 2007; 2016). This major disadvantage is the main reason why the ANN technique has not been widely applied (Thomas, 2010).

Third, other researchers suggest that future researchers may seek to use a combination of techniques (see for example, Thomas, 2010; West, Dellana and Qian, 2005; Yu, Li, Tang, Zhang and Kou, 2015). Two combination methods in which researchers may find more efficient to build a CSM are the ensemble and hybrid methods (Lu, Liyan, & Hongwei, 2013; Thomas, 2010; Tsai, 2014; Tsai & Chen, 2010; West et al., 2005). As any global recession continues, banks need to increase profits while minimising losses. They need better credit evaluation methods that enable them to identify 'good' and 'bad' borrowers as well as a better method of arriving at credit granting decisions. Researchers have developed various methods to improve the overall accuracy of CSM with clear interpreting relationships between critical factors and borrowers' creditworthiness. Nevertheless, the use of conventional statistics or advanced computing or a combination of techniques to build CSMs that produce greater accuracy and/or classificatory power remains the subject of ongoing debate and challenges (Lessmann, Baesens, Seow, & Thomas, 2015; Thomas, 2010; Yu et al., 2015).

In summary, retail banking is a significant opportunity that Vietnam banks can embrace for long-term growth. However, as this chapter has suggested, the internal credit-granting system (or CSM) used by many Vietnam banks has flaws and is considered "poorly-designed". This is a major reason for the high level of non-performing loans in Vietnam banks. Therefore, understanding the retail bank lending process, the critical factors that affect the borrowers' creditworthiness, and investigating more efficient techniques for building a CSM in the retail banking sector are essential for the survival of Vietnam's banking sector.

Our retail banking results show that the behaviour of borrowers has changed. When a loan is issued after one year, the probability of default of this loan decrease by 0.92%. Loan value is a risk factor. An increase in loan value increases the default rate by 0.25%. An increase of one year in the duration also increases the probability of default by 0.18%. The result also shows that the integrated model (a combination of logit regression and MLP) outperforms logit regression. In short, the hybrid technique is a better option for bank executives to build a more efficient CSM.

Problem Statement and Objectives.

Vietnam's economy has been experiencing a period of slower development. As a result of a decline in economic growth, the Vietnam banking sector faces challenging circumstances, such as stagnation in the credit growth rate and high NPL levels. The "Restructuring Financial Institutions 2011-2015" programme (Decision no.254/QD-TTG) was designed to improve bank performance using international banking standards (Le, 2014; Nguyen et al., 2014; Nguyen et al., 2016a;). The key element of the plan is M&As that have impacted significantly on bank competitiveness and efficiency (Le, 2014). M&As decrease the number of banks, lead to concentrated bank size and hence may result in the rise of a group of strong (high market power) banks (or a monopolists) in the market. This group of banks may impede the competitive environment of the Vietnam banking industry (Hoang et al., 2016). However, no study has provided evidence for the existence of such a group of strong (high market power) banks.

The consumer retail market is increasingly important for Vietnam banks in terms of their long-term growth. However, after the 2008 GFC, bank risk management systems, particularly the internal rating system or the CSM, have been found to be poorly designed which has contributed, in part, to the extraordinary level of the non-performing loans in Vietnam banks. Bank executives have identified the link between risk management and profitability and agreed to allocate investment to improve the credit risk management systems (KPMG, 2013). Therefore, it is vital for Vietnam banks to investigate the critical factors in bank lending and develop a robust credit risk model to minimise the default rate and failures.

The CSM has recently played a significant role and is widely used as the most successful mechanism for banks wanting to evaluate consumer credit risk (Abdou & Pointon, 2011; Chuang & Lin, 2009; Lee & Chen, 2005; Šušterši, Mramor, & Zupan, 2009). However, the conventional statistical and advanced computing techniques used to build the best discriminant CSMs are challenging (Thomas, 2010). In terms of the classification power among conventional statistical methods, logit regression is considered the most accurate and therefore is the preferred technique (Abdou et al., 2008; West, 2000). Alternative advanced computing techniques, such as ANN, have recently been shown to be more accurate in predicting 'bad' borrowers. However, the technique has not been used widely because of its inability to interpret input variables' contributions to output (Thomas, 2010). Instead, scholars suggest that a combination of techniques such as hybrid and ensemble techniques may be more efficient (Thomas, 2010; West et al., 2005). Increasing the overall accuracy of CSMs is important for financial institutions to both optimise profit and minimise losses. However, an optimising methodology to build robust CSMs that produce far better classificatory power than current methods is subject to ongoing debate (Lessmann et al., 2015; Thomas, 2010; Yu et al., 2015).

Using the SOM technique, this study investigates the existence of groups of strong (high market power) banks (or monopolists) that may harm bank competitiveness. This study also identifies the determinants that influence bank market power and efficiency. Understanding the determinants of market power and efficiency can aid bank managers in selecting appropriate strategies for long-term development. This study also explores the critical factors in lending within Vietnam retail banks and proposes a more accurate method that can clearly interpret the relationship between the input and output variables when building a CSM. Understanding the critical factors in bank lending decisions will enhance credit-granting decisions and prevent losses through wrongly evaluating borrowers' creditworthiness.

Research Objectives

I. To investigate financial trajectories and categorise Vietnam banks using the SOM technique.
II. To measure and compare different groups of banks' market power and efficiency as categorised by the SOM technique.
III. To investigate the determinants that impact Vietnam bank market power and efficiency.
IV. To identify critical factors influencing the lending decision in Vietnam retail banking.
V. To investigate the classificatory power of the CSM based on a hybrid technique using a combination of artificial neural network and logit regression.
VI. To investigate factors that determine the interest rate charged on retail loans.
VII. To examine how bank ranked their customers to offer a loan price

Research questions:

1. What are the financial trajectories and categorise Vietnam banks using the SOM technique?
2. Are there any statistical differences in Vietnam bank's market power and efficiency based on the SOM technique?
3. What determinants impact Vietnam bank market power and efficiency?
4. What factors influence the lending decision in Vietnam retail banking?
5. What is the classificatory power of the CSM based on a hybrid technique using a combination of artificial neural network and logit regression?
6. What factors determine the interest rate charged on retail loans?
7. How banks bank ranked their customers offer a loan price?

1.4 Conceptual framework

Based on the New Empirical Industrial Organisation Theory (NEIO), this thesis developed a conceptual framework to address the research problems in the Vietnam banking industry (see appendix A0). The NEIO assumes that entering/exiting fencing rules and existing players (or banks) will affect the competitive environment (Liu et al., 2012; Nguyen et al., 2016b). On 1 March 2021, the Vietnamese government released the "Restructuring Financial Institutions 2011-2015" programme, which encourages M&A activities as a key strategy to improve the competitiveness and efficiency of Vietnam banks (Hoang, Phan, & Bandaralage, 2016; Le, 2014). These M&A deals decreased the number of Vietnam banks from 43 to 32 between 2008 and 2017 (Hoang et al., 2016). According to the NEIO thoery, a decrease in the number of banks has a significant impact on market competition levels within the Vietnam banking industry. Moreover, a monopoly or a group of monopolies may transpire as a result of M&As because of a decrease in the number of banks and harm the competitive environment of the banking industry (Hoang et al., 2016).

In addition, successful M&A deals may have impacted the efficiency of Vietnam domestic banks. For example, a merger between large bank (SCB) and small banks (Tin Nghia Bank and First Commercial Bank) occurred in 2011. This merger increased the assets (a new SCB) 2.5 times (from 60,183 billion VND to 149,205 billion VND), and 2.4 times for charter capital (from 4,710 billion VND to 11,370 billion VND). The new SCB almost doubled its staff numbers (from 2075 to 3959 employees). The NPL rate (of SCB) dramatically increased from 1.28% to 7.25% from 2009 to 2012 (Hoang et al., 2016; SCB, 2012). As a result, this merger may increase bank size, but impact negatively on bank efficiency because of increases in NPL and operating costs. In other words, M&A deals can modify all determinants impact on bank market power and efficiency.

1.5 Research Contributions

Most previous studies on bank market power and efficiency differentiate between state-owned and joint stock commercial banks (Nguyen & Nghiem, 2018; Vu & Turnell, 2010). However, these studies do not show any significant difference in bank market power and efficiency between these bank groups (state-owned and joint stock banks). Therefore, this study uses the SOM technique to categorise Vietnam banks into super-class groups. This technique is more appropriate because it is based on an unsupervised algorithm and does not distinguish between state-owned and joint stock commercial banks. This is the first study to use a categorising methodology to examine differences in bank market power and efficiency in the Vietnam banking industry. In addition, this is the first study to use the SOM technique to build the financial trajectories of the Vietnam banking industry.

Previous studies have examined the determinants that affect Vietnamese bank efficiency (Gardener, Molyneux, & Nguyen-Linh, 2011; Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016). However, very few studies investigate the factors impacting bank market power. This study identifies the determining factors that impact both bank market power and efficiency. Identifying these factors will enable bank managers to boost their market power and efficiency. This study applies dynamic OLS and the strict exogeneity test to examine the dynamic relationships and endogeneity problems. A dynamic GMM estimator is used to estimate consistent, unbiased coefficients. Finally, this study uses a robustness check to verify the plausible and robust determinants impacting market power and efficiency.

The study contributes to the credit risk assessment method literature by using a hybrid technique (a combination of ANN and logit regression) to build a more robust CSM. The hybrid method increases the overall accuracy of the classification power of credit granting systems and the interpretation of the relationships between critical factors and borrowers' probability to default. These research findings can be used by bank executives to improve the overall accuracy of their credit granting systems.

Limited studies have been conducted on retail lending in Vietnam, particularly on risk-based pricing strategies. This study investigates whether Vietnam banks use risk-based pricing strategies to determine customer interest rates.

1.6 Organisation of the Thesis

This thesis is divided into six chapters. Having provided the background in Chapter 1, Chapter 2 discusses the literature on the SOM technique and methodologies commonly used to measure bank market power and efficiency. This chapter also discusses methodologies used to build a CSM. Chapter 3 reviews the methods used to build financial trajectories and explore the determinants of bank market power and efficiency using a hybrid technique that integrates logit and ANNs to explore the risk factors for default borrowers and enhance classificatory power. Chapter 4 discusses the empirical outcomes for the Vietnam banking industry, including the empirical bank financial trajectory patterns, bank market power (using the Lerner index), bank efficiency (using technical and cost efficiency scores) and determinants' impact on bank market power and efficiency. Chapter 5 presents the results of the Vietnam retail banking analysis, including risk factors associated with default borrowers, a hybrid technique that enhances classificatory power and the risk-based pricing strategy. Chapter 6 summarises the study's key findings and discusses policy implications.

Chapter 2

Literature review

2.1 Introduction

This chapter analyses the literature on the banking industry, focusing on the self-organisation map (SOM) technique for building financial trajectories and measuring bank market power and efficiency. This study chapter discusses the literature on retail banking, such as the credit scoring model (CSM) and risk-based pricing (RBP) strategy.

The chapter is arranged as follows. Section 2.2 reviews the SOM technique that is used in this study to build financial trajectories. Section 2.3 analyses the literature on bank competition and efficiency in industrialised and emerging countries, including Vietnam. Section 2.4 discusses the determinants that affect bank market power and efficiency. Section 2.5 analyses the concept of the CSM, highlighting its benefits and its shortcomings, and the techniques used to build a CSM. Section 2.6 analyses literature on the RBP strategy in banks in developed countries. Section 2.7 summarises the chapter.

2.2 The SOM Technique For Tracking Financial Trajectories

This section discusses the literature linked to research objective one; it reviews the literature that focuses on the SOM technique and its use in building financial trajectory patterns. As Du Jardin and Séverin (2011) emphasised, bankruptcy forecasting is a dynamic process. The process is more correctly positioned as continuous forecasting rather than a snap point prediction (Du Jardin & Séverin, 2011). This is a major problem when evaluation is limited to a single year, but debt is repaid over a longer period, such as several years (Du Jardin & Séverin, 2011). The reason is that the risk of a default borrower or bankruptcy may transpire more than a year after the appraisal (Du Jardin & Séverin, 2011). For example, company executives might ask investors to give them more time to improve their financial health but, after the grace period, they may find it impossible to recover, leading to bankruptcy (Du Jardin & Séverin, 2011). The major shortcoming of forecasting techniques (such as linear, nonlinear or classification regression) is that they have a horison time that is very short and does not exceed one year (Du Jardin & Séverin, 2012). Moreover, when forecasting for over one year, the accuracy of the technique decreases dramatically. For example, Altman's model has an accuracy of 95% in one-year forecasting but drops to 48% for three year forecasts (Du Jardin & Séverin, 2011).

Because these limitations, researchers developed a combination technique that uses the SOM (or Kohonen map) to track a company's financial trajectory, also known as *the SOM trajectory technique* (Chen et al., 2013; Du Jardin & Séverin, 2012). Those authors' primary purpose is to improve the accuracy of the forecasting technique over a specific period, not just the accuracy in snapshot forecasting (Du Jardin & Séverin, 2011). There are several advantages to using the SOM technique to build a financial trajectory; they are explained below:

First, the SOM trajectory is a user-friendly imagining technique for exploring financial reports (Chen, 2012; Chen et al., 2013). The SOM is frequently used to visualise clustering applications and analyse high-dimensional financial data in a 2-dimensional (2D) plot (Chen, 2012). The SOM technique enables researchers/executives to observe a company's changing position on a trajectory; in short, the major difference is that the SOM allows a dynamic view of changed financial status rather than snapshot forecasting (Chen, 2012; Du Jardin & Séverin, 2011). This method can aid experts analyse trajectory patterns of a bankrupt or healthy corporation and better understand the phenomenon of bankruptcy (Chen et al., 2013).

Second, the SOM trajectory method has an advantage, in the medium term, of forecasting and detecting financial threats (Du Jardin & Séverin, 2011). This is a valuable tool for a financial institutions to use when assessing the risk of debt that is longer-term than the evaluation (Du Jardin & Séverin, 2011). In addition, this method may help executives measure the financial health of their company and take immediate corrective actions (Chen et al., 2013). This technique also assists experts identify downward financial trends, over time, and enables them to anticipate the risk of bankruptcy (Chen et al., 2013).

Table 2.1 The Literature on The SOM Trajectory and Its Application to Build Financial Trajectories

Author(s)	**Dataset**	**Study period**	**Methodology** Multi-period		Single period	**Results**
Du Jardin and Séverin (2011)	French companies	1995 – 2003	SOM trajectory	Cox	discriminant, logistic regression, MLP	SOM trajectory is more stable
Du Jardin and Séverin (2012)	French companies	1991 – 2009	SOM trajectory	Cox	Discriminant, logistic regression, MLP	SOM trajectory is more stable
Chen (2012)	Taiwanese companies	1999 - 2010	Hybrid SOM-SVR			Hybrid SOM-SVR more accuracy
Chen et al. (2013)	French companies	2003 - 2006	SOM-based trajectory			

Note: SOM trajectory = the SOM technique's use for building financial trajectories; MLP = Multilayer perceptron technique; Cox = Cox survival analysis, a method designed to handle time-series data; SVR = support vector regression machines which was designed by Drucker, Burges, Kaufman, Smola and Vapnik (1997)
Source: Author's summary of previous literatures

Table 2.1 lists the studies that used the SOM technique to create financial trajectories. Chen et al. (2013) and Du Jardin & Séverin (2011; 2012) all collected sound and unsound firm samples from French companies from 1991 to 2009. The authors discover that the SOM trajectory technique is not only more stable than other techniques (discriminant analysis, logit regression and multilayer perceptron) over time but is also a valuable tool for banks to assess investment risks in the medium-term (longer than one year) (Du Jardin & Séverin, 2011, 2012). In addition, Chen (2012) gathered data from Taiwanese firms listed on the TSEC from 1999 to 2010 to track their financial trajectories. The results show that the hybrid SOM-SVR technique is more accurate and easier to use than other methods, such as single SOM, support vector machine, and learning vector quantisation (Chen, 2012).

The SOM trajectory technique involves two steps: the static and dynamic phases (Chen, 2012; Du Jardin & Séverin, 2011; 2012). In the static phase, the SOM technique is used to cluster a company's financial status. The technique locates a company into a neuron in the 2D SOM map based on the company's financial statements (Chen, 2012; Du Jardin & Séverin, 2011). The number of neurons might be huge and need to be grouped into super-class clusters (Du Jardin & Séverin, 2011; 2012). These clusters might be healthy or bankrupt groups (Chen, 2012; Du Jardin & Séverin, 2011).

Based on their financial status in consecutive years, companies are positioned in every 2D map over continuous years in the dynamic phase. When overlapping these maps into one 2D SOM map, the movement of a company can be viewed within the clusters over time. For example, if a company A has annual financial reports from three consecutive years, the SOM technique will locate the company's financial status to be X1(a1, b1) in a 2D map of year 1, X2(a2, b2) in a 2D map of year 2 and finally X3(a3, c3) in a 2D map of year 3. The three 2D maps of years 1, 2, and 3 are overlaid to create one 2D map. Thus, Chen (2012) joins three points, X1 to X2 and X2 to X3, to draw the dynamic trajectory for the company A over three years. Analysing companies' financial trajectories over consecutive years can reveal healthy or bankruptcy patterns (Chen, 2012). As a result, the SOM trajectory technique is a good tool for observing and rating companies' dynamic financial situation (Chen, 2012).

No studies use the SOM trajectory technique to build the financial trajectories of Vietnam's banks. This is the first analysis that attempts to examine the dynamic financial status of Vietnam's banks using the SOM trajectory technique. The SOM trajectory technique can provide a user-friendly image with which to explore bank financial reports. The results are expected to support bank executives in observing their bank's financial condition over time by analysing financial trajectory patterns. This will ultimately help bank executives understand their financial circumstances and identify potential risks, which means they can implement corrective action (Chen, 2012; Du Jardin & Séverin, 2011; 2012). In addition, the SOM trajectory technique can categorise Vietnam banks into super-class groups. Most previous studies classify Vietnam banks into state-owned and joint stock banks to compare bank market power and efficiency (Nguyen & Nghiem, 2018; Vu & Turnell, 2010). However, their results demonstrate that market power and efficiency are not different (insignificant statistical difference) between these groups (state-owned and joint stock banks). This study provides a new grouping approach for categorising Vietnam banks. It can be used for further analysis in terms of benchmarking market power and efficiency between these super-class bank groups. The next section discusses the methodologies that measure bank market power and efficiency.

2.3 Bank Market Power and Efficiency

This section examines the literature relevant to the study's second objective, which is to quantify competitiveness and efficiency levels in the banking industry. This section covers the literature on both emerging and industrialised nations and then focuses on Vietnam banks.

2.3.1 Measuring Bank Market Power

There are two ways to quantify bank competitiveness level in the literature: (1) the structural and (2) non-structural approach (Ab-Rahim, 2017; Liu, Molyneux, & Nguyen, 2012; Nguyen, 2018). The structural model is based on the "structure–conduct–performance" and "efficient structure" theories (Ab-Rahim, 2017; Nguyen, 2018; Nguyen et al., 2016b). Both theories argue that market concentration determines the competitiveness level in the market (Ab-Rahim, 2017; Adjei-Frimpong et al., 2016; Liu et al., 2012). These theories assume that banks are in markets that are composed of a few large players can offer a higher price for their financial products than banks in markets that have many players (Liu et al., 2012). The Herfindahl–Hirschman (HHI) index or the concentration (CR) score have been used to assess the competitiveness level in the structural approach (Nguyen, 2018; Nguyen & Nghiem, 2018; Nguyen et al., 2016b). The advantage of the concentration ratios (HHI or CR) is that they require low levels of data and are easily valued at industrial or national levels (Nguyen, 2018). However, there are disadvantages associated with these concentration ratios. First, a higher level of concentration might lead to a collusive increase among banks (Nguyen, 2018). Efficient banks tend to grow their market share that results in a higher market concentration. For this reason, the concentration index becomes endogenous (Nguyen, 2018). As a result, the structural approach has recently been applied to research on the banking industry in several countries. However, this approach also has its limitations (Nguyen, 2018; Nguyen & Nghiem, 2018).

The non-structural model basically depends on the New Empirical Industrial Organisation Theory (NEIO) as a new approach to serve banks in emerging markets (Nguyen, 2018; Nguyen & Nghiem, 2018). The non-structural approach measures bank market power at a single level without considering market organisation (Adjei-Frimpong et al., 2016; Nguyen, 2018). The non-structural approach assumes that entering/exiting fencing rules and existing players (or banks) will affect the competitive environment (Liu et al., 2012; Nguyen et al., 2016b). Two main indicators are used to determine the competition level under the non-structural approach: the H-statistic and the Lerner index (Nguyen, 2018; Nguyen & Nghiem, 2018; Nguyen et al., 2016b). The Boone index was proposed by Boone (2008) to quantify competitiveness level. It concentrates on the market's dynamics (Nguyen, 2018). The major disadvantage of these two methods (H-statistic and Bond index) is that these indicators do not measure market power continuously (Nguyen & Nghiem, 2018). This is because estimating these indicators requires data from the whole study period to identify different types of competition (such as monopolistic competition, monopoly or perfect competition) (Nguyen, 2018; Nguyen & Nghiem, 2018). Hence, these indicators are not appropriate to monitor competition level over time (Nguyen & Nghiem, 2018). In contrast, the Lerner index indicates market power at a continuous, individual level (Adjei-Frimpong et al., 2016; Nguyen, 2018; Nguyen & Nghiem, 2018). The Lerner index represents market power, which has the opposite meaning from competition; higher market power suggests lower competition (Nguyen et al., 2016b).

Table 2.2 Panel A summarises the literature that measures bank market power in emerging and developed nations and on Vietnam's banks (Panel B). The Lerner score is frequently used to quantify the market power of banks. This indicator quantifies the market power of banks in industrialised countries such as European countries, Italy, Spain and America (Angelini & Cetorelli, 2003; Delis, 2012; Delis & Erler et al., 2017; Pagoulatos, 2009). It is also used to appraise bank market power in emerging nations such as China, Bangladesh, Ghana, Indonesia, India, Malaysia, the Philippines and Zambia (Abel & le Roux, 2017; Adjei-Frimpong et al., 2016; Phan et al., 2016; Simpasa, 2013). The Lerner index has the following advantages:

- Compared with other concentration measures that evaluate competition at the industry level, the Lerner index can be used to appraise each bank or provide a continuous measurement for each year (Abel & le Roux, 2017; Adjei-Frimpong et al., 2016; Nguyen, 2018; Nguyen & Nghiem, 2018). Hence, the results can be used as a dependent variable in a subsequent analysis to evaluate the determinants that impact bank power (Delis & Pagoulatos, 2009).

- The Lerner index can be used to identify the evolution of power level (growth, fall, or continuous increase of market power) (Fernández De Guevara & Maudos, 2007). This indicator also reflects the profitability of an individual bank because the indicator is processed by the change in the 'output price - cost margin' divided by output price (Nguyen, 2018). The 'output price-cost margin' can be used to assess profitability (Nguyen, 2018). Higher profitability might imply that the bank has higher market power or lower competition.

Table 2.2 Use of the Lerner Indicator to Measure Bank Market Power in Emerging and Industrialised Nations

Author(s)	Study period	Country(ies)	Measure market power	Value	Note
PANEL A – Measuring Bank Market Power (or Lerner Index) in Various Countries					
Angelin and Cetorelli (2003)	1984-1997	Italy	Lerner		
Fernández De Guevara & Maudos (2007)	1986-2002	Spain	Lerner		
Delis and Pagoulatos (2009)	1984-2005	EU countries	Lerner		
Soedarmono, Machrouh, and Tarazi (2011)	1999-2007	12 Asian countries	Lerner		
Delis (2012)	1987-2005	84 countries	Bond & Lerner		
Simpasa (2013)	1998-2011	Zambia	H-statistic, Lerner		
Adjei-Frimpong et al. (2016)	2001-2010	Ghana	Lerner		
Erler et al. (2017)	2004-2013	America	Lerner		
Abel and le Roux, (2017)	2009-2014	Zambia	Lerner		
PANEL B – Measuring Bank Market Power (or Lerner Index) in Emerging Countries (including Vietnam)					
Coccorese (2014)	1995-2002	87 countries	Lerner	0.166	
Phan et al. (2016)	2005-2012	Bangladesh, India, Indonesia, Malaysia, Philippines, Vietnam	Lerner	0.221	
Nguyen et al. (2016b)	1995-2011	Vietnam	Lerner	< 0.45	
Nguyen and Nghiem (2018)	2000-2014	Vietnam	Lerner	0.158	Compared market power between state and J.S. banks.
Nguyen (2018)	2006-2016	Vietnam	Lerner	0.2007	
Fred (2019)	2008-2015	Vietnam	Lerner	0.215	

Note: EU = European countries; J.S. = joint stock
Source: Author's summary of previous studies.

Previous studies have measured competition in Vietnam banks using the Lerner index (Coccorese, 2014; Fred, 2019; Nguyen, 2018; Nguyen and Nghiem, 2018; Nguyen et al., 2016b; Phan et al., 2016) (see Table 2.2, Panel B). These studies have shown that for different time periods from 1995 to 2016 the mean Lerner index for Vietnam banks ranges from 0.158 to 0.215. Nguyen and Nghiem's (2018) study is the only one that offers a comparison of that market power of commercial state and joint-stock Vietnam banks (see Table 2.2, Panel B). The authors used the two-sample t-test to check whether the market power of state and joint-stock Vietnam banks is different. Their results reveal that the difference in market power between commercial state and joint-stock Vietnam banks is *insignificant* (Nguyen & Nghiem, 2018). In short, the commercial state and joint-stock Vietnam banks exhibit equal market power and have similar competitiveness levels.

To date, no study has evaluated the market power of various bank groups (the super-class bank groups) using the SOM trajectory technique. This study compares the market power of the super-class bank groups in Vietnam using the two-sample t-test. If the p-value of the test is statistically significant, it can be assumed that the SOM trajectory technique has identified a difference in market power within the super-class bank groups.

2.3.2 Measuring Bank Efficiency

Bank efficiency is of interest to various bank stakeholders (bank excecutives and policy makers) as well as academic researchers (Kočišová, 2014; Nguyen et al., 2016a). This is because the banking industry has a central role in national development (Nguyen et al., 2016a). Efficient banks stabilise the banking industry and a nation's monetary system (Kočišová, 2014). Moreover, bank excecutives are always interested in benchmarking or comparing their bank's efficiency with the best operating bank to improve their operations (Nguyen et al., 2016a). Pure technical efficiency (PTE) and cost efficiency (CE) are frequently used to quantify bank efficiency. The PTE method appraises executives' managerial skills in terms of how they use inputs to maximise outputs (Adjei-Frimpong, Gan & Hu, 2014). In contrast the CE method is used to capture how banks manage their costs compared with the optimal costs produced by best-practice (Kočišová, 2014; Nguyen, Skully & Perera, 2012). Hence, the PTE and CE scores are the best economic indicators for quantifying bank efficiency (Nguyen et al., 2016a).

A bank serves as a financial *intermediary*, using personnel and tangible capital to transform funds or financial resources from depositors to loans or earning assets (Ab-Rahim, 2017; Adjei-Frimpong et al., 2014; Sealey & Lindley, 1977). Hence, deposits or financial resources from the market, personnel and tangible capital are regarded as inputs, whereas financial products and earning assets are considered outputs (Ab-Rahim, 2017; Nguyen et al., 2016b; Phan et al., 2016).

There are two common methods used to quantify bank efficiency: Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA) (Ab-Rahim, 2017; Kočišová, 2014; Phan et al., 2016). SFA is a parametric technique that generates a function of the expense, income or production (Ab-Rahim, 2017). The function defines the relationship among inputs, outputs, and environmental factors (Ab-Rahim, 2017). In contrast, DEA is a non-parametric procedure. DEA does not use a mathematical form but uses a linear program method to estimate bank efficiency (Ab-Rahim, 2017; Adjei-Frimpong et al., 2014; Barth, Lin, Ma, Seade, & Song, 2013). Table 2.3 Panel A summarises the literature on bank efficiency measurement in various countries and Panel B includes Vietnam banks.

Table 2.3 The Methodology Used to Measure Bank Efficiency in Various Countries

Author(s)	Study period	Country(ies)	Measurement Method	EFF Score	Note
PANEL A – Measuring Bank Efficiency in Various Countries					
Ataullah and Le (2006)	1992-1998	India	DEA		
Havrylchyk (2006)	1997-2001	Poland	DEA		
Delis and Papanikolaou (2009	1994-2005	EU countries	DEA		
Kaplan (2010)		Africa	SFA		
Staub, Souza, and Tabak (2010)	2000-2007	Brazil	DEA		
Al-Gasaymeh (2016)	2007-2014	GCC countries	SFA		
Adjei-Frimpong et al. (2014)	2001-2010	Ghana	DEA		
Anwar (2018)	2002-2010	Indonesia	SFA, C92		
PANEL B – *Measuring Bank Efficiency in Emerging Countries Including Vietnam*					
Vu and Tunell (2010)	2000-2006	Vietnam	Bayesian approach	0.87	Compared cost efficiency of state and J.S. banks
Gardener et al. (2011)	1998–2004	Indonesia, Malaysia, Philippine, Thailand, Vietnam	DEA	0.61	
Nguyen et al. (2014)	1995-2011	Vietnam	SFA/ DEA	0.90	
Phan et al. (2016)	2005 - 2012	Bangladesh, India, Indonesia, Malaysia, Philippines, Vietnam	SFA	0.863	
Stewart et al. (2016)	1999-2009	Vietnam	DEA	0.74 0.81	CRS VRS
Nguyen et al. (2016b)	1995-2011	Vietnam, China, India	SFA	0.81	
Nguyen et al. (2016a)	2000-2014	Vietnam	DEA	0.93	
Nguyen and Nghiem (2018)	2000-2014	Vietnam	SFA	0.924	Compared cost efficiency of state and J.S. banks

Note: EFF = efficiency score; J.S. = joint-stock;
Source: Author's summary of past studies

Table 2.3 provides an overview of the studies that used SFA and DEA to quantify bank efficiency in industrialised countries (EU countries and Poland) and emerging countries (African countries, Bangladesh, Brazil, China, Indonesia, India, Ghana, Malaysia, the Philippines and Vietnam). The DEA technique has several advantages over the SFA technique:

- DEA performs well with small sample size compared with the SFA method. This is because the SFA technique is a statistical method that needs a huge dataset to create unbiased estimate coefficients (Adjei-Frimpong et al., 2014; Gardener et al., 2011; Havrylchyk, 2006). In 2017, there were 32 domestic banks (4 commercial state and 28 joint-stock Vietnam banks). Our dataset contains 258 acceptable observations taken from a 10-year period (2008 – 2017). As prior studies indicate, a dataset is considered small if the number of observations ranges from 211 to 392 observations (see for example, Adjei-Frimpong et al., 2014 (211 observations); Stewart et al., 2016 (379 observations); Nguyen et al., 2014 (377 observations); Vu and Tunell, 2010 (392 observations)). All these studies use the DEA method. Because of the size of this study's dataset, DEA was chosen as the most appropriate method.

- The DEA technique focuses on single and individual observations compared with the SFA method that concentrates on all observations in the sample (Barth et al., 2013). Thus, DEA can identify *efficient units* or the best practice units. This technique also specifies *inefficient units* and assists in improving inefficient units (Sherman & Zhu, 2006).

- The SFA technique uses a mathematical formula to measure efficiency. Therefore, the accuracy in measuring efficiency depends on the suitability of the chosen mathematical formula (Barth et al., 2013). In contrast, the DEA technique uses the linear program approach to predict efficiency scores. As a result, the DEA technique does not suffer from the problem of choosing a functional form (Barth et al., 2013).

Studies measuring the efficiency of Vietnam banks include Gardener et al. (2011), Nguyen and Nghiem (2018), Nguyen et al. (2016a, 2016b), Phan et al. (2016) and Vu and Turnell (2010) (Table 2.3, Panel B). They show that the cost-efficient scores (of Vietnam banks) range from 0.61 to 0.93 by different techniques (DEA, SFA, or Bayesian) in different study periods from 1995 to 2014. The cost-efficient values indicate that Vietnam banks can save in their costs from 7% (1-0.93) to 39% (1-0.61) to produce the same output with a given input. Only Nguyen and Nghiem's (2018) and Vu and Turnell's (2010) studies compare the cost-efficient scores between commercial state and joint-stock banks. These authors note that state banks are mostly performing well, i.e., have a higher efficiency score than private banks in transforming the economy (Vu & Turnell, 2010). The authors used the two-sample t-test to verify the differences in efficiency in these bank groups. However, their results show that the difference in cost-efficiency scores between commercial state and joint-stock banks remains statistically *insignificant* (Nguyen & Nghiem, 2018; Vu & Turnell, 2010). In other words, Vietnam banks compete strongly to cut costs and maintain their efficiency levels (Vu & Turnell, 2010). As yet, no study has compared the cost-efficiency scores of Vietnam banks within super-class groups as categorised by the SOM trajectory technique. This study fills this knowledge gap, using the two-sample t-test to compare efficiency scores within the super-class bank groups.

The nonparametric DEA method performs under the constant return to scale (CRS) assumption (Nguyen et al., 2014; Othman, Mohd-Zamil, Rasid, Vakilbashi, & Mokhber, 2016). However, the DEA method has been modified under the variable return to scale (VRS) assumption (Kočišová, 2014; Othman et al., 2016). The VRS assumes that banks work at a less than optimal/efficient scale because of an imperfectly competitive atmosphere, facing economic constraints and a regulatory system (Adjei-Frimpong et al., 2014; Nguyen et al., 2014; Othman et al., 2016). Vu and Turnell (2010) inspected the economies of scale of domestic banks in Vietnam. The authors note that the Vietnam commercial state and joint-stock banks need to cut their costs to achieve an optimal size and be more cost-efficient (Vu & Turnell, 2010). Nguyen et al. (2014) contend that Vietnam banks may include some banks on the inefficient scale. As a result, the CRS assumption is not a suitable option for measuring bank efficiency in Vietnam banks. The VRS assumption is an appropriate choice for this study to avoid the impact of suboptimal scales in several banks.

In conclusion, this study engages the Lerner index to quantify the market power of domestic banks in Vietnam. This is because it can measure individual and continuous values. The two-sample t-test is used to test the differences in market power (Lerner index values) within super-class bank groups. The DEA technique is also used to appraise the efficiency of Vietnam banks, including their PTE and CE scores. The VRS assumption is used with the DEA technique. The t-test is also used to check differences in efficiency (the PTE or CE scores) within the super-class bank groups.

2.4 The Determinants Impacting on Bank Market Power and Efficiency

This section discusses the literature associated with this study's third objective: factors determining bank market power and efficiency.

2.4.1 The Determinants Impacting Bank Market Power

To explore determinants' effects on bank market power, the Lerner index is used as a predicted variable in an estimating function. Table 2.4 shows the prediction methods used to examine the determinants of market power in various countries. Fixed-effect estimation is regularly used (see for example, Abel & le Roux (2017), Adjei-Frimpong et al. (2016) and Simpasa (2013)). All variables are assumed to have strict exogeneity under fixed-effect estimation. Some authors consider the endogeneity problem and use an instrumented estimation approach, such as the generalised method of moment (GMM) or two-stage least squares (2SLS). For example, Erler et al. (2017) identify endogenous variables that could be simultaneously determined in bank market power. These endogenous variables are *inefficiency*, *fee income*, *bank size* and *non-performance loans* (Erler et al., 2017). In Delis and Pagoulato's (2009) study, bank-specific variables such as *capitalisation* and *bank size*, are treated as endogenous determinants. The fixed effect estimator is biased and inconsistent because of endogeneity issues (Adjei-Frimpong et al., 2016; Erler et al., 2017; Wintoki, Linck, & Netter, 2012). In contrast, the well-developed GMM estimator is used to address the endogeneity problem (Adjei-Frimpong et al., 2016; Delis, 2012; Delis & Pagoulatos, 2009; Erler et al., 2017).

Table 2.4 The Determinants that Influence Bank Market Power

Author(s)	Study period	Country(ies)	Techniques use			Note
			OLS	Fixed effect	Instrument estimation	
(1)	(2)	(3)		(4)		(5)
Erler et al., (2017)	2004-2013	The U.S.		✓	SysGMM	Lerner (-1)
Abel and le Roux (2017)	2009-2014	Zambia		✓		
Adjei-Frimpong et al. (2016)	2001-2010	Ghana	✓	✓	SysGMM	Lerner (-1)
Simpasa (2013)	1998-2011	Zambia		✓		
Delis (2012)	1987-2005	84 countries			GMM and 2SLS	Lerner (-1)
Soedarmono et al. (2011)	1999-2007	12 Asian countries		✓	GMM	
Delis and Pagoulatos (2009)	1984-2005	European countries			GMM	Lerner (-1)
Fernández De Guevara and Maudos (2007)	1986-2002	Spain		✓		
Angelin and Cetorelli (2003)	1984-1997	Italy		✓		

Note: Lerner (-1) = one lagged Lerner index variable is positively statistically significant related to bank market power.
Source: Author's summary of past studies

Bank market power tends to persist over time. This is because the banking industry has particular characteristics, such as informational opacity, networking, and relationship lending (Adjei-Frimpong et al., 2016; Delis, 2012; Delis & Pagoulatos, 2009). The positive relationship between the one-year lagged and current market power provides evidence of persistent bank market power (Delis, 2012). This dynamic relationship has been confirmed and reported in various studies (see for example, Adjei-Frimpong et al., 2016; Delis, 2012; Delis & Pagoulatos, 2009; Erler et al., 2017) (see Table 2.4, Column 5).

However, no studies have examined the determinants that affect the market power of Vietnam banks. Understanding the determinants could aid bank executives to select appropriate strategies to boost their market power. Moreover, identifying these determinants could enable policymakers to better understand and implement appropriate policies to maintain a stable, competitive banking environment.

2.4.2. The Determinants Influencing Bank Efficiency

To investigate the determinants that affect bank efficiency, PTE or CE scores are used as predicted variables in regression estimations. Table 2.5, Panel A, lists previous studies that have examined the determinants impacting bank efficiency in various countries; Panel B lists those that focus specifically on Vietnam.

Table 2.5 The Determinants That Influence Bank Efficiency

Author(s)	**Study period**	**Country(ies)**	**Methodology**			**Dynamic**
			TOBIT	Fixed effect	Instrument estimators	
(1)	(2)	(3)		(4)		(5)
Panel A - In Various Countries						
Ataullah and Le (2006)	1992-1998	India	OLS		GMM	EEF(t-1)
Havrylchyk (2006)	1997-2001	Poland	✓			
Denis & Papanikolaou (2009)	1994-2005	EU countries	✓			
Staub et al. (2010)	2000-2007	Brazil	✓		GMM	EEF(t-1)
Adjei-Frimpong et al. (2014)	2001–2010	Ghana		✓	two steps SGMM	EEF(t-1)
Al-Gasaymeh (2016)	2007–2014	GCC countries			GMM	EEF(t-1)
Anwar (2018)	2002-2010	Indonesia	✓			
Panel B - In Vietnam						
Gardener et al. (2011)	1998–2004	including Vietnam	✓			
Stewart et al. (2016)	1999-2009	Vietnam				
Phan et al. (2016)	2005-2012	including Vietnam	✓		2SLS	
Nguyen and Nghiem (2018)	2000-2014	Vietnam			SGMM	

Note: EFF (t-1) – one lagged PTE or CE is positive statistically significant to current bank efficiency.
Source: Author's summary of past studies

As the bank efficiency (PTE or CE) score ranges from zero to one, the TOBIT technique is preferred in various studies (Anwar, 2018; Delis & Papanikolaou, 2009; Gardener et al., 2011; Havrylchyk, 2006; Phan et al., 2016; Staub et al., 2010). All variables are assumed to have strict exogeneity under TOBIT regression. However, some authors consider the endogeneity problem and address the matter in their study. For instance, Phan et al.'s (2016) study considered *concentration* and *competition* as endogenous variables. The authors use the Wald test to check the 'reverse causation' that bank efficiency can impact concentration and competition. The p-value reached an acceptable 1% significance level to reject the null hypothesis of exogenous concentration and competition regressors (Phan et al., 2016). Thus, the TOBIT estimator can generate biased results and instrumented estimators (such as a 2SLS estimator) are used instead (Phan et al., 2016). *Capitalisation* and *non-performance loan* ratio are also treated as endogenous variables so a well-developed GMM estimator is used to investigate the determinants influencing bank cost efficiency (Adjei-Frimpong et al., 2014). The GMM estimation is used in various studies (see for example, Al-Gasaymeh, 2016; Ataullah & Le, 2006; Nguyen & Nghiem, 2018; and Staub et al., 2010). GMM uses the lagged variables as instruments and generates unbiased, consistent results (Ataullah & Le, 2006; Wintoki et al., 2012).

Some studies investigate the dynamic relationships in bank cost efficiency. This is because banks can accumulate knowledge and technological endowment in the previous year that helps them become more efficient the following year (Adjei-Frimpong et al., 2014; Al-Gasaymeh, 2016). Therefore, a dynamic relationship can be realised in a positive relationship between a one-year lagged cost-efficiency and the current cost-efficiency. Several studies have reported a positive nexus between a one-year lagged dependent variable and bank cost efficiency (Al-Gasaymeh, 2016; Ataullah & Le, 2006; Staub et al., 2010) (see Table 2.5, Column 5). In the Vietnam banking context, numerous studies have examined the determinants that impact on bank efficiency (for examples, see Table 2.5, Panel B) (Gardener et al., 2011; Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016). These authors use different techniques, such as TOBIT, 2SLS, and one-step system GMM regression, to analyse factors that impact on bank efficiency (see Panel B in Table 2.5).

In conclusion, previous studies have investigated the determinants that influence bank efficiency of Vietnam banks (Gardener et al., 2011; Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016). However, limited studies have examined the factors that impact on these banks' market power. Therefore, this study will investigate the critical factors that impact both bank market power and efficiency. Identifying such determinants will enable managers to improve both bank market power and efficiency. This study uses dynamic OLS and the exogeneity test to investigate dynamic relationships and endogeneity problems. An instrumented GMM estimator is used to generate consistent, unbiased coefficients.

2.5 The CSM and its Application in Retail Banking

The previous sections have discussed studies on the Vietnam banking industry. This section reviews studies associated with this study's fourth and fifth objectives - the CSM and its application to the retail banking industry.

2.5.1 An Introduction to CSM

CSM is the most successful application of statistical modelling methodology to evaluate consumer creditworthiness; it is a key element in successful credit evaluation by financial institutions (Abdou & Pointon, 2011; Abrahams & Zhang, 2008; Šušterši et al., 2009). The basic goal of the CSM is to build a quantitative model that can assess risks linked to the creditworthiness of applicants based on the information they, and others, provide (financial and demographic information, and reputation in a given period). CSM is also used to monitor the financial health of borrowers' ongoing accounts (Abdou, Tsafack, Ntim, & Baker, 2016; Huang & Scott, 2007).

The fundamental mechanisms of CSM are *scorecard* and *cut-off score* (Abrahams & Zhang, 2008). The scorecard includes sub-classifications, extract scores from statistical models. Each score is a response to the characteristics of the individual loan application (Abrahams & Zhang, 2008). CSM works in as follows. First, information in the loan application is collected from the credit application form. Second, the loan officer assigns a score for each category compared with the scorecard. For example, the category covering demographic information (age, education, and gender) and financial information (disposal income and collateral) will be given different scores. A total score is provided for each category. Third, the total score of each applicant is matched with a cut-off score to decide if the loan is accepted, rejected, or referred for review (Equifax, 2014). Some strategies are used to determine a cut-off score; these include a break-even point to ensure that the bank makes a profit (Abrahams & Zhang, 2008; Equifax, 2014) or an efficient frontier developed by Oliver and Wells (2001). Based on the efficient frontier, banks can select an optimal cut-off score that depends on trade-offs among conflicting business purposes, such as market share, expected profit and expected loss (Oliver & Wells, 2001).

The benefits of using CSM include:

- Credit applications are appraised consistently, uniformly and quickly; risk is quantified as the probability of default (Abdou et al., 2016; Abrahams & Zhang, 2008).
- CSM helps the financial institution to perform better in terms of a measurable system and enhanced management control of existing credit accounts (Schreiner, 2004).
- In addition, financial institutions have a chance to improve their services and avoid future risks by applying CSM (Baklouti & Bouri, 2014).

Some shortcomings associated with the CSM include:

- An applicant being unable to repay their loan because of a natural catastrophe which cannot be calculated by the CSM (Van Gool et al., 2012).
- The most important difficulty in building a CSM relates to the lack of available data. CSM might have decreased accuracy to predict the creditworthiness of borrowers if there are missing data (Dinh & Kleimeier, 2007).
- The CSM method needs constant updating and consistently trained credit officers who use it to predict borrowers' creditworthiness (Van Gool et al., 2012).

In industrialised countries such as Australia, Austria, Sweden and the United States, CSM is used as the primary tool by banks to verify the creditworthiness of debtors because data are available. In contrast, banks in emerging countries may face challenges in designing an effective CSM because of the unavailability of information and payment histories (Dinh & Kleimeier, 2007). There are important features in building a CSM method, such as statistical and computer intelligent techniques and efficiency evaluation criteria (Abdou & Pointon, 2011). These features are discussed in the next section.

2.5.2 CSM Techniques

There is a widespread choice of techniques available to assemble a CSM. These approaches may be divided into two categories: conventional statistical and advanced computing techniques (Abdou & Pointon, 2011). Conventional statistical techniques include linear regression, discriminant regression, decision tree, probit, and logistic regression. The advantages of traditional statistical methods are that they are easy to use and to validate (Dinh & Kleimeier, 2007). Advanced computing intelligence techniques are considered alternative opportunities to improve CSM to help financial institutions have better credit-granting appraisal systems, particularly in emerging economies (Trinkle & Baldwin, 2016). Advanced computing approaches consist of neural networks, support vector machines and fuzzy systems (Abdou & Pointon, 2011; Paliwal & Kumar, 2009). Advanced computing intelligence approaches may assist banks to improve their creditworthiness assessments, therefore, increasing their accuracy in credit decision making (Trinkle & Baldwin, 2016).

Table 2.6 CSM with Statistical and Intelligence Unit Techniques

Author(s)	**Country(ies)**	**Type**	**Conventional statistical approach**			**Artificial Int**	**Note**
			LDA	QDA	LOGIT	ANN	
West (2000)	Australia Germany	Retail	✓		✓	☑	
Limsombunchai et al. (2005)	Thailand	Retail	✓		☑	✓	
West et al. (2005)	Australia Germany	Retail					Ensemble
Lee and Chen (2005)	Taiwan	Retail	✓		✓	✓	Hybrid
Dinh and Kleimeier (2007)	Vietnam	Retail			☑		
Pham and Lensink (2008)	Vietnam	Retail			☑		
Tsai and Chen (2010)	Taiwan	Retail					Hybrid
Kocenda & Vojtek (2011)	Czech	Retail			☑		
Jagric and Jagric (2011)	Slovenian	Retail			✓	☑	
Van Gool et al. (2012)	Bosnia	MFI			☑		
Vasanthi and Raja (2011)	Australia	Retail			☑		
Blanco et al. (2013)	Peruvian	MFI	✓	✓	✓	☑	
Lu et al. (2013)	German	Retail			✓	✓	Hybrid
Baklouti and Bouri (2014)	Tunisian	MFI			☑		
Tsai (2014)	European	Retail			✓	✓	Hybrid & Ensemble
Bekhet and Eletter (2014)	Jordanian	Retail			✓	✓	No model
Tran and Ho (2015)	Vietnam	Retail			☑		
Kiruthika and Dilsha (2015)	India	MFI			✓	✓	No model
Nguyen (2015)	China	Retail			☑		
Abdou et al. (2016)	Cameroon	Retail			✓	☑	

Note: ✓ Model is used by researcher; ☑ Model is highlighted as a superior model; **LDA/QDA** = Linear/quadratic discriminant analysis; **LR** = Logistic regression; **ANN** = artificial neural networks; **MFI** = Microfinance institution.
Source: Author's summary of previous studies

Table 2.6 summarises the techniques researchers have used to build CSM model to evaluate the creditworthiness of loan applicants. Some researchers, such as Baklouti and Bouri (2014), Dinh and Kleimeier (2007), Tran and Ho (2015), Nguyen (2015) and Pham and Lensink (2008), used traditional statistical techniques, such as logit regression, in CSM to categorise debtors into 'default' and 'non-default' groups in retail banks or micro-institutions. Their results reveal that the model helps financial institutions reduce 'default' borrowers. For example, Dinh and Kleimeier (2007) report a decrease in default rates from 3.3% to 2% in a Vietnam retail bank. Baklouti and Bouri (2014) report a decline in default ratio from 39.7% to 21.58% in Tunisian microfinance banks.

Tran and Ho (2015) suggest that the logit regression technique is far superior to the z-score technique in building a CSM in a Vietnam retail bank. The technique is used to calculate a z-score value that is used to predict the likelihood of bankruptcy (Altman, 1968). The authors point out that logit regression is easier and more flexible to cut/add/adjust variables than the z-score to investigate specific variables (or determinants) in credit risks (Tran & Ho, 2015). Chi and Hsu (2012) and Crook et al. (2007) note that logistic estimation is broadly applied in financial and economic practices because the technique is easy to calculate and interpret. West (2000) concludes that logistic regression is the best approach among conventional methods to construct a CSM.

The major shortcoming of the logit regression technique is that the overall accuracy classification power is lower than advanced computing and various hybrid approaches (Abdou & Pointon, 2011; Crook et al., 2007; Trinkle & Baldwin, 2016; Yu et al., 2015). Van Gool et al. (2012) argue that the CSM technique (using logit regression) cannot replace bank officers and suggest it should be used only for semiautomatic processing or as a modification tool.

Another technique that has emerged recently, based on the computer revolution, is the ANN technique (Marque, García, & Sánchez, 2013). Researchers build a CSM for retail banks or micro-institutions using conventional statistical techniques and later benchmark it with the ANN technique (Abdou et al., 2016; Blanco et al., 2013; Jagric & Jagric 2011; Jagric, Kracun & Jagric, 2011; West, 2000). The authors argue that the ANN technique is stronger than traditional statistical techniques, such as logit regression. The advanced computational technique generates more accurate results in classifying borrowers into 'default' and 'non-default' groups. For example, West (2000) finds that ANN improves classification rates by 0.5% to 3%. Abdou et al. (2016) conclude that ANN outperforms logit regression in terms of accuracy. Their results reveal that default rates drop from 15.69% under the ANN technique to 7.68% under the logit regression technique.

By using a more accurate classifier, banks' profits would increase as a result of a decreased default rate. Lee and Chen (2005) highlight that a 1% improvement in an accurate classifier would reduce losses and save a significant amount of money in a large loan portfolio. Misclassification costs, or the cost of credit score errors, have significantly decreased to 13.78% after applying ANN in place of the logit regression technique (Blanco et al., 2013). After reviewing 96 studies over 1990 – 2007 on a variety of subjects, Paliwal and Kumar (2009) find that 58% of the studies that used the ANN approach were more accurate in overall classificatory power than the conventional approach (namely logit regression). In summary, previous studies provide evidence that ANN outperforms conventional statistics such as logit estimation.

Some researchers argue that there is no difference or only a slight enhancement in accuracy in classification power between logistic regression and ANN (Bekhet & Eletter, 2014; Kiruthika & Dilsha, 2015; Limsombunchai et al., 2005). Bekhet and Eletter's (2014) study indicates that both logit regression and neural networks provide practical results but are unable to conclude which technique was better. Limsombunchai et al. (2005) argue that ANN might not necessarily replace logit regression for accuracy in building a better CSM. This is a challenge that researchers may confront: find a better way to build a more accurate risk-assessment system such as CSM (Thomas, 2010). The challenge is that some methodologies slightly outperform other techniques, such as neural networks; but the differences are small and the results generated by the methods might not be discriminant enough to determine which cases should be rejected or accepted (Thomas, 2010). In addition, the major drawback of ANN is that it is a 'black box'. The method itself cannot clearly explain the magnitude of the nexus between the exploratory and dependent variables (Trinkle & Baldwin, 2016). This shortcoming leads to a less developed ANN method (Crook et al., 2007; Thomas, 2010).

Abdou and Pointon (2011) note that there is no unique "best technique" for all conditions unless additional information is available. Generally, the technique chosen is based on the problem's details, data availability and the selected variables. The best technique to build a CSM is one that generates a low, stable, consistent classification rate result (Kiruthika & Dilsha, 2015). Paliwal and Kumar (2009) note that both techniques, ANN and logit regression, are forms of discriminant analysis. These are not only non-competing methodologies but are also techniques that support and overlap each other (Paliwal & Kumar, 2009). Thomas (2010) and West et al. (2005) suggest that researchers may use a combination of techniques. These are ensemble or hybrid methods (Lu et al., 2013; Thomas, 2010; Tsai, 2014; Waidyarathne & Samarasinghe, 2014; West et al., 2005). An ensemble method combines multiple machine learning classifiers (Tsai, 2014). A hybrid method combines with different classifier techniques; for example, a combination of clustering and classification (Tsai & Chen, 2010; Tsai, 2014).

West et al. (2005) argue that although researchers have begun to concentrate on complicated computing, non-linear methodology such as ANNs, the technique often is referred to as a single methodology. The authors demonstrate that ensemble or combination methods may potentially decrease the generalisation error by 3% - 5% compared with a single technique to build a CSM (West et al., 2005). Tsai (2014) confirms that classifier ensembles are better predictors than many single classifiers. The author explains that a classifier ensemble is more accurate because the ensemble technique is integrated from many less accurate classifiers. This combination can lessen the overall errors made by single, different techniques. In other words, the classifier ensemble performs better than a single methodology (Tsai, 2014).

Hybrid methodology shows promising results (Tsai & Chen, 2010). There are various combinations for developing a hybrid technique. Huysmans, Baesens, Vanthienen and Gestel (2006) developed a hybrid credit score classifier by combining a clustering technique, the SOM technique, and a neural network for classification, namely supervised MLP technique. Their results demonstrate that a hybrid model is more accurate than single classifiers (Huysmans et al., 2006). Waidyarathne and Samarasinghe (2014) suggest that the merger of SOM and MLP can create a more robust nonlinear classifier than a single technique. Tsai and Chen's (2010)'s experimental results illustrate that a 'Classification + Classification' hybrid model that combined logistic regression and MLP, provides more accurate results in predicting a borrower's creditworthiness.

Lu et al.'s (2013) study shows that a hybrid model gave promising results. The authors combined logit regression and MLP to build a hybrid method. They use logistic regression to first identify significant and insignificant variables and, then, they use the Wald test to remove insignificant variables. Next, the selected significant variables are inputs for the MLP. Their results reveal that a hybrid model outperforms the logit model and improves the effect of the MLP technique (Lu et al., 2013). The authors' methodology provides a solution for the drawbacks of ANN techniques, such as being a 'black-box' method. In other words, hybridizing logic regression with an MLP technique leads to improved interpretable input variables and decreased noise in the data (Lu et al., 2013).

In conclusion, CSM has recently received much attention from financial institutions looking to better appraise borrowers' creditworthiness (Abdou & Pointon, 2011). However, the techniques used to build a CSM, such as conventional statistics (logit regression) or advanced computing (ANNs) or combinations of the two (a hybrid model) have been re-evaluated by researchers (Lessmann et al., 2015; Thomas, 2010; Yu et al., 2015). This study uses a hybrid technique that integrates logit regression and ANN techniques (including SOM and MLP) to build an efficient CSM. The logistic regression can interpret variables and gives a logical pass for the selected variable (Lu et al., 2013). The SOM technique, with its easy-to-use visualisation capacity, can help to identify potential clusters with non-linear boundaries and reveal dataset trends (Huysmans et al., 2006; Waidyarathne & Samarasinghe, 2014). The MLP technique can deal with a non-linear pattern to increase the accuracy of classificatory power (Abdou et al., 2016). The merging of SOM and MLP can create a robust non-linear classifier (Huysmans et al., 2006; Waidyarathne & Samarasinghe, 2014). The combination of logit regression and

MLP can generate better accuracy in predicting creditworthiness and provide better interpretability for input variables (Lu et al., 2013; Tsai, 2014). For these reasons, this hybrid technique produces a robust CSM.

2.5.3 Other Important Features of CSM

Three important features (evaluation criteria, sample size, and validation method) are associated with the CSM.

2.5.3.1 Performance Evaluation Criteria

Table 2.7, Column 2, presents regular evaluations of CSM performance. A confusion matrix, misclassification cost and area beneath the receiver operating character (ROC) curve (*AUC* in short) are the standards frequently used to appraise CSM implementation (Abdou et al., 2016; Limsombunchai, 2005; Van Gool et al., 2012; West, 2000).

The *confusion matrix* is used to measure the discriminant implementation the CSM. It is measured by the accuracy rate (AC rate), Type I and Type II errors. The AC rate is the ratio of the total correctly categorised as good plus bad cases in dataset (Abdou & Pointon, 2011). Type I error measures the ratio of really good cases predicted badly. Type II error evaluates the ratio of really bad items predicted good. Predicting a poor borrower as a good one costs more than seeing a good creditor as a risk debtor. Thus, the Type II error must be carefully considered, together with the overall accuracy rate (Lu et al., 2013; West, 2000).

The *AUC* is a second criterion commonly used in the literature to evaluate the classification power of a CSM. It refers to the area underneath the ROC curve. The ROC curve is defined as a 2D axis that depicts the percentage of actual bad items predicted as actual bad cases versus the percentage of real bad cases which are projected as good applicants. The ROC curves give a view of the CSM discriminatory power (Baklouti & Bouri, 2014). The AUC varies from 0.5 to 1. However, an AUC equal to at least 0.7 or in the range between 0.75 – 0.9 is used to claim the good discriminatory power of CSM (Chi & Hsu, 2012; Cholongitas et al., 2006; Van Gool et al., 2012).

Misclassification cost is another evaluation criterion used to assess methodologies for building a CSM. This is the only criterion that considers the costs generated by wrongfully selecting “default” or “non-default” borrowers (West, 2000). A technique that generates lower misclassification costs is seen as a better technique.

Gini coefficient is another criterion used to assess the discriminant performance of a CSM. This criterion measures the partition among good and bad creditors (Abdou et al., 2016). Gini < 0.25 implies low discriminant ability; 0.25 ≤ Gini < 0.6 implies an average discriminant ability; and Gini ≥ 0.6 implies a very good discriminant ability.

Table 2.7 Frequently Used Performance Evaluation Criteria, Sample Size, and Validation Methods

	Performance Evaluation Criteria						**Sample Size**	**Validation Method**
	Matrix	Mis. Cost	MSE/RMSE	AUC	K-S	Gini	(No. of Obs)	(Train: Test)
(1)	(2)						(3)	(4)
West (2000)	✓	✓					1 000 690	90 : 10
Lee and Chen (2005)	✓	✓					510	-
Dinh and Kleimeier (2007)	✓						56 037	55 : 45
Abdou et al. (2008)	✓	✓					581	100 : 0
Van Gool et al. (2012)	✓			✓	✓		6 722	70 : 30
Jagric et al. (2011)	✓						1 904	50 : 50
Blanco et al. (2013)	✓	✓		✓			5 500	75 : 25
Baklouti and Bouri (2014)	✓			✓			5 022	67 : 33
Kiruthika and Dilsha (2015)	✓			✓			520	60 : 40
Tran and Ho (2015)	✓						850	82 : 28
Abdou et al. (2016)	✓	✓		✓	✓	✓	599	80 : 20

Note: **Matrix** = Confusion matrix; **MSE/RMSE** = Mean square error, square root of MSE.
Source: Author's summary of previous studies

Table 2.7, Column 2, shows that the confusion matrix, AUC and misclassification cost are the three most popular criteria used to measure CSM performance. Mean square error (MSE)/square root of MSE, Kolmogorov-Smirnoff (K-S) coefficient and the Gini coefficient are less commonly used.

2.5.3.2 Sample Size

Table 2.7, Column 3, lists the sample size used by other researchers. The size ranges from as small as 520 observations (Kiruthika & Dilsha, 2015) to as large as 56,037 observations (Dinh & Kleimeier, 2007). A larger sample size leads to greater CSM accuracy (Abdou & Pointon, 2011). However, the determination of sample size depends primarily on the data provided by financial institutions and the nature of the credit markets. According to Anderson (2007), the minimum number of cases/observations should be 1,500 observations.

2.5.3.3 Validation Method

The validation method is another issue to consider when building a CSM. Table 2.7, Column 4, shows the type of validation technique or in-sample or out-of-sample forecast. The technique typically segregates the original sample into training and test subsamples. The purpose of segregating the original sample is to avoid upward bias (Limsombunchai et al., 2005). In some cases, the model predicts very well for in-sample but its performance may be quite poor on out-of-sample cases. Therefore, a test on a subsample may help assess the future applicability of the technique. Column 4 also shows several kinds of segregating sample. A ratio of training includes test subsample ranges from 50:50 (Jagric et al., 2011), to 55:45 (Dinh & Kleimeier, 2007), up to 80: 20 (Limsombunchai et al., 2005) or no segregation (Abdou et al., 2008).

In conclusion, this study will investigate evaluation criteria namely, confusion matrix, AUC, misclassification cost, K-S coefficient and the Gini coefficient, to appraise the classification power of the hybrid technique. These evaluating criteria show improved accuracy of the hybrid technique compared with conventional statistical techniques such as logit regression.

2.6 The Risk Based Pricing Strategy

This section discusses studies linked to the study's sixth and seventh objectives: the Risk Based Pricing Strategy (RBP) strategy and its implementation on the industrialised banking industry.

2.6.1 Introduction of the RBP Strategy

In the early 1990s, banks offered only constant interest rates for loans and banned most high probability of default borrowers (Edelberg, 2003). Changing underwriting standards inspired more US banks to provide credit during the 1990s (Getter, 2006). Further, with developments in automated CSM and a reduction in data storage costs, banks started to predict borrowers' creditworthiness to approve different lending interest rates for loans (Edelberg, 2003, 2006).

The primary purpose of credit scoring has changed dramatically (Scroggins et al., 2004). In the beginning, CSM was used to categorise borrowers and determine whether to approve a loan (Scroggins et al., 2004). Recently, credit scoring has been improved and can now assist banks to price loans (Thomas, 2010). A common theme in the literature is that the interest rate should reflect a borrower's individual default risk (Magri, 2018). Each potential credit customer should be offered an interest rate that matches his/her specific credit risk (Magri, 2018). Higher frequency of default borrowers might be given a higher lending interest rate to reimburse for the possibility of defaulting, and the highest credit risk may result in a rejection of credit (Getter, 2006). This approach is commonly known as the RBP strategy (Magri, 2018).

By implementing CSM for pricing loans, banks can group their likely customers based on their creditworthiness and enhance more effective distribution of finance (Magri & Pico, 2011; Scroggins et al., 2004). For this purpose, pricing credit is more important than discriminating or rejecting credit applicants (Getter, 2006). Using historical payment data, banks can negate asymmetric information and allow for price credit risks (Getter, 2006; Scroggins et al., 2004). The next section explains the major benefits for banks of implementing the RBP.

2.6.2 The Benefits of Implementation of the RBP strategy

First, Edelberg (2006) suggests that if the banks refuse high-risk quality customers, these borrowers might be removed from the credit market. With the RBP, the banks can charge high-risk customers a higher interest rate to cover an unexpected loss (Edelberg, 2006; Magri & Pico, 2011). Banks may also find that they can extend market share to high-risk borrowers using credit scoring without increasing default rates (Scroggins et al., 2004). In other words, this strategy provides a more effective distribution of financial resources (Magri & Pico, 2011). Consequently, this strategy may have positive effects on the bank's profitability and market share, as well as their stockholders' wealth (Magri & Pico, 2011; Scroggins et al., 2004).

Second, the change from constant to various loan prices leads to an increase in a bank's competitiveness. The reason could be that banks that offer loans to high risk borrowers may gain a competitive advantage over those that do not (Edelberg, 2006; Getter, 2006). As a result, banks can develop their position in the credit market based on how large their market share is (or how large they want it to be) or as how much they are willing to lose (Oliver & Wells, 2001).

Third, from a borrowers' perspective, fewer customers are refused credit because banks use the RBP strategy (Getter, 2006). Banks categorise their customers into dissimilar groups; riskier groups have to pay a higher loan price as a penalty for poor payment performance instead of outright rejection of their applications (Getter, 2006). High-risk customers are thus given an opportunity to participate in the credit market so long as if they are willing to accept a greater loan price to compensate for additional risk (Getter, 2006). In other words, making credit available to riskier consumers through different strategies can enhance the functioning of the credit market (Getter, 2006).

2.6.3 The Implementation of the RBP Strategy in Industrialised Countries

This section analyses the practical implementation of the RBP strategy in industrialised countries (see Table 2.8).

American banks have used the RBP strategy since the mid-1990s (Edelberg, 2006). American banks began to categorise their customers into different groups and determine loan pricing and credit terms depending on individual customer's creditworthiness (Scroggins et al., 2004). In America, differences between the loan prices of high and low-risk borrowers became significantly larger from 1989 to 1998 (Edelberg, 2003). For example, an increase of 0.01 in a loan's default risk, the lending interest rate increases by over 100% for a first mortgage from 0.16 to 0.38 points. In addition, the charges increase up to five times for a second mortgage and double for car loans. Edelberg (2003) shows that American banks increasingly used RBP after the mid-1990s, but only for secured loans.

Table 2.8 The RBP Strategy Applied in Banks of Industrialised Countries

Author(s)	Country	RBS	Since	Evident findings
Edelberg (2003, 2006)	America	☑	Mid-1990s	An increase of 1% in default risk, interest rate charged increases nearly double for first mortgage (0.16 to 0.38 basis points).
Scroggins et al. (2004)	America	☑	1993, 2004	Credit risk and potential profitability are two critical factors. The banks use these critical factors to separate their customers into groups and offer different interest rates that reflect default risks.
Getter (2006)	America	☑	2006	Borrowers with riskier credit profiles will be charged a higher interest rate. Credit prices are influenced by customer behaviour.
Magri and Pico (2011)	Italy	☑	2005 2007	The interest rate will increase 21 basis point when 1%-point increases the probability of bad borrowers in mortgage loans granted from 2000 to 2007.
Magri (2018)	Italy	☑	2008	Three critical factors (net wealth, education and searching for better terms) influence both interest rate charged and probability of default.

Source: Author's summary of previous studies

Scroggins et al. (2004) administered a survey questionnaire to the Chief Executive Officers of 1000 American banks. The questionnaire asked whether their bank separated customers into groups. The banks were also asked about the factors they used to divide their potential customers (Scroggins et al., 2004). The survey generated 156 useable responses. Only 150 banks confirmed the use of factors to split their customers into groups. *Credit risk* and *potential profitability* were two critical factors for American banks separating their customers (Scroggins et al., 2004).

Getter (2006) attempted to re-examine whether US borrowers' interest rates reflect the accuracy of their credit risk. The author indicates that borrowers face inaccessible credit because the banks do not have information to appraise credit risk correctly; banks are also apprehensive about offering abnormally high rates to risky loans (Getter, 2006). Some critical factors are impact interest rates significantly (Getter, 2006). Borrowers with 30-day or 60-day delinquencies are charged higher interest rates. Getter (2006) concludes that loan prices are shaped by customers' behaviour. Finally, Getter shows that American banks implementing RBP strategies lead to high-risk borrowers (low and smaller group) increasing over the past 20 years.

Magri and Pico (2011) investigated credit risk and mortgage rate pricing in Italy. The authors indicate that credit risk of mortgage loans can be estimated more accurately than consumer loans because data are available (Magri & Pico, 2011). Italian banks have implemented CSM since 2003. However, the CSM is used to measure a borrower's creditworthiness for acceptance or rejection more than pricing a loan (Magri & Pico, 2011). Approximately 5% of household mortgages were defaulted in 2007 in Italy (Magri & Pico, 2011). The authors also investigated the critical determinants that impact the probability of mortgage default and find that *higher education* and *household income* reduce default rates. Conversely, *bad health, unemployment* and *immigration status* (if they were a migrant) increase the likelihood of defaulting on mortgage repayments. Magri and Pico (2011) also examined how interest rates of mortgages are linked to household credit risk. The authors note that an increase in mortgage credit risk leads to a higher interest rate. Principally, between 2000 and 2007, for an increase of 1% in a loan's default risk, the mortgage lending interest rate increases 21 basis points (Magri & Pico, 2011).

Magri (2018) investigated the RBP in relation to consumer loans (car loans and furniture loans) by Italian banks. First, the author examined critical factors that influence consumer loan pricing. The results show that the richest borrowers in the first quartile of net wealth are paid 1.2 basis points more than others. The second net wealth quartile borrowers are paid 0.8 basis points more. Having a college degree reduced a loan price by 0.5 basis points (Magri, 2018). Younger borrowers (less than 44 years) are offered a lower interest rate by 0.6-0.9 basis points than older borrowers and a larger loan is charged a lower interest rate. The last significant factor is a 'search' variable. Borrowers who searched for better terms paid 0.4 basis points lower than those who did not. Secondly, Magri (2018) appraised whether the critical factors impact on the probability of borrowers defaulting. The author shows that borrowers with lower net wealth have a higher credit risk (5 basis points) than those with the highest net wealth. Conversely, more educated borrowers have a 4.8 basis points lower probability of defaulting than others (Magri, 2018). Three critical factors (net wealth, education and searching for better terms) have been found to influence both interest charged and the probability of borrowers' defaulting. Consequently, Magri (2018) and Magri and Pico (2011) reveal Italian banks' increasingly rely on CSM for loan pricing since the 2008 GFC.

There are few studies that address the role of CSM in loan pricing by Vietnam banks. Therefore, this study investigates whether the RBP has been used in Vietnam banks. First, this study investigates the critical factors that influence lending interest rates. Next, the study examines whether the critical interest rate factors are based on the credit risk factors of default borrowers.

2.7 Chapter Conclusion

This is the first study to explore the financial status of Vietnam banks using the SOM trajectory technique. It covers 10 years from 2008 to 2017. The findings are expected to support bank executives observe their financial situation over time by analysing financial trajectory patterns. This also helps them understand their financial condition and potential risks which would allow them to take corrective action (Chen et al., 2013). In addition, the study categorises Vietnam banks into super-class groups whereas most previous studies classify Vietnam banks into commercial state and joint-stock banks (Nguyen & Nghiem, 2018; Vu & Turnell, 2010). Hence, this study provides a new approach for categorising Vietnam banks that can be used to further analyse market power and efficiency among these super-class bank groups.

This study uses the Lerner index to quantify the market power of Vietnam banks. This is because it can provide individual and continuous values. The two-sample t-test will be used to check the differences in market power (the Lerner index) in the super-class bank groups. If the p-value of the two-sample t-test is statistically significant, it implies that the SOM trajectory technique has identified differences in market power within the super-class bank groups. The DEA technique is used to quantify the efficiency of Vietnam banks, represented by the PTE and CE scores. The VRS assumption is also used with the DEA technique. The two-sample t-test will be used to test differences in efficiency (PTE or CE) scores within the super-class bank groups. Previous studies have investigated the determinants that influence bank efficiency in Vietnam banks (Gardener et al., 2011; Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016). However, few studies examine the factors that impact on the banks that you are exploring with an alternative methodology that one hopes will be useful to banks' market power. Therefore, this study will investigate the critical determinants that impact on both bank power and efficiency. Studying such determinants will enable bank managers to improve both bank market power and efficiency. This study uses dynamic OLS and the exogeneity test to assess the dynamic nexus and endogeneity problems. An instrumented GMM estimator is used to generate consistent, unbiased coefficients.

CSM has recently received much attention from financial institutions that use it to appraise borrowers' creditworthiness (Abdou & Pointon, 2011). However, the techniques used to build a CSM, such as conventional statistics (logit regression) or advanced computing (ANN) or a combination (hybrid or ensemble) have been re-evaluated recently by researchers (Lessmann et al., 2015; Thomas, 2010; Yu et al., 2015). This study uses a hybrid technique that combines logit regression and ANN techniques (the SOM and MLP techniques) to build an efficient CSM. Evaluating criteria (confusion matrix, AUC ratio, misclassification cost, K-S coefficient and Gini coefficient) are used to appraise the classification power of the hybrid technique. These evaluating criteria improve the accuracy of the hybrid technique compared with conventional statistical techniques such as logit regression. Finally, this study investigates whether RBP has been used in the Vietnam retail banking sector because limited studies provide evidence on the use of the CSM for loan pricing. The next chapter details the methodologies use in this study.

Chapter 3

Research Methodology

3.1 Introduction

This chapter discusses the study's research methodology. Section 3.2 provides an overview of the SOM trajectory technique used to build the financial trajectory of Vietnam banks. Section 3.3 reviews methodologies that measure the banks' market power and efficiency. A two-sample t-test is used to compare banks' market power and efficiency between super-class bank groups, as categorised by the SOM trajectory technique. Section 3.4 provides an overview of the techniques used to investigate the determinants of banks' market power and efficiency.

Sections 3.5 to 3.8 explain the methodology used to examine the Vietnam retail banking sector. Section 3.5 presents the logit regression method that is used to investigate factors that determine the probability of borrowers defaulting on their repayments. Section 3.6 examines the proposed hybrid technique that integrates logit regression and an ANN to boost the classificatory power of CSM. This section also evaluates the classificatory power of the hybrid technique based on evaluation criteria: the confusion matrix, the AUC ratio, the K-S statistic ratio, the GINI coefficient and the misclassification cost. Section 3.7 discusses the critical factors that have impact the interest rate charged by banks. Section 3.8 examines whether Vietnam's domestic banks rank customers according to their risk levels when they determine their loan prices. Section 3.9 summarises the chapter.

3.2 Objective 1: Investigate Vietnam Banks' Financial Trajectory Using the SOM Technique

3.2.1 Data from Vietnam Banks

Data for the study were gathered from 32 domestic banks operating in Vietnam. Appendix Table A1 lists the Vietnam banks in operation during the study period (2008 to 2017). The 10-year timeframe was chosen for the following reasons:

- This study investigates the influence on Vietnam banks of the "Restructuring Financial Institutions 2011-2015" program (Decision no.254/QD-TTg), in particular, M&A deals. Therefore data were collected from the beginning of the restructuring program in 2011. However, panel data, which covers fewer than 10 years, may generate biased results (Nadeem, Gan & Nguyen, 2017; Wintoki et al., 2012). This bias happens because statistical conclusions cannot be realised if the data study period is too short (Vu & Turnell, 2010). To meet the 10-year timeframe, bank data need to be collected from 2011 to 2021, which is impossible, given that the study was conducted in 2019. At the time of study, data were available only from 2011 to 2017. Therefore, bank data need to be gathered from 2008 to 2017 to cover a 10-year study period. GPD growth rates range from 5.25% to 6.81% over 2008 to 2017. The inflation rate dropped from 23.1% in 2008 to 3.5% in 2017. These macroeconomic factors indicate that the Vietnam banking industry was relatively stable during the study period, especially after the 2008 GFC.

- Two banks in our sample dataset of 32 banks were established in Vietnam in 2008. The Tien Phong Bank was established in March 2008 and the LienViet Postbank opened in May 2008. These two banks are currently operating in Vietnam.

Bank financial data were obtained from the *Bloomberg* database and bank websites for the study period. Macroeconomic indicators, such as GDP and the inflation rate, were sourced from the *World Bank* database. Five of the listed banks did not provide financial data during the period. These were: PVcom bank, Seabank, Bao Viet bank, Co-op Bank, and Vietcapital bank. In addition, several banks have missing data. SCB did not provide data in 2011. The Bac A bank has no financial data from 2008 to 2010. Vietbank has data for only 2016 and 2017. Considering these data deficiencies, there are 27 banks with unbalanced panel data (258 observations) over the study period. The 27 banks represent approximately 85% of the Vietnam banking industry. Appendix Table A2 summarises the 27 banks' financial data for 2008 to 2017.

3.2.2 Unsupervised Learning: The SOM Technique

The unsupervised SOM technique is a feed-forward ANN that includes input and output layers (Chen et al., 2013; Samarasinghe, 2006). The output neural layer is usually a low-dimension grid that is, a one- or two-dimensional grid (see Figure 3.1). Each unit of the input layer is connected to all neurons in the output layer by weight. The training process for n-input neurons and m-output neurons is to (Chen, 2012):

1. randomly initialise the weight vectors $\mathbf{w}_i$ for all neurons i = 1, ..., m;

2. choose an input vector $\mathbf{x}$: $\mathbf{x} = [x_1, x_2, x_3, \ldots, x_n]$; and

3. compare $\mathbf{x}$ with weights $\mathbf{w}_i$ in each neuron i to decide the winner. The winning vector is the one nearest to the input vector or which has the smallest Euclidean distance (see Equation 3.1).

$$||w_k - x|| = \min_i ||w_i - x|| \text{ (Euclidean distance)} \qquad [3.1]$$

4. Update the winning node so that the winner becomes closer to $\mathbf{x}$, together with neighbours around the winner. The SOM is called a topology preserving map (Huysmans et al., 2006). Weight vectors of the neighbourhood ('neurons i') of the winner ('neuron k') are updated as follows (see Equation 3.2):

$$w_i = w_i + \mu \cdot \varphi(i, k) \cdot (x - w_i) \qquad [3.2]$$

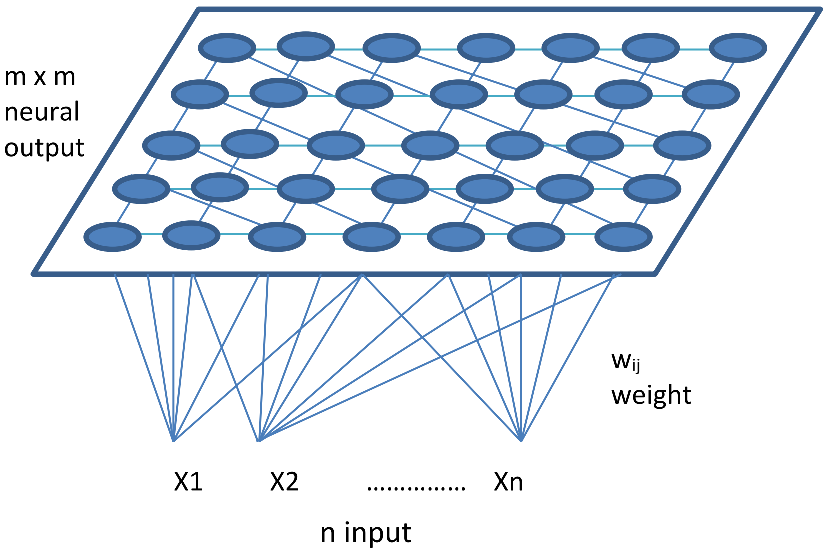
Where:

μ = learning rate:

φ (i, k) = neighbourhood function, a Gaussian function is often used.

5. Repeat steps (2) to (4) until the map has converged (that is, the weights of neurons do not change) or a pre-defined number of epoch trainings has been reached.

Figure 3.1 SOM Architecture with n Neurons Inputs and m × m Neuron Outputs



Source: Adapted from Chen (2012)

After training, the SOM topology map is compressed and visualised into a 2-D SOM map that retains only the potential cluster and trend within the dataset (Waidyarathne & Samarasinghe, 2014). In other words, the SOM helps to outline the nature of the input pattern dataset (Tsai, 2014).

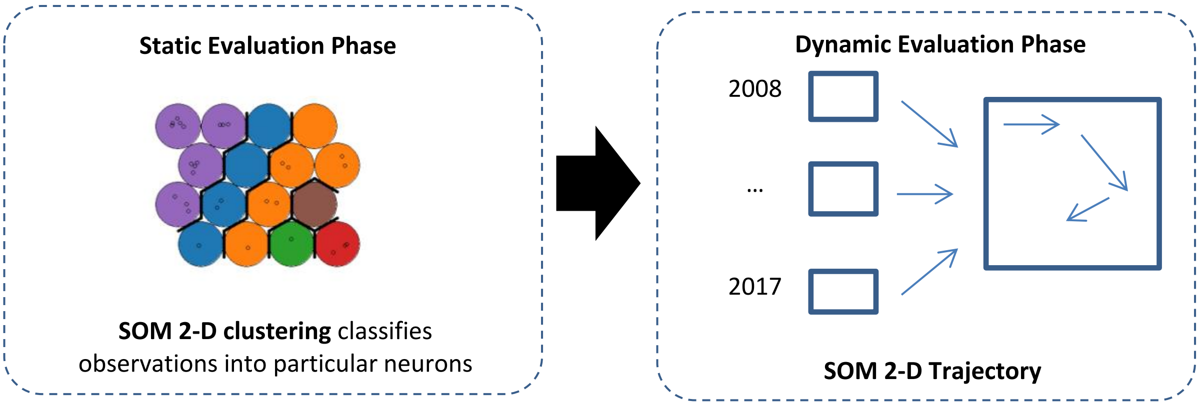
3.2.3 Developing Financial Trajectories Using the SOM Technique

This study draws on the methodology from Chen's (2012) study to answer the first research objective – to investigate Vietnam banks' financial trajectory patterns. Building the trajectory includes static and dynamic phases (see Figure 3.2) as follows:

In the *static* phase, Vietnam banks' financial statement data are screened using the SOM technique for every year from 2008 to 2017. After screening, each bank is located within specific neurons in the 2D SOM map. Each year will have a different 2D SOM map to represent a bank's location. Cluster analysis is applied to optimise groups of neurons.

In the *dynamic* phase, ten 2D SOM maps from every year are overlapped into one 2D map. Immediately, bank locations reveal changes over time. For example, bank A is in group I in year 1, then moves to group II in year 2, and group III in year 3. Therefore, bank A's trajectory is represented by connecting the bank's location from year 1 to year 3. By observing the trajectory of every bank in the banking industry, we can recognise important trajectory patterns as well as categorise the banks into super-class groups (Chen, 2012; Du Jardin & Séverin, 2011).

Figure 3.2 Building the Financial Trajectories of Vietnam Banks using the SOM Technique



Source: Author's computation

This study expects to find common patterns in the financial trajectories. That outcome could help Vietnam bank executives identify any risky financial situations and take corrective action (Chen, 2012; Chen et al., 2013). Vietnam banks can be categorised into super-class groups (Du Jardin & Séverin, 2011; 2012). These outcomes can be used for further analysis in terms of benchmarking a bank's market power and efficiency among these super-class bank groups. The next section explains the methodologies that measure and compare bank market power and efficiency among the super-class groups.

3.3 Objective 2: To Measure and Compare Different Groups of Vietnam Banks' Market Power and Efficiency as Categorised by The SOM Technique.

3.3.1 Measuring Vietnam Bank Market Power

This study uses the non-structural Lerner index to quantify the market power of Vietnam domestic banks. The Lerner index reveals the changes between the price charged (or interest rate and fees) by the banks and their marginal cost of total assets (MC). A higher Lerner index indicates greater market power. This relationship can be expressed mathematically, using equation (3.3) (Ab-Rahim, 2017; Adjei-Frimpong et al., 2016; Demirguc-Kunt & Martínez-Pería, 2010):

$$Lerner_{it} = \frac{P_{it} - MC_{it}}{P_{it}} \qquad [3.3]$$

Where:

P_{it} = Price output of the i^{th} bank at time t, which is designed using total revenue (non-interest income plus interest income) divided by total asset value; and

MC_{it} = Marginal cost of i^{th} bank at time t that is determined using the derivative of the translog-cost function.

The Lerner index ranges from negative infinity to 1. If a bank's Lerner index value is close to 1, it shows the higher power of that bank. When the Lerner index has a value of 1, the bank is a monopolist. In contrast, if the Lerner score is close to zero, it indicates a greater competition. If the value is 0, this reflects perfect competition. All banks cannot set their prices above marginal costs. The degree of bank competitiveness is calculated using the value of one minus the Lerner score (Nguyen et al., 2016b). When the Lerner index is negative, it indicates that a bank has reduced its prices below cost because of external influences including economic crises (Abel & le Roux, 2017; Coccorese, 2009).

The MC of the ith bank at time t is determined using the derivative function of the translog cost formula, which is expressed as follows (Demirguc-Kunt & Martínez-Pería, 2010):

$$\begin{aligned} Ln(C_{it}) = {} & a_0 + b_0 Ln(Q_{it}) + b_1 0.5[Ln(Q_{it})]^2 + a_1 Ln(W_{1it}) + a_2 Ln(W_{2it}) + a_3 Ln(W_{3it}) + \\ & b_2 0.5 Ln(Q_{it}) \times Ln(W_{1it}) + b_3 0.5 Ln(Q_{it}) \times Ln(W_{2it}) + b_4 0.5 Ln(Q_{it}) \times Ln(W_{3it}) + \\ & a_4 Ln(W_{1it}) \times Ln(W_{2it}) + a_5 Ln(W_{1it}) \times Ln(W_{3it}) + a_6 Ln(W_{2it}) \times Ln(W_{3it}) + \\ & a_7 0.5[Ln(W_{1it})]^2 + a_8 0.5[Ln(W_{2it})]^2 + a_9 0.5[Ln(W_{3it})]^2 + \\ & d_1 Trend + d_2 Trend^2 + d_3 Trend \times Ln(Q_{it}) + d_4 Trend \times Ln(W_{1it}) \\ & d_5 Trend \times Ln(W_{2it}) + d_6 Trend \times Ln(W_{3it}) + u_{it} \end{aligned} \quad [3.4]$$

Where:

C_{it} = the total cost of the i^{th} bank at time t computed using total operating expenses plus interest expenses;

Q_{it} = the total assets of the i^{th} bank at time t;

W_{1it} = the unit price of staff costs of the i^{th} bank at time t computed using staff costs divided by total assets;

W_{2it} = the unit price of funding of the i^{th} bank at time t calculated using interest expenses divided by deposits;

W_{3it} = the unit price of capitalisation of the i^{th} bank at time t calculated using working expenses, deducting staff costs and dividing by fixed assets;

$Trend$ = the time trend which captures technical change; and

$a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, b_1, b_2, b_3, b_4, d_1, d_2, d_3, d_4, d_5, and\ d_6$ = the coefficients to be estimated.

The banks were not randomly chosen. Therefore, a fixed-effect estimator is used to predict the coefficients of the translog-cost formula (equation 3.4) (Abel & le Roux, 2017; Adjei-Frimpong et al., 2016; Demirguc-Kunt & Martínez-Pería, 2010; Nguyen & Nghiem, 2018).

The MC_{it} of the i^{th} bank (at time t) is estimated using the first derivative of equation (3.4) as follows:

$$MC_{it} = \frac{TC_{it}}{\mathrm{Q_{it}}}[a_0 + b_1 Ln(Q_{it}) + b_2 0.5Ln(w_{1it}) + b_3 0.5Ln(w_{2it}) + b_4 0.5Ln(w_{3it}) + d_3 Trend] \quad [3.5]$$

The coefficients $a_0, b_1, b_2, b_{3,} b_4, d_3$ are estimated from equation (3.4) and plugged into equation (3.5) to compute (MC_{it}) of the i^{th} bank at time t.

After the market power of each Vietnam bank (represented by the Lerner index score) is determined at time t using equation (3.3), the two-sample t-test is employed to compare the different market powers of the super-class bank groups. If the p-value indicates significance at all significance levels (between 1% and 5%), this indicates a unique situation within Vietnam banks. There are two groups of bank with different levels of market power that exist separately in the Vietnam banking industry. This outcome also provides evidence for the existence of a group of strong banks because of bank size concentration in the post M&A period. These strong banks tend to be monopolists and damage the perfectly competitive environment of the Vietnam banking industry (Hoang et al., 2016).

3.3.2 Measuring Vietnam Bank Efficiency

This study uses the DEA technique with the VRS assumption to compute bank efficiency. PTE and CE scores are used to characterise bank efficiency. The PTE and CE scores use linear programs to generate a number of frontier observations. The PTE value can be computed mathematically as follows (Ab-Rahim, 2017; Kočišová, 2014; Nguyen et al., 2016a; Othman et al., 2016):

$$\begin{aligned}
&min && \theta_q && \\
&subject\ to && \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_q x_{iq} && i = 1,2,\cdots m \qquad [3.6]\\
& && \sum_{j=1}^{n} y_{rj}\lambda_j \le y_{rq} && r = 1,2,\cdots s\\
& && \sum_{j=1}^{n} \lambda_j = 1 && \\
& && \lambda_j \ge 0 && j = 1,2,\cdots n
\end{aligned}$$

Where:

θ_q = the efficient ratio (PTE score) of the decision-making unit $(DMU)_q$. If θ_q = 1, the bank is classified as pure technically efficient. If θ < 1, the bank is classified as inefficient.

y_{rq} = the amount of the r^{th} output ($r = 1,2,\cdots s$) of DMU_q;

x_{iq} = the amount of the i^{th} input $(i = 1,2,\cdots m)$ of DMU_q;

y_{rj} = the amount of the r^{th} output ($r = 1,2,\cdots s$) of $DMU_j (j = 1,2,\cdots n)$;

x_{ij} = the amount of the i^{th} input $(i = 1,2,\cdots m)$ of DMU_j $(j = 1,2,\cdots n)$; and

λ_j = the weight of DMU_j $(j = 1,2,\cdots n)$.

The cost minimisation DEA model is solved to calculate the CE score (Ab-Rahim, 2017; Kočišová, 2014):

$$
\begin{array}{lll}
min & \sum_{i=1}^{m} w_{iq} x_{iq}^{*} & \\
subject\ to & \sum_{j=1}^{n} x_{ij}\lambda_j \leq x_{iq}^{*} & i = 1,2,\cdots m \\
 & \sum_{j=1}^{n} y_{rj}\lambda_j \leq y_{rq} & r = 1,2,\cdots s \\
 & \sum_{j=1}^{n} \lambda_j = 1 & \\
 & \lambda_j \geq 0 & j = 1,2,\cdots n
\end{array} \qquad [3.7]
$$

Where:

w_{iq} = an input price of DMU_q; and

x_{iq}^{*} = the frontier or cost-minimising unit of DMU_q , given input prices w_{iq} and output y_{rq}; x_{iq}^{*} is calculated by solving the linear program.

Following the minimum cost, the cost efficiency of $Bank_q(CE_q)$ is computed as a proportion being the minimum cost ($w_{iq}x_{iq}^{*}$) divided by the actual cost($w_{iq}x_{iq}$) in the form:

$$
CE = \frac{w_{iq}x_{iq}^{*}}{w_{iq}x_{iq}} \qquad [3.8]
$$

Therefore, cost efficiency, as outlined in equation (3.8), assesses the difference between a bank's actual operating cost (in relation to the best/or minimum cost in the market), given the input price and outputs (Kočišová, 2014). CE has a value between zero and one. If CE = 1 for DMU_q , it means that $Bank_q$ is cost-efficient. If CE < 1 forDMU_q, the $Bank_q$ is inefficient. The value $(1 - CE)$ for the $Bank_q$ is also important. This number exhibits the cost that the $Bank_q$ can collect in producing the same output with a given input (Nguyen et al., 2016a).

After domestic banks' efficiencys (illustrated by PTE and CE scores) are calculated using equations (3.6) and (3.8) for each bank at time t, the two-sample t-test is used to compare the different PTE and CE scores within the super-class groups of Vietnam banks. If the p-value indicates significance at all significance levels (between 1% and 5%), the result indicates that there are variations in the efficiency levels of Vietnam banks.

3.4 Objective 3: To Investigate the Determinants that Impact Vietnam Bank Market Power and Efficiency

Using two models, this study explores the determinants that impact the market power and efficiency of Vietnam banks: an empirical model of a bank's market power (equation 3.9) and an empirical model of a bank's cost efficiency (equation 3.10). The Lerner index and CE scores calculated in section 3.3, are used as predicted variables in the two empirical models (equations 3.9 and 3.10). Two groups of exploratory variables (or determinants) are used. They include bank-specific characteristics and macroeconomic indicators. The relationships can be expressed mathematically as follows:

The empirical model of a bank's market power is given as:

$$LER = f(Bank\ specific\ characteristics, Macroeconomic\ indicators) \qquad [3.9]$$

The empirical model of a bank's cost efficiency is given as:

$$CE = f(Bank\ specific\ characteristics, Macroeconomic\ indicators) \qquad [3.10]$$

Where:
LER = Lerner index;
CE = Cost efficiency score;
the bank specific characteristics = total assets, equity ratio and non-performance ratio of banks;
the macroeconomic indicators = GDP growth rate and the inflation rate; and
f = a regression function.

3.4.1 Pooled OLS Regression Estimates For Panel Data

The dataset is unbalanced panel data (258 observations) with 27 banks over 2008 – 2017. It is necessary to check data stationarity to avoid spurious regression (Gujarati & Porter, 2009). If the dataset is stationary (or no unit root), a pooled OLS estimator can be used to estimate equations (3.9) and (3.10). The results obtained from the pooled OLS estimator are used to check robustness. Under the pooled OLS, all variables are assumed to be strictly exogeneous. The pooled OLS regression can be expressed as follows:

$$y_{it} = \beta_0 + X'_{it}\beta + \mu_{it} \qquad [3.11]$$

Where:
i = characterises the particular bank and *t* denotes year;
y_{it}= dependent variables, including the CE and LER scores;
X_{it} = independent variables (bank characteristics and macroeconomic indicators); and
μ_{it}= the error term.

The coefficient β can be computed as follows (Wooldridge, 2012):

$$\beta_{OLS} = \left(\sum_{i,t}^{n,T} X'_{it}X_{it}\right)^{-1} \sum_{i,t}^{n,T} X'_{it}y_{it} \qquad [3.12]$$

The major problem of the OLS estimator is that it cannot differentiate between various banks or time. In other words, the OLS estimator does not illustrate whether the response of market power or efficiency (CE and LER scores) to the exploratory variables (characteristics and macroeconomic indicators) is the same or different banks over time. As a consequence, the pooled OLS may suffer from the heterogeneity because this method does not distinguish various banks at different times (Gujarati & Porter, 2009). The fixed effect estimator is used to overcome this problem because it allows individual banks to have their own intercepts (Gujarati & Porter, 2009).

3.4.2 The Fixed-Effect Estimator Predictions for Panel Data

The fixed-effect estimator presumes that unobserved effects are constant within individual banks but change across banks (assuming heterogeneity across banks) at a given in time (Gujarati & Porter, 2009). Unlike the pooled OLS estimator, fixed effect regression controls for unobserved heterogeneity. Fixed effect regression is used to assess the determinants that influence the market power and efficiency of Vietnam banks. This can be expressed mathematically as follows:

$$y_{it} = \beta_0 + X'_{it}\beta + \eta_i + \mu_{it} \qquad [3.13]$$

Where:

I = characterises the particular bank and *t* denotes year;

y_{it}= the dependent variables, including the CE and LER scores;

X_{it} = the independent variables (bank characteristics and macroeconomic indicators);

η_i = unobserved individual effects; and

μ_{it}= the error term.

To remove an unobserved individual effect η_i, the fixed effect estimator (equation 3.13) first subtracts the time mean within each variable. Then equation (3.13) is rewritten as follows (Wooldridge, 2012):

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta' + (\mu_{it} - \bar{\mu}_i)$$

Where: [3.14]

$$\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it} \ and \ \bar{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} \mu_{it}$$

$$\widehat{X_{it}} = (X_{it} - \bar{X}_i) and \ \widehat{y_{it}} = (y_{it} - \bar{y}_i)$$

Therefore, the OLS method is chosen to estimate β, as follows (Wooldridge, 2012):

$$\widehat{\beta_{FE}} = \left(\sum_{i,t}^{T} \hat{X}'_{it}\hat{X}_{it}\right)^{-1} \sum_{i,t}^{T} \hat{X}'_{it}\hat{y}_{it} \qquad [3.15]$$

3.4.3 A Dynamic Estimator for Panel Data

Previous studies have reported that the market power and efficiency of banks tends to remain over time (Adjei-Frimpong et al., 2016; Al-Gasaymeh, 2016; Anwar, 2018; Delis, 2012; Erler et al., 2017). This relationship can be expressed with one-year lagged market power (or efficiency) of banks, which appears as a regressor and has a positive significant relationship with the current market power of banks (or current bank efficiency). To capture the dynamic nature of the relationship between market power and efficiency of bank, a one-year lagged market power and efficiency ($y_{i,t-1}$) are included in the market power and efficiency of banks' models (equations 3.9 and 3.10) among the other explanatory variables. These dynamic empirical models are rewritten as follows:

$$y_{i,t} = \alpha y_{i,t-1} + x'_{i,t}\beta + \varepsilon_{it} \qquad \text{i=1, 2…N and t=1,2…T} \qquad [3.16]$$

Where:

$\varepsilon_{it} = \eta_i + \mu_{it}$

$E(\eta_i) = 0, E(\mu_{it}) = 0 \; and \; E(\mu_{it}\eta_i) = 0$ for all I = 1, 2…N and t=1, 2…T

I characterises the particular bank and *t* denotes the year;

$y_{i,t}$= the dependent variables including LER and CE scores;

$y_{i,t-1}$= one-year lagged LER and CE score variables;

$x_{i,t}$ = the independent variables (bank characteristics and macroeconomic indicators); and

η_i = unobserved heterogeneity; and

μ_{it}= the error term.

This dynamic empirical model (equation 3.16) has two specific features:

- First, the one-year lagged value of the dependent variables (the Lerner and CE scores) are also independent regressors. This means that a regressor (one-year lagged) is correlated with dependent variables at different times. In other words, *autocorrelation* exists (Baltagi, 2008). This means that the pooled OLS and fixed effect estimation are *inefficient* (Antoniou, Guney, & Paudyal, 2008; Baltagi, 2008; Gujarati & Porter, 2009).

- Secondly, the η_i term in the error equation (ε_{it}) is linked with the dependent variable ($y_{i,t}$) (in equation 3.16). But the lagged dependent variable ($y_{i,t-1}$) is also the exploratory variable. This means that the error equation ε_{it} is interrelated with the exploratory variable. In the other words, an *endogeneity* problem potentially exists (Hill, Griffiths & Lim, 2011). In addition, bank characteristics variables $x_{i,t}$ are weakly exogenous and may be endogenous. For example, inefficiency, fee income, bank size, non-performance loans, capitalisation and concentration were considered endogenous variables in previous studies (Adjei-Frimpong et al., 2016; Erler et al., 2017; Nguyen & Nghiem, 2018; Phan et al., 2016). Hence, the OLS and fixed effect estimates will generate biased and inconsistent parameter estimates (Baltagi, 2008; Wintoki et al., 2012).

Therefore, an instrumented GMM regression is used to calculate the dynamic model (equation 3.16). This qualified GMM estimator can deal with the autocorrelation and endogeneity problems to generate unbiased, consistent parameters (Wintoki et al., 2012; Zeitun, Temimi & Mimouni, 2017).

3.4.3.1 The First-differencing GMM Regression

The first-differencing variables in equation (3.16) are used to eliminate bias from the unobserved bank-specific effects (unobserved heterogeneity) (Arellano & Bond, 1991). Equation (3.16) is changed to first-differencing form as follows:

$$y_{i,t} - y_{i,t-1} = \alpha(y_{i,t-1} - y_{i,t-2}) + \beta\left(x_{i,t} - x_{i,t-1}\right)' + (\varepsilon_{i,t} - \varepsilon_{i,t-1}) \qquad [3.17]$$

Arellano and Bond (1991) suggest using instrumented variables by lagged levels of exploratory and predicted variables for the first-differencing equation (equation 3.17). Lagged exploratory and predicted variables are used as valid instruments under the presumption that the error term is not linked to the instrument variables. Therefore, the moment conditions for each endogenous variable must satisfy the assumptions below (Arellano & Bond, 1991):

$$E\left[y_{i,t-n}(\Delta\varepsilon_i)\right] = 0 \qquad \text{for n} \geq 2,\ \text{t} = 3\ldots\text{T} \qquad [3.18]$$

$$E\left[x_{i,t-n}(\Delta\varepsilon_i)\right] = 0 \qquad \text{for n} \geq 2,\ \text{t} = 3\ldots\text{T} \qquad [3.19]$$

Where:

$y_{i,t-n}$ and $x_{i,t-n}$ = the instrumented sets used.

Equations (3.18) and (3.19) can be rewritten in a more compact form as:

$$E[Z'_{di}(\Delta\varepsilon_i)] = 0 \qquad \text{for i} = 1\ldots\text{N} \qquad [3.20]$$

Where Z_{di} is the $(\mathrm{T}-2) \times \mathrm{L}$ matrix of instruments of the individual ith and L = (T-2) (T-1)0.5+(T-2) TK for strictly exogenous variables (Greene, 2012) is given as follows:

$$Z_{di} = \begin{bmatrix} y_{i1}x'_{i1}x'_{i2}\cdots x'_{iT} & 0 & \cdots & 0 \\ 0 & y_{i1}y_{i2}x'_{i1}x'_{i2}\cdots x'_{iT} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_{i1}y_{i2}\cdots y_{iT-2}x'_{i1}x'_{i2}\cdots x'_{iT} \end{bmatrix} \qquad [3.21]$$

$$\Delta\varepsilon_i = \begin{bmatrix} \Delta\varepsilon_{i3} \\ \Delta\varepsilon_{i4} \\ \vdots \\ \Delta\varepsilon_{iT} \end{bmatrix}$$

Then the matrix with all the instruments in the model are represented as $Z_d = (Z'_{d1}, Z'_{d2}, \cdots Z'_{dN})$.

The GMM parameters $\widehat{\Phi} = (\alpha, \beta)'$ for the first-differencing in equation (3.17) under the assumption of moment conditions in equation (3.20) are created as follows (Baltagi, 2008):

$$\widehat{\Phi} = (\Delta X' Z_d A_d Z'_d \Delta X)^{-1} \Delta X' Z_d A_d Z'_d \Delta Y \qquad [3.22]$$

Where:
$X_i = ((y_{i2}, x_{i3})', (y_{i3}, x_{i4})', \cdots, (y_{iT-1}, x_{iT})')'$
$Y_i = (y_{i3}, y_{i4}, \cdots, y_{iT})'$

However, a major shortcoming of the first-differencing GMM estimator is weak instruments because of weakly exogenous lagged variables, particularly if the variables have values similar to a random walk shape (α close to one) (Baum, 2013; Bond, 2002). When there are gaps in the panel data (unbalanced panel data), difference GMM cannot calculate all observations in the orthogonal deviation (Wintoki et al., 2012). In short, differencing GMM is less efficient in such situations.

3.4.3.2 The System GMM Regression

The system GMM (sysGMM*)* regression was introduced to reduce bias and inaccuracies associated with the first-differencing GMM regression because of the problem of weak instruments (Arellano & Bover, 1995; Blundell &Bond, 1998). The sysGMM uses additional instruments, including lagged levels and lagged differences of variables (Bond, Hoeffler & Temple, 2001; Presbitero, 2005). The sysGMM estimator is thus an enlargement of the first-differencing GMM estimator. This technique has enhanced accuracy and efficiency compared with the first-differencing GMM regression in situations where the variables have values similar to a random walk shape (or α close to one) or there are gaps in the panel data (Bond, 2002; Presbitero, 2005; Wintoki et al., 2012). As a result, sysGMM has a set of added moment conditions:

$$E\left[(\Delta y_{i,t-k})\varepsilon_{i,t}\right] = 0 \qquad \text{for k} \geq 1, \text{t = 3…, T} \qquad [3.23]$$

$$E\left[(\Delta x_{i,t-k})\varepsilon_{i,t}\right] = 0 \qquad \text{for k} \geq 1, \text{t = 3…, T} \qquad [3.24]$$

Equations (3.23) and (3.24) can be written compactly as follows:

$$E[Z'_{li}\varepsilon_i] = 0 \qquad [3.25]$$

The instruments $Z_{li}$for the equations in the levels for each individual *i* are given in the 2(T-2) (T- 2) matrix as follows:

$$Z_{li} = \begin{bmatrix} \Delta y_{i2}\Delta x_{i2} & 0 & \cdots & 0 \\ 0 & \Delta y_{i3}\Delta x_{i3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Delta y_{iT-1}\Delta x_{iT-1} \end{bmatrix} \qquad \varepsilon_i = \begin{bmatrix} \varepsilon_{i3} \\ \varepsilon_{i4} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} \qquad [3.26]$$

Finally, all additional instruments are illustrated as follows:

$$Z_l = (Z'_{l1}, Z'_{l2}, \cdots Z'_{lN}) \quad [3.27]$$

The overall moment conditions of the sysGMM can be written compactly from the moment conditions of equations (3.18), (3.19), (3.23) and (3.24):

$$E[Z'_{si} p_i] = 0 \quad [3.28]$$

Therefore, the whole set of instruments is given by the 2N(T-2) (2(T-2+(T-1) (T-2)) matrix:

$$Z_{si} = \begin{bmatrix} Z_{di} & 0 \\ 0 & Z_{li} \end{bmatrix} \qquad \mathrm{p}_i = \begin{bmatrix} \Delta\varepsilon_i \\ \varepsilon_i \end{bmatrix} \quad [3.29]$$

Where:

$Z_s = (Z'_{s1}, Z'_{s2}, \cdots Z'_{sN})$.

The coefficients of the system GMM with moment conditions in equation (3.29) are computed as follows:

$$\widehat{\Phi_s} = (X'_s Z_s A_s Z'_s X_s)^{-1} X'_s Z_s A_s Z'_s Y_s \quad [3.30]$$

Where:

Xs = the stacked matrix of independent variables in differences and levels; and

Ys = the stacked vector of the dependent variables in differences and levels.

3.4.3.3 One-Step and Two-Step Procedures

The GMM estimator has two different choices, a one-step or two-step procedure. The one-step GMM estimator uses:

$$A_s = \left(\frac{1}{N} \sum_{i=1}^{N} Z'_s H Z_s \right) \quad [3.31]$$

Where:

$$H = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{bmatrix} \quad [3.32]$$

The two step uses:

$$A_s = \left(\frac{1}{N} \sum_{i=1}^{N} Z'_s \Delta\hat{\varepsilon}_i \Delta\hat{\varepsilon}'_i Z_s \right)^{-1} \quad [3.33]$$

The one-step GMM estimator uses a weighted matrix that is only competent under homoskedasticity and is uncorrelated with error terms (Presbitero, 2005). The two-step estimator is preferred because it generates more efficient results when heteroscedasticity is present in the dataset (Nguyen, 2018; Presbitero, 2005).

3.4.3.4 Other Tests

Excessive instruments

The problem of many instruments used (over-fitting) in the GMM estimator is that it can generate biased results (Roodman, 2009b). When there are numerous instruments, the Hansen test (the test of effective instruments) shows that the effectiveness of instruments will be weakened (Roodman, 2009b). To solve this problem, Roodman (2009b) suggests the following techniques:

i. Use fewer lagged variables as instruments than all available lagged variables.
ii. Use a collapsed instruments matrix to diminish the number of instruments without dropping data information.

Test of Correlation in the Error Term

Tests for the first order (AR1) and second order (AR2) show serial correlation in the error term. The AR2 test assesses the null hypothesis of non-second connection in the error term. The null hypothesis must be fulfilled to meet the model specification. Therefore, the p-value (of AR2) must exceed conventional significance (10%) levels to accept the null. This result indicates that the GMM regression is consistent (Baltagi, 2008).

Test for Exogeneity of the Instruments (or Over-identifying Restrictions Test)

The over-identifying restriction test (Hansen (1982) J-test) is used to check exogenous instruments (Bond et al., 2001; Roodman, 2009a). The null of exogenous instruments is assumed. Therefore, the p-value (of the Hansen J-test) must exceed the 10% significance level to accept the null of exogenous instruments. In other words, it indicates that the instruments are effective.

Test for Exogeneity of the Subset Instrument (or Difference-in-Hansen Test)

The sysGMM estimate also uses lagged differencing variables as additional instruments (Presbitero, 2005). These instruments need to be checked for exogeneity using a difference-in-Hansen test (Bond et al., 2001; Roodman, 2009a). The null of the exogenous subset instruments is assumed. Hence, the p-value (the difference-in-Hansen J-test) must exceed the 10% significance level to accept the null exogeneity of the subset instrument.

In summary, this study uses the advanced instrumented GMM estimator to determine the dynamic empirical models (equation 3.16). The sysGMM estimator is chosen to reduce probable biases and inaccuracies linked to the first-differencing GMM estimator because of weak instruments (Presbitero, 2005; Wintoki et al., 2012; Zeitun et al., 2017).

3.4.4 Robustness Check

This study compares the OLS, fixed effect and the sysGMM estimates. If any coefficients are plausible and robust (that is, these coefficients have similar signs and are statistically significant), that indicates structural validity (Lu & White, 2014). In other words, these are variables to which bank executives must pay attention to boost both their bank's market power and efficiency.

3.5 Objective 4: To Identify Critical Factors Influencing The Lending Decision in Vietnam Retail Banking

To identify critical factors associated with default borrowers in Vietnam retail banks, this study uses customers' data from retail banks. It uses a cross-section dataset that has a binomial outcome (good and bad loans). Therefore, logit regression is used to investigate the critical factors that impact borrowers' creditworthiness. Binary logit regression has two possible values for the dependent variable; in this case, they are default/non-default borrowers.

The probability that the loan applicant i_{th} is a 'bad' applicant (y = 1) with the given input variables X_{ij} is given as follows (Hill et al., 2011):

$$p_i = \Lambda(Z_i) = E(y_i = 1|X_{1i}, X_{2i} \dots X_{ni}) = \frac{1}{1 + e^{-Z_i}} = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{j=1}^{24} \beta_j X_{ij} + \varepsilon_i\right)}} \qquad [3.34]$$

Where:

p_i = the probability of a bad borrower (high p_i value indicates a high default risk);

y_i = 1 if the loan defaulted (a bad borrower); and 0 if loan is not defaulted (a good borrower);

β_0 = the intercept'

β_j: = the regression coefficient of applicant i_{th}; and

X_{ij} = are independent variables related to the loan-applicant i_{th} .

The probability that the loan applicant i_{th} is a 'good' applicant ($y_i = 0$) with the given input variables X_{ij} is given by (Hill et al., 2011):

$$1 - p_i = \frac{1}{1 + e^{Z_i}} \qquad [3.35]$$

The odds ratio – in favour of 'good' loans (Gujarati & Porter, 2009) is given by:

$$odds\ ratio = \frac{p_i}{1 - p_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i} \qquad [3.36]$$

We take the natural log of equation (3.36) which creates a *logit function,* as follows (Gujarati & Porter, 2009):

$$logit_i = \ln(\frac{p_i}{1 - p_i}) = \mathrm{Z}_i = \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} \qquad [3.37]$$

The logit model in equation (3.37) is the logarithm of the 'odds' ratio which is a *linear* parameter, and the partial derivative is given as:

$$\frac{\partial Logit_i}{\partial_{X_{ij}}} = \frac{\partial \ln\left(P_i/1 - P_i\right)}{\partial_{X_{ij}}} = \beta_j \qquad [3.38]$$

Equation (3.38) allows us to interpret $\beta_j$as a unit of change in X_{ij}, the logit of outcome is expected to change by β_j units, holding all other variables constant. We can calculate the marginal effect on a one-unit change in Xij on the probability of a 'bad' loan (Y = 1) by calculating the partial equation (3.34) (Hill et al., 2011)

$$\frac{\partial p_i}{\partial_{X_{ij}}} = \frac{\partial \Lambda(Z_i)}{\partial_{Z_i}} \times \frac{\partial Z_i}{\partial X_{ij}} = \beta_j \times \lambda(Z_i) = \beta_j \times \lambda\left(\beta_0 + \sum_{j=1}^{n} \beta_j X_{ij}\right) \qquad [3.39]$$

$$= \beta_j \times \frac{e^{-Z_i}}{(1+e^{-Z_i})^2} = \beta_j \times \underbrace{\frac{1}{(1+e^{-Z_i})}}_{p_i} \times \underbrace{\left[1 - \frac{1}{(1+e^{-Z_i})}\right]}_{(1-p_i)} = \beta_j \times p_i \times (1-p_i)$$

Equation (3.39) has the following characteristics (Hill et al., 2011):

- p_i is the probability of a bad loan. Values range from 0 to 1. Thus, the term $p_i \times (1 - p_i)$ is always positive with a maximum p value of 0.5. Consequently, the sign of $\partial p_i / \partial_{X_{ij}}$is decided by the sign of β_j.

- In contrast, if Z_i is very large (either positive or negative), the term $p_i \times (1 - p_i)$ is near zero. X_{ij} will have little effect on the p_i value. This result of the logit regression can be used to forecast the probability of a borrower defaulting. If p_i equals one, banks predict that the applicant will default and the loan will not be approved, although they can impose additional conditions such as extra collateral or a higher interest rate.

Comparing p_i against a threshold value (or cut-off value) such as 0.5, we can assign an i^{th} applicant to one of two groups: a good or bad borrower (Hill et al., 2011):

$$p(y_i) = \begin{cases} bad\ borrower,\ p \geq Threshold = 0.5 \\ good\ borrower,\ p < Threshold = 0.5 \end{cases} \qquad [3.40]$$

To obtain the efficient parameters β_j the maximum likelihood estimation is calculated for the logit regression (Gujarati & Porter, 2009):

$$\mathcal{L} = \prod_{y_i=1} p_i \prod_{y_i=0} (1 - p_i) \qquad [3.41]$$

Logit regression is evaluated for goodness of fit through the Homer-Lemeshow test (Bartlett, 2014; Tran & Ho, 2015). A Pearson χ^2 goodness of fit statistic is calculated. The null hypothesis assumed that the model fits the data (Tran & Ho, 2015). Therefore, the Homer-Lemeshow test's p-values must be higher than the significance level (5%) to accept the null hypothesis. In short, small χ^2 values with large test p-values indicate that logit regression is a good fit to observed data (Bartlett, 2014; Tran & Ho, 2015).

The major disadvantage of ogit regression is that the overall accuracy classificatory power is lower than advanced computing and various hybrid approaches (Trinkle & Baldwin, 2016; Yu et al., 2015). We use a hybrid technique to enhance the model's classificatory power. This method is addressed in the next section.

3.6 Objective 5: To Investigate The Classificatory Power of The CSM Based on a Hybrid Technique Using a Combination of Artificial Neural Network and Logit Regression.

3.6.1 Supervised Learning: The Multilayer Perceptron Technique

The Multilayer Perceptron (MLP) technique is classified as a supervised paradigm (Samarasinghe, 2006). The MLP network has three layers: the input, hidden and output layers (see Figure 3.3):

Figure 3.3 The MLP Architect with an Input, Hidden and Output Layer

Image removed for Copyright compliance

The processes of the MLP network with an input pattern (x = [x1, x2… xn]) and output pattern (z = [0,1]) can be summarised as follows:

- Step (i) : Begin with an initial random weight, the sum weight u_j is counted

$$u_j = a_{0j} + \sum_{i=1}^{n} a_{ij} x_i \quad [3.42]$$

▪ Step (ii) : The sum result (u_j) is transformed via logistic activation function $f(.)$ (in the hidden layer)

$$y_j = f(u_j) = \frac{e^{u_j}}{1+e^{u_j}} \quad [3.43]$$

▪ Step (iii) Continue calculating the sum of the hidden layer v_j

$$v_j = b_{0j} + \sum_{i=1}^{m} b_{ij} y_i \quad [3.44]$$

▪ Step (iv) : if the sum result v_j is smaller than the threshold then the output *y* is zero. If the sum result is higher or equal to the threshold the *y* output is one.

$$z = g(v_j) = \begin{cases} 1, & v_j \geq Threshold \\ 0, & v_j < Threshold \end{cases} \quad [3.45]$$

▪ Step (v) the error is computed (Target – output). If the error is zero, the weight is not modified. The weights are modified according to a back-propagation algorithm under the gradient descent learning rule (or delta rule) (see equations (3.46a) and (3.46b)) (Samarasinghe, 2006). This back-propagation algorithm was first introduced in 1986 by Paul J. Werbos and is widely used for training the ANN network (Matroushi, 2011).

Weights b_0 and b_i are modified by the back-propagation algorithm based on the gradient descent learning rule, as follows:

$$\begin{aligned} b_{0_{new}} &= b_{0_{old}} + \beta \frac{\partial E}{\partial b_0} = b_{0_{old}} + \beta \times p \\ b_{1_{new}} &= b_{1_{old}} + \beta \frac{\partial E}{\partial b_1} = b_{1_{old}} + \beta \times p \times y_1 \\ &\dots \\ b_{m_{new}} &= b_{m_{old}} + \beta \frac{\partial E}{\partial b_m} = b_{1_{old}} + \beta \times p \times y_m \end{aligned} \quad [3.46a]$$

with $\quad p = (z-t)z(1-z) = Ez(1-z).$

Weight $a_{0i}\ for\ i = 1 \dots m$ and $a_{ij}\ for\ i = 1 \dots m\ and\ j = 1 \dots n$ is modified by the back-propagation algorithm based on the gradient descent learning rule as follows:

$$[a_{01_{new}} = a_{01_{old}} + \beta \times q_1] \vdots [a_{02_{new}} = a_{02_{old}} + \beta \times q_2] \vdots [\dots] \vdots [a_{0m_{new}} = a_{0m_{old}} + \beta \times q_m]$$
$$[a_{11_{new}} = a_{11_{old}} + \beta \times q_1] \vdots [a_{12_{new}} = a_{12_{old}} + \beta \times q_2] \vdots [\dots] \vdots [a_{1m_{new}} = a_{1m_{old}} + \beta \times q_m]$$
$$\dots$$
$$[a_{n1_{new}} = a_{n1_{old}} + \beta \times q_1] \vdots [a_{n2_{new}} = a_{n2_{old}} + \beta \times q_2] \vdots [\dots] \vdots [a_{nm_{new}} = a_{nm_{old}} + \beta \times q_m]$$

With $q_1 = pb_1y_1(1-y_1) \vdots q_2 = pb_2y_2(1-y_2) \vdots \cdots \vdots q_m = pb_my_m(1-y_m)$ [3.46b]

Where:

x_i = the attributes of an i_{th} applicant;

n = the input neurons, m is the number of hidden neurons and k = 1 output neuron;

a_{ij} = the weighting coefficient connected to the input i and neuron j;

a_{0j} = the bias weighting coefficient of input neuron I;

b_{ij} = the weighting coefficient connected to the hidden i and neuron k;

b_{0j} = the bias weighting coefficient of the hidden neuron j;

β = the learning rate; and

E = the error term.

- Step (vi): steps (ii) to (v) are repeated until the error is minimised to the point where there is no improvement or a pre-specified number of training batch iterations (epochs) has been reached.

3.6.2 The Hybrid Technique

To improve the classification power of the CSM, our study proposes a hybrid empirical technique. The hybrid empirical technique integrates both logit regression and ANNs (the SOM and MLP techniques). The hybrid technique can be expressed using the general form (equation 3.47):

$$\boldsymbol{Lending\ decision} = \boldsymbol{H}(\boldsymbol{X_1}, \boldsymbol{X_2}, \cdots, \boldsymbol{X_n}, \boldsymbol{\varepsilon}) \qquad [3.47]$$

Where:

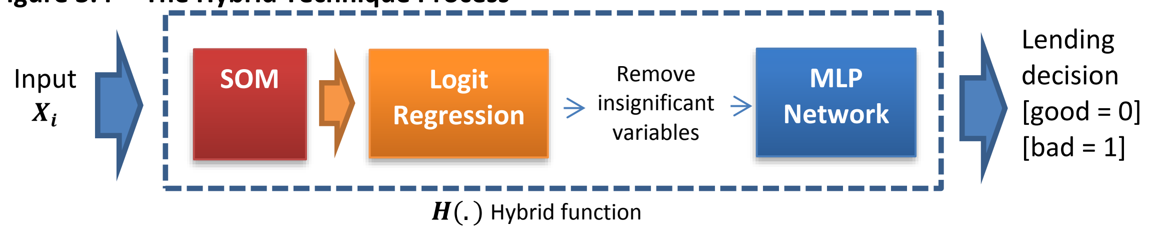
Lending decision = 1 if the loan is in 'default' (bad loan); 0 if loan is repaid (good loan);

X1, X2 … Xn = are independent variables;

ε = the error term; and

$\boldsymbol{H}(.)$ = a hybrid function explained in Figure 3.4.

Figure 3.4 The Hybrid Technique Process



Source: Author's computation

The hybrid function works in three steps. The *first* step involves screening the dataset using the SOM technique. Neuron grids such as 15 x 15, 20 x 20, to 40 x 40 are trained to obtain a minimum mean distance and clearly show cluster boundaries (Samarasinghe, 2006; Waidyarathne & Samarasinghe, 2014). The topology grid ("hexagon") was selected because it provides more neighbouring connections (Waidyarathne & Samarasinghe, 2014). The SOM is initially given a random weight value. The map is trained repeatedly in "batch mode" with 200 iterations and learning rates that vary between 0.05 and 0.1. The Euclidean distance is used to find a winning neuron. The Gaussian neighbourhood formula is used to correct weight smoothly with distance. This process is continued until the SOM is thoroughly trained. The mean within group sum of squares (WSS) is used to decide the optimal clusters based on the elbow method (Thorndike, 1953). As the WSS value decreases, the number of clusters increases (from two to the maximum number of clusters). The optimal number of clusters can be determined by establishing which has the greatest effect on the WSS value (Waidyarathne & Samarasinghe, 2014). For example, the WSS equals 10, 6, 5, 4 and 4 in responding to numbers of clusters that are 1, 2, 3, 4, and 5. The optimal number of clusters is two because it has the highest reduction in WSS value of four (10 - 6). Hierarchical analysis is used to confirm the optimal clusters decided by the WSS value. Finally, each weight vector generated by the SOM technique denotes a group input data vector. Therefore, a final 2D SOM map displays a clear visualisation of a group of cluster input vectors. The final 2D SOM map results identify any ordinary clusters, which may determine the outputs (default or non-default borrowers). In other words, this step helps to understand the dataset and presents the relationships between inputs/outputs for the next steps. The SOM technique also creates component plans that reflect correlations between inputs and outputs (default/non-default borrowers). Analysing these correlation coefficients investigates the most or least input variable that influences the outputs.

Second, the original dataset is estimated using logit regression. We expect to find significant variables associated with the probability of default borrowers. We use the Wald test to determine whether to add variables that improve the results (Lu et al., 2013). Insignificant variables are removed from the input variables. Removing insignificant variables helps lessen the noise in the data and may generate better results.

Third, the remaining significant variables included in the MLP network are used to classify 'non-default' and 'default' borrowers. The Levenberg-Marquardt algorithm is used to train the networks. The numbers of neurons in the input and output layers are identified, but the structure of optimal layers and neurons in the hidden layer are still ambiguous and need further research (Samarasinghe, 2016; Waidyarathne & Samarasinghe, 2014). The accuracy of the MLP technique varies in the choice of layers and neurons in the hidden layer (Samarasinghe, 2016; Trinkle & Baldwin, 2016). There are no formal guidelines for choosing the layers or neurons in the hidden layer. Therefore, this study selects random neurons with one hidden layer and runs a trial. The appropriate number of neurons is determined using the most accurate results (those that generate maximum values of evaluation criteria, such as AC rate, AUC ratio, K-S ratio and Gini ratio) and considering the acceptable error levels.

By deleting insignificant variables (from the logit regression results), the result is expected to be more accurate in the overall classification rate, especially regarding Type II errors; that is, the number of 'bad' borrowers classified as 'good' borrowers', will be reduced significantly as documented in Lu et al.'s (2013) study. In other words, the hybrid technique builds a more efficient CSM (Lu et al., 2013).

3.6.3 The Evaluating Criteria for Examining Classification Power

This study uses several criteria to measure the classificatory power of the hybrid technique. These are the confusion matrix, the misclassification cost, the area beneath the receiver operating characteristics curve (AUC), the Kolmogorov–Smirnov (KS) statistic, and the GINI coefficient. Evaluation criteria are calculated for both techniques to show improvements in classificatory power between logit regression and the hybrid technique. The criteria are explained in the following subsection.

3.6.3.1 The Confusion Matrix

The 'confusion matrix' displays actual and predicted values as shown in Table 3.1 (Tsai, 2014). This matrix enables researchers to recognise errors quickly and to evaluate whether the technique estimates correctly.

Table 3.1 The Confusion Matrix

	Predicted Observation		
	G	**b**	
Actual Observations			
G	Gg	Gb(II)	TG
B	Bg(I)	Bb	TB
	Tg	Tb	TN
SENS	Gg/Tg		
SPEC		Bb/Tb	

Where:

G = the actual observations that are good (non-default) borrowers;

g = the predicted observations that are good (non-default) borrowers;

B = the actual observations that are bad (default) borrowers;

b = the predicted observations that are bad (default) borrowers;

TG = the total actual observations that are good (non-default) borrowers;

TB = the total actual observations that are bad (default) borrowers;

Tg = the total predicted observations that are good (non-default) borrowers;

Tb = the total predicted observations are bad (default) borrowers;

TN = the total observations in the dataset;

SPEC = quantifies the percentage of correctly actual bad to total projected default observations ($SPEC = B_b/T_b$); and

SENS = quantifies the percentage of correctly actual good to total projected non-default observations ($SENS = G_g/T_g$).

Important criteria can be used to assess the classificatory power of the technique: the overall accuracy rate (AC rate), and Type I and II errors (see equation 3.48).

$$AC\ rate = \frac{G_g+B_b}{TN};\ Type\ I\ error\ = \frac{G_b}{TG};\ Type\ II\ error\ = \frac{B_g}{TB} \quad [3.48]$$

A technique (between logit and hybrid) that generates a higher AC rate and lower Type I & II errors can be seen as the better technique.

3.6.3.2 The Misclassification Cost

The second evaluating criterion is the misclassification cost. This criterion is defined in equation (3.49) (Abdou, 2009; West, 2000):

$$EMC = C_{21}P_{21}\pi_1 + C_{12}P_{12}\pi_2 \quad [3.49]$$

Where:

$\pi 1$ and $\pi 2$ = the prior probabilities of a default/non-default borrower;

P_{21} and P_{12} = quantify the probability of Type I and II errors, respectively; and

C_{21} and C_{12} = the corresponding costs of misclassifying of Type I and II errors.

The ratio of misclassification cost related to Type I and II errors, ranges from 1:1 to 5:1 in West's (2000) study and 1:1 to 10:1 in Abdou's (2009) study. Therefore, this study will provide a sensitivity analysis within a ratio range of 1:1 to 10:1. A technique that generates lower misclassification costs is seen as a better technique.

3.6.3.3 The AUC Ratio

The third criterion for evaluating classification power is the AUC ratio. This ratio is computed by equation (3.50) (Van Gool et al., 2012)

$$AUC = \int_0^1 SENS(1 - SPEC) \times \partial(1 - SPEC) \quad [3.50]$$

Where:

SPEC = quantifies the percentage of correctly actual bad to total projected default observations ($SPEC = B_b/T_b$); and

SENS = quantifies the percentage of correctly actual good to total projected non-default observations ($SENS = G_g/T_g$).

The AUC value fluctuates from 0.5 to 1. However, the AUC value must be between 0.7 to 0.9 to provide good classificatory power (Chi & Hsu, 2012; Van Gool et al., 2012).

3.6.3.4 The K-S Statistic

Other important criterion for evaluating model performance is the K-S statistic (Crook et al., 2007). The K-S statistic quantifies the partition between the cumulative distribution of 'good' and 'bad' borrowers (Chi & Hsu, 2012; Crook et al., 2007). The point that generates the largest partition among the distribution of 'good' and 'bad' reflects the cut-off value for accepting or rejecting an applicant (Chi & Hsu, 2012). A higher K-S ratio indicates a better technique (Chi & Hsu, 2012; Crook et al., 2007). The K-S ratio is calculated as follows (Chi & Hsu, 2012):

$$K - S = max_s[F(s|Good) - F(s|Bad)] \quad [3.51]$$

Where:

F(s|Good) and F(s|Bad) are spreading functions of 'Goods' and 'Bads'.

3.6.3.5 The Gini ratio

A related measuring classificatory power is the Gini ratio, which is computed as follows (Equation 3.52) (Abdou et al., 2016):

$$AUC\ value = \frac{Gini_{cofficient} + 1}{2} \quad [3.52]$$

The classificatory power is considered low if the Gini ratio is lower than 0.25 and very good if the Gini ratio is higher than 0.6 (Abdou et al., 2016).

The evaluation outcomes generated from the logit regression and hybrid techniques are compared to determine both techniques' classificatory power. The hybrid technique is better than logit regression in classificatory power when the criteria (AC rate, AUC ratio, the K-S statistic and the Gini ratio) generated by the hybrid technique are higher than those criteria produced by logit regression. Misclassification costs generated by the hybrid technique are lower than the costs created by logit regression. As the hybrid technique is better, it can be used to build a more accurate CSM.

3.7 Objective 6: To Investigate Factors That Determine The Interest Rate Charged on Retail Loans.

To investigate the factors that influence a customer's interest rate, the interest rate is used as a dependent variable in a regression function. OLS regression is used to estimate the parameters. The interest rate model is expressed mathematically as follows (equation 3.53):

$$Y_i = \beta_0 + \sum \beta_i X_i + \varepsilon_i \quad [3.53]$$

Where:

Y_i = the interest rate charged on retail loans by a bank;

βo = the intercept

βi = the estimating coefficients;

X_i = the independent variables' (factors) impact on loan prices; and

ε_I = the error term.

The coefficient β can be worked out as follows (Gujarati & Porter, 2009):

$$\beta_{interest\ rate} = \left(\sum_{i=1}^{n} X'_i X_i \right)^{-1} \sum_{i=1}^{n} X'_i Y_i \qquad [3.54]$$

Lending interest rate factors, which are significant in the lending interest rate model (equation 3.53), are used to benchmark the default risk factors that were identified in section 3.5. Benchmarking between these two factor groups reflects whether the bank ranks customers according to their risk level or charges them different lending rates. The next section explains, in detail, benchmarking between lending interest rate factors and default risk factors.

3.8 Objective 7: To Examine Bank Ranked Customers According to Their Risk Level to Offer a Loan Price

Based on the critical risk factors associated with the probability of default borrowers (identified in section 3.5), banks can divide borrowers into 'low-risk' and 'high-risk' groups. Banks can offer different lending interest rates based on the RBP strategy; that is, lower-risk borrowers are offered lower loan prices and higher risk borrowers are given higher loan prices. To examine whether the bank segments customers before deciding loan prices, we compare the risk factors indicated as good/bad borrowers and the critical factors determining the lending interest rate. There are two possible outcomes:

I. If the risk factor has the same sign (negative or positive) as the factor that impacts the lending interest rate, we can conclude that banks uses risk factors to determine lending interest rates. In short, low-risk borrowers are offered a lower loan price, and high risk or 'bad' borrowers are charged more.

II. If any critical risk factor is different from the factors that determine the lending interest rate, we can conclude that banks do not use risk factors to decide the loan price. In other words, the RBP strategy is not applied for this risk factor.

Based on how many factors are used with the RBP rule, we can assess the effectiveness of the RBP strategy deployed by banks.

3.9 Chapter Conclusion

This study uses the SOM technique to build the financial trajectories of Vietnam banks to answer research objective one. It is hoped that this study will provide typical financial trajectory patterns. These patterns could be helpful to bank executives who could use them to identify potential risky financial situations and take corrective action (Chen et al., 2013). Vietnam banks can be categorised into super-class groups that can be used for further investigation.

To answer research objective two, this study quantifies bank market power using the Lerner index. The market power (represented by the Lerner index) of each bank is computed at time *t*. The two-sample t-test is used to compare market power between super-class groups of banks. If the test's p-value reaches an accepted significance level, the outcome indicates a specific situation in the Vietnam banking industry. There are two bank groups, with different levels of market power (low and high), that exist in the Vietnam banking industry. A group of strong banks (or high market power) exists that is isolated from other banks. This strong bank group tends to comprise monopolists and damages the perfectly competitive atmosphere of Vietnam banks. This study also examines bank performance (or efficiency) using the DEA technique. After the efficiency (represented by the PTE and CE scores) of Vietnam banks is computed for each bank at time *t*, the two-sample t-test is used to compare the differences (PTE and CE scores) among the various super-class bank groups. If the p-value indicates an acceptable significance level, this indicates differences in efficiency of the Vietnam banking industry

Two empirical models, market power and cost efficiency of banks, are used to examine the determinants that impact the market power and efficiency of Vietnam banks to answer research objective three. Pooled OLS and fixed-effect estimators are used to predict coefficient parameters of the two empirical models. The qualified sysGMM estimator is also used to examine the dynamic relationship between market power and cost efficiency of banks. This qualified GMM estimator deals with heteroskedasticity, autocorrelation and endogeneity problems and generates unbiased, consistent results (Presbitero, 2005; Wintoki et al., 2012; Zeitun et al., 2017). Finally, the results obtained from the OLS, fixed effect, and GMM estimators are compared to verify plausible and robust determinants. These robust determinants (same sign and significance) provide strong evidence that bank executives can use to enhance both the market power and efficiency of banks.

To respond to research objective four, logit regression is used to investigate the determinants associated with default borrowers in the Vietnam retail banking industry. Binary logit regression has two possible outcomes: good or bad borrowers. The hybrid empirical technique integrates both logit regression and ANN (including the SOM and MLP techniques) to improve the CSM's classificatory power to answer research objective five. The confusion matrix, the misclassification cost, the AUC, the KS statistic, and the GINI ratio are used as evaluation criteria to measure classificatory power of both the logit regression and hybrid techniques. If the criteria confirm that the hybrid technique is better, then the hybrid technique will result in a more efficient CSM.

OLS regression is used to estimate critical factors that impact the lending interest rate to answer research objective six. By comparing the critical risk factors indicating 'good' versus 'bad' borrowers and the critical factors that determine the interest rate charged by the banks, we can assess whether banks segment their customers before determining loan price in answering research objective seven.

In other words, we can evaluate whether the bank implements the RBP strategy. Based on how many factors comply with the RBP rule, (that is, lower-risk borrowers are offered a lower loan price and higher-risk borrowers receive a higher loan price), we can assess the effectiveness of the RBP strategy. The next chapter discusses the empirical results of our study models.

Chapter 4

Research Results and Findings of Domestic Vietnam Banks

4.1 Introduction

This chapter presents the results for objectives one to three of the study. More specifically, it presents the empirical financial trajectory patterns of the sampled banks; fundamental economic indicators representing bank market power and efficiency; and, finally, the determinants of market power and efficiency. The chapter is organised as follows: section 4.2 illustrates the financial trajectory patterns of domestic Vietnam banks over the study period 2008 to 2017. Section 4.3 discusses the fundamental economic indicators used to measure bank market power (represented by the Lerner index) and bank efficiency (represented by PTE and CE scores). Section 4.4 examines the determinants that directly impact banks' market power and efficiency. Section 4.5 summarises the chapter's key findings.

4.2 Objective 1: To Investigate Financial Trajectories and Categorise Vietnam Banks Using The SOM Technique

4.2.1 Using the SOM Technique to Detect the Banks' Financial Locations in 2D SOM Maps

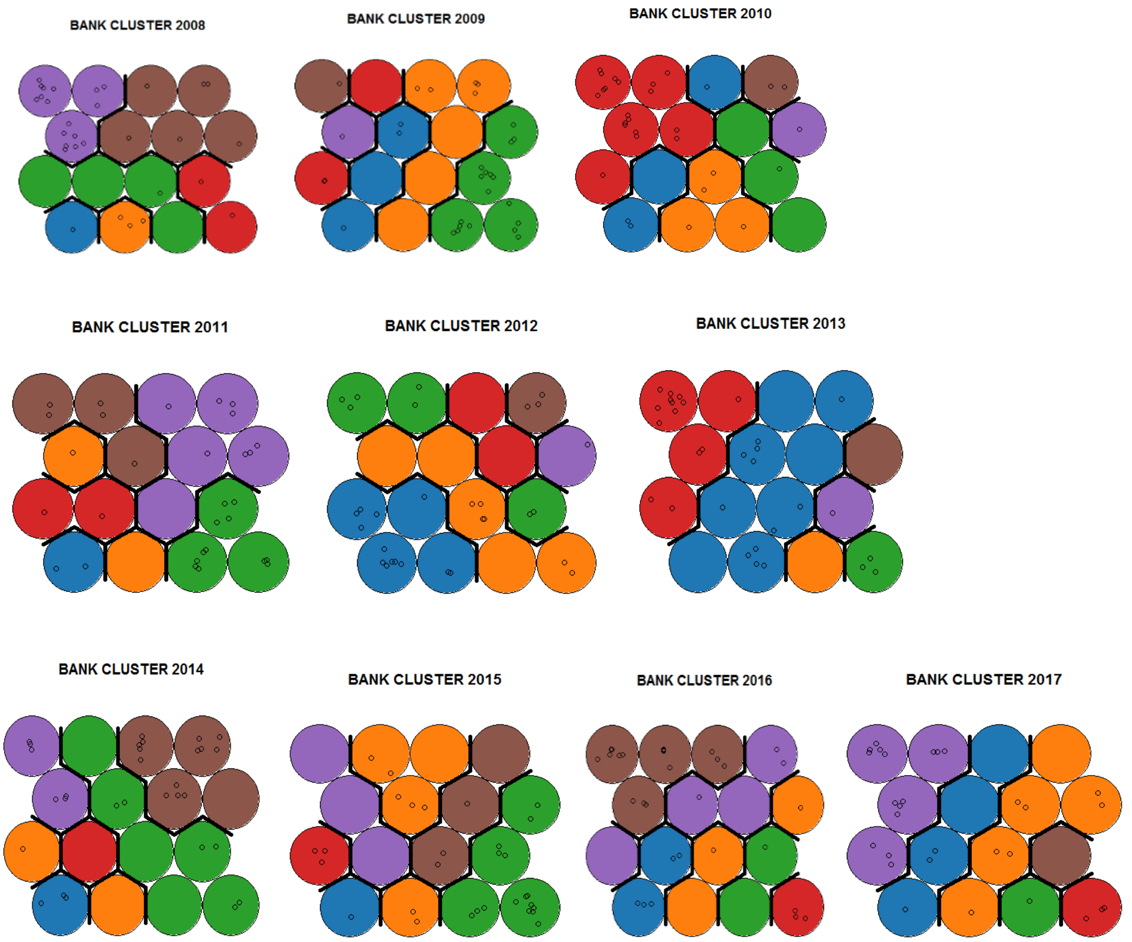
As explained in Chapter 3, this study includes four state-owned banks and 28 Vietnam-owned commercial joint-stock banks (see Appendix Table A1). It does not include data from the nine foreign-owned banks (HSBC, Standard Chartered, ANZ, Shinhan Bank, Public Bank Bhd, Hong Leong Bank, Woori, UOB and CIMB). With the data deficiencies, there are 25 to 27 banks with unbalanced panel data (258 observations) over the 10-year period 2008 to 2017. These 25 to 27 banks (four state-owned banks and 21 to 23 Vietnam-owned commercial joint-stock banks) are referred to throughout the study as Vietnam domestic banks.

The first step in finding a bank's trajectory involves locating it within a 2-dimensional (2D) SOM map, based on its provided financial information. Appendix Table A2 summarises Vietnam domestic banks' financial data over the study period.

SOM is trained using a variety of neuron grids such as 3 x 3, 4 x 4, 5 x 5, to obtain a map that best illustrates the probability distribution of the data and clearly shows cluster boundaries (Samarasinghe, 2006; Waidyarathne & Samarasinghe, 2014). The SOM 4x4 grid was selected for capturing bank locations because this SOM grip had a minimum quantisation error (Waidyarathne & Samarasinghe, 2014). The 'hexagonal' structure was selected because it gives more neighbouring connections and results in smoother maps (Samarasinghe, 2006). The SOM is initially given a random weight value and the map is trained repeatedly with 100 iterations and a learning rate from 0.01 to 0.02. The Euclidean distance is used to select a winning neuron and the Gaussian neighbourhood function is used to adjust weight smoothly across distance. This procedure is repeated multiple times until the SOM is thoroughly trained.

Ten SOM results correspond to the financial data for the ten years from 2008 to 2017. Appendix Figure B1 shows the training process of the SOM technique, from iteration 1 to 100, from 2008 to 2017. The mean distance reaches a maximum and drops to the minimum at iteration 100. This indicates that the 4x4 neuron grid is closest to the data information. Each weight vector generated by the SOM technique represents input data (Samarasinghe, 2006). Consequently, the final SOM maps show the locations of Vietnam domestic banks in the 2D neuron map (see Figure 4.1). The 2D mapping plot in Figure 4.1 displays the sum of banks in each neuron in each year from 2008 to 2017. There are some empty neurons and some neurons that contain many banks. This implies that the SOM technique has categorised the local Vietnam banks into different groups. In other words, the SOM results provide a clear picture of which banks are categorised into each specific neuron.

Figure 4.1 The 2D SOM Map Results Separating Vietnam Banks into Categories at Each Specific Neuron and Group from 2008 to 2017



Note: Orange, purple, red, brown, green and blue colour indicate groups 1 to 6.
Source: Author's computation

4.2.2 Clustering Analysis

This study categorises 16 neurons into six groups. The six groups are shown in orange, purple, red, brown, green, and blue in Figure 4.1. The SOM results show that group 1 always contains at least 9 banks (ABBank, NVB, OCB, Viet A bank, Nam A Bank, Vietbank, KienLongBank, PGbank and Saigonbank) from 2008 to 2017. In contrast, groups 2 to 6 contain one to five banks (see Figure 4.1).

Based on the elbow method, the mean within group sum of squares (WSS) is used to select the optimum number of clusters. The optimum number of clusters is selected at a cluster that has the largest change in WSS value (Potočnik et al., 2012; Thorndike, 1953; Waidyarathne & Samarasinghe, 2014). Appendix Figure B2 lists the WSS values based on the number of clusters each year from 2008 to 2017. The WSS value drops dramatically when the number of clusters increases from 1 to 2. This suggests that the 16 neurons may be categorised into two clusters. Hierarchical analysis, using the "Canberra" linkage method, was used to identify the optimum number of clusters. Appendix Figure B3 contains the dendrogram of the 16 neurons from 2008 to 2017. The dendrograms confirm that two separate groups exist. The endrograms also suggest that groups 2 to 6 are grouped into one cluster. Therefore, groups 2 to 6 are clustered to create a super-class group 2 bank; group 1 is super-class group 1 banks. As a result, the 16 neurons from the 2D mapping plots are grouped in two clusters, or two super-class bank groups. In short, super-class bank groups 1 and 2 represent two clusters of 16 neurons, which represent 27 banks (see Table 4.1).

Table 4.1 demonstrates how the 27 Vietnam domestic banks are distributed among the 16 neurons and two super-class groups (1 or 2); in short *group (1 or 2) banks*, over the period 2008 to 2017. An individual bank may be in different neurons but settles in the same group (1 or 2) over 2008 to 2017. For instance, the Saigonbank is located in various neurons (V9, V8, V3, V13, and V16) in the study period. However, all these neurons are in group 1 so the Saigonbank belongs to group 1 banks. In contrast, several banks move from group 1 banks to group 2 banks and vice versa.

Through observing dynamic changes in each bank's financial location in the 2D SOM results, we can build financial trajectory patterns. These trajectory patterns provide information about the bank's financial situation over time. Trajectory patterns enable bank executives to recognise potential risks and respond to any warning signs (Chen, 2012; Du Jardin & Séverin, 2011; 2012).

Figure 4.2 also shows the number of banks in each specific bank group (either 1 or 2). The number of banks in groups 1 and 2 fluctuates over the 10-year period. In 2017, group 1 had 13 banks and group 2 had 14. Group 2 banks have four commercial state banks and several joint-stock banks whereas group 1 consists of only commercial joint-stock banks. Appendix Table B4 provides the statistical/financial data for both groups of banks. The mean of total assets over the 10 years is 38,053 billion VND for group 1 banks and 280,194 billion VND for group 2 banks. In other words, group 2 banks are larger (have greater total assets) than group 1 banks.

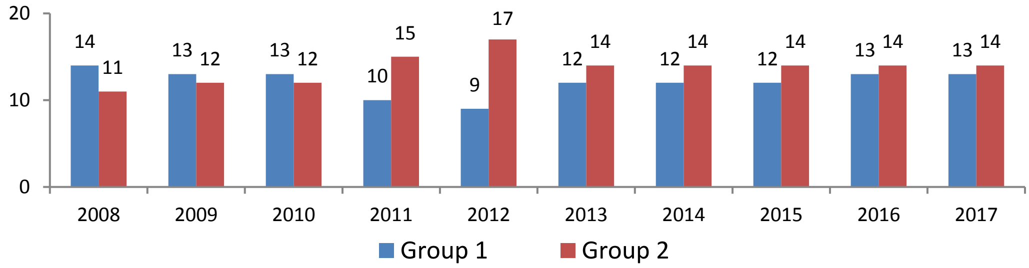
Table 4.1 Distribution of Local Banks into Neurons and Groups in the 2D SOM Maps (2008 to 2017)

	2008		2009		2010		2011		2012		2013		2014		2015		2016		2017	
	V	G	V	G	V	G	V	G	V	G	V	G	V	G	V	G	V	G	V	G
BIDV	2	2	9	2	16	2	5	2	16	2	4	2	1	2	5	2	4	2	4	2
Agribank	1	2	13	2	12	2	1	2	16	2	4	2	1	2	5	2	4	2	4	2
Vietinbank	2	2	5	2	16	2	1	2	16	2	4	2	1	2	5	2	4	2	4	2
VCB	2	2	5	2	8	2	6	2	12	2	8	2	5	2	1	2	8	2	3	2
SCB	16	2	16	2	10	1	n.a.		8	2	2	2	9	2	14	2	7	2	12	2
Sacombank	4	2	10	2	7	2	10	2	14	2	2	2	9	2	14	2	12	2	12	2
Mbbank	11	2	15	2	15	2	14	2	13	2	2	2	9	2	10	2	1	2	11	2
Techcombank	8	2	10	2	7	2	13	2	13	2	7	2	13	2	2	2	6	2	7	2
SHB	9	2	12	1	10	1	11	2	8	2	7	2	13	2	10	2	1	2	2	2
ACB	7	2	1	2	3	2	9	2	14	2	2	2	13	2	10	2	1	2	11	2
VPBank	10	2	12	1	5	1	16	2	4	2	6	2	4	2	2	2	6	2	7	2
HDBank	9	1	8	1	14	1	12	2	7	2	10	2	10	2	7	2	16	2	1	2
LienViet	14	1	8	1	14	1	12	2	7	2	10	2	10	2	7	2	16	2	6	2
Eximbank	12	1	15	2	2	3	13	2	13	2	16	2	4	2	11	2	10	2	6	2
TPBank	13	1	3	1	9	1	3	1	5	2	13	1	15	1	8	1	14	1	5	1
VIB	16	2	16	2	1	2	16	2	7	2	5	1	11	1	3	1	15	1	5	1
Maritimebank	15	2	16	2	1	2	15	2	4	2	10	1	8	1	12	1	14	1	9	1
BacAbank			n.a.				8	1	2	1	9	1	11	1	3	1	14	1	9	1
ABBank	9	1	12	1	14	1	12	1	2	1	9	1	11	1	8	1	14	1	9	1
NVB	9	1	8	1	9	1	8	1	1	1	13	1	15	1	4	1	9	1	14	1
OCB	9	1	8	1	9	1	8	1	6	1	14	1	15	1	8	1	9	1	9	1
Viet A bank	9	1	8	1	9	1	8	1	1	1	13	1	15	1	4	1	9	1	14	1
Nam A Bank	14	1	3	1	13	1	3	1	1	1	13	1	15	1	4	1	13	1	14	1
Vietbank							n.a.								4	1	13	1	13	1
KienLongBank	13	1	3	1	13	1	3	1	1	1	13	1	16	1	4	1	13	1	13	1
PGbank	14	1	3	1	9	1	3	1	1	1	13	1	16	1	4	1	13	1	13	1
Saigonbank	9	1	8	1	9	1	3	1	1	1	13	1	16	1	4	1	13	1	13	1

Note: V = neuron number; G = group number. n.a. = Not available

Source: Author's computation

Figure 4.2 The Annual Number of Banks in Bank Groups 1 and 2

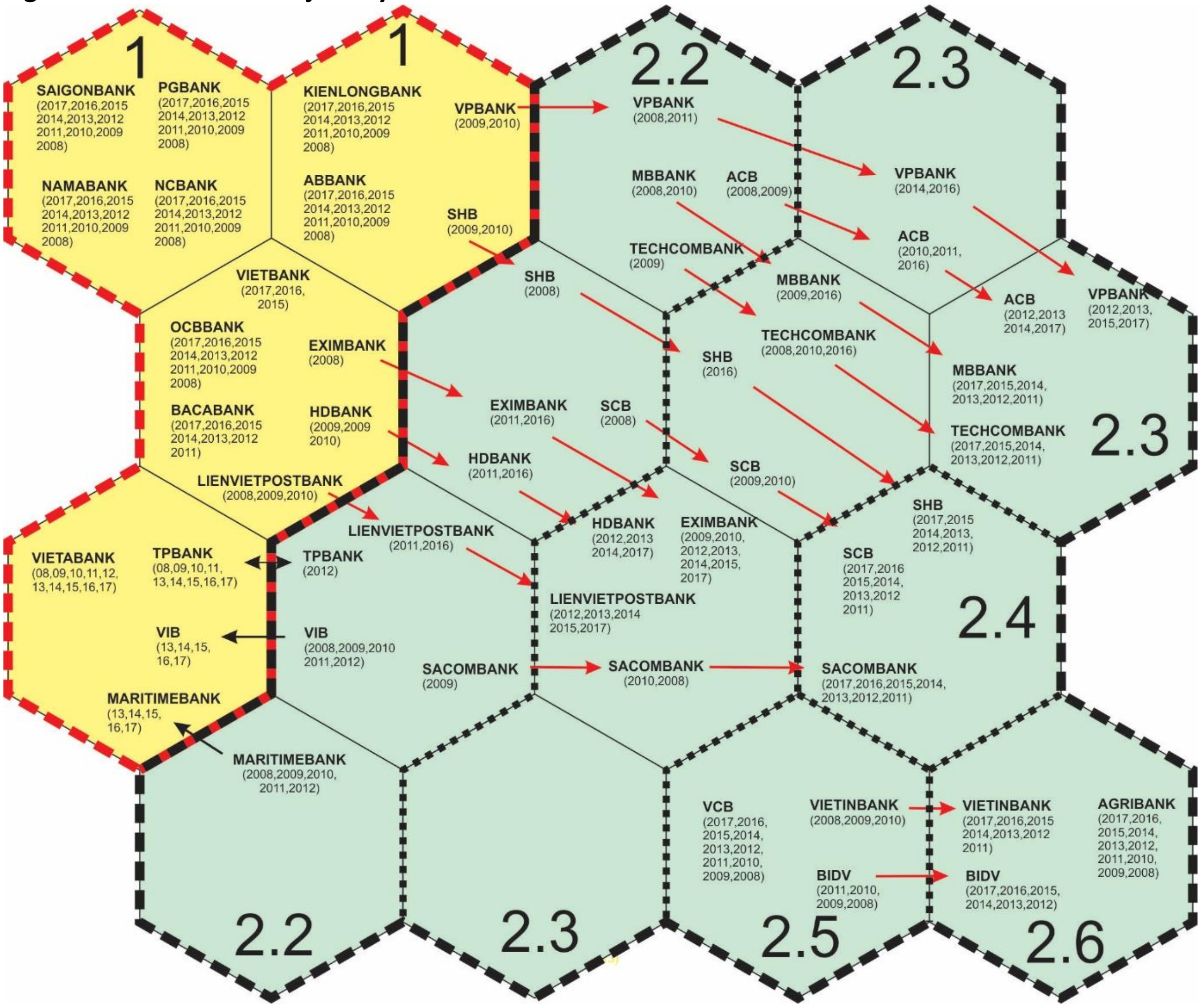


Source: Author's computation

4.2.3 Dynamic Evaluation Phase

In the second step of building bank trajectory patterns, ten 2D mapping plots (see Figure 4.1) from 2008 to 2017 are overlapped on each other. Each bank is observed and connected from neuron to neuron; group to group, to find its specific trajectory. The trajectory pattern results are shown in Figure 4.3. Table 4.2 shows how Vietnam domestic banks are located within super-class groups 1 and 2 from 2008 to 2017.

Figure 4.3 The Final Trajectory Patterns of Vietnam Banks from 2008 to 2017



Note: '1' = super-class group 1 bank; '2' = super-class group 2 bank;
2.2 to 2.6 = group 2 to 6 are grouped to a super-class group 2 bank
Source: Author's computation

Figure 4.3 and Table 4.2 show that some banks remained in the same group over the study period. (from 2008 to 2017). Ten banks maintained their position as group 1 banks. These were: BacAbank, ABBank, NVB, OCB, Viet A bank, Nam A Bank, Vietbank, KienLongBank, PGbank and Saigonbank. Similarly, nine banks were always group 2 banks. These were: BIDV, Agribank, Vietinbank, VCB, SCB, Sacombank, Mbbank, Techcombank and ACB (see Table 4.2 and Figure 4.3). However, five banks shifted position from group 1 banks to group 2 banks. They were SHB, VPBank, HDBank, LienVietPostbank and Eximbank. These banks (SHB, VPBank, HDBank, LienVietPostbank and Eximbank) were in group 1 banks from 2008 to 2010 but moved to group 2 banks from 2011 to 2017. The shift from group 1 banks (the small assets bank group) to group 2 banks (the bigger assets bank group) indicates growth or development. The remaining banks moved from group 2 banks to group 1 banks, or from the big to the smaller bank asset value bank group. For instance, VIB and Maritime were in group 2 from 2008 to 2012 but changed to group 1 from 2013 to 2017. This movement indicates that they had decreased in size in the Vietnam banking market.

Table 4.2 Bank Group Results (1 or 2) for the Sampled Banks from 2008 to 2017

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
BIDV	2	2	2	2	2	2	2	2	2	2
Agribank	2	2	2	2	2	2	2	2	2	2
Vietinbank	2	2	2	2	2	2	2	2	2	2
VCB	2	2	2	2	2	2	2	2	2	2
SCB	2	2	2	n.a	2	2	2	2	2	2
Sacombank	2	2	2	2	2	2	2	2	2	2
Mbbank	2	2	2	2	2	2	2	2	2	2
Techcombank	2	2	2	2	2	2	2	2	2	2
SHB	2	1	1	2	2	2	2	2	2	2
ACB	2	2	2	2	2	2	2	2	2	2
VPBank	2	1	1	2	2	2	2	2	2	2
HDBank	1	1	1	2	2	2	2	2	2	2
LienViet	1	1	1	2	2	2	2	2	2	2
Eximbank	1	2	2	2	2	2	2	2	2	2
TPBank	1	1	1	1	2	1	1	1	1	1
VIB	2	2	2	2	2	1	1	1	1	1
Maritimebank	2	2	2	2	2	1	1	1	1	1
BacAbank		n.a.		1	1	1	1	1	1	1
ABBank	1	1	1	1	1	1	1	1	1	1
NVB	1	1	1	1	1	1	1	1	1	1
OCB	1	1	1	1	1	1	1	1	1	1
Viet A bank	1	1	1	1	1	1	1	1	1	1
Nam A Bank	1	1	1	1	1	1	1	1	1	1
Vietbank				n.a.					1	1
KienLongBank	1	1	1	1	1	1	1	1	1	1
PGbank	1	1	1	1	1	1	1	1	1	1
Saigonbank	1	1	1	1	1	1	1	1	1	1

Note: n.a. = not available. ‘1’= group 1 banks; ‘2’ = group 2 banks
Source: Author’s analysis

4.2.4 Conclusion

In this section, the SOM technique was used to categorise Vietnam domestic banks into super-class bank groups. There are two super-class banks: group 1 and group 2 banks. Group 1 banks have lower assets than group 2 banks. Group 2 banks contained the four commercial state banks and several joint-stock banks during the study period. Group 1 consists of only commercial joint-stock banks.

The 2D self-organisation map provides a clear picture of the financial trajectory patterns in the Vietnam banking industry for 2008 to 2017. Some common financial trajectory patterns have been identified. 19 of 27 banks have retained their position in group 1 or group 2 banks over the entire study period. In contrast, eight banks shifted position from group 1 (small group banks) to group 2 (the bigger banks) or vice versa. This result implies that the presence of rigidity in the Vietnam banking industry may make it difficult for weak banks to compete. Most previous studies did not show any significant difference in bank market power and efficiency between these groups of banks (state-owned and joint stock banks)

(Nguyen & Nghiem, 2018; Vu & Turnell, 2010). This study uses the SOM technique to categorise Vietnam banks into super-class group 1 bank and group 2 banks. Thus, we hypothesise the following relationship:

H_{1_0}: The mean difference in market power between groups 1 and 2 banks is zero.

H_{2_0}: The mean of difference in cost efficiency between groups 1 and 2 banks is zero.

These hypothesises are tested in the next section.

4.3 Objective 2: To Measure and Compare Different Groups of Banks' Market Power and Efficiency as Categorised by The SOM Technique (the Super-Classes Groups 1 and 2 Banks).

4.3.1 Measuring Group 1 and 2 Bank's Market Power

4.3.1.1 Dataset

This study uses the non-structural Lerner index to assess the market power of Vietnam domestic banks. Table 4.3 lists the financial information is used to quantify the Lerner index from 2008 to 2017. These are unbalanced panel data with 258 observations.

Table 4.3 The Financial Data Used to Compute the Lerner Index of Vietnam Banks 2008 - 2017

Variables			Obs	Mean	SD	Min	Max
Total Costs	TC	Billion VND	258	10,873	13,659	106	66,993
Total Assets	Q	Billion VND	258	169,857	227,337	2,419	1,202,284
Total Revenue	TR	Billion VND	258	13,795	17,505	195	88,560
Price of Staff Cost	w1		258	0.0078	0.0028	0.0031	0.0184
Price of Deposits	w2		258	0.0851	0.0455	0.0316	0.3927
Price of Capital	w3		258	1.0500	1.3496	0.0792	17.3288

Note: Obs = observation; SD = Standard deviation
Source: Author's computation

4.3.1.2 Empirical Results and Discussion

The Lerner index results are calculated and are displayed in Table 4.4. The Table 4.4 shows the average price is 0.089, which is greater than the average marginal cost value (0.07). This result implies that market power exists in Vietnam domestic banks. The average Lerner index score of Vietnam domestic banks, from 2008 to 2017, is 0.210. According to Delis & Pagoulatos (2009), moderately competitive behaviour exists when the Lerner index ranges between 0.1 and 0.4. Thus, the Vietnam banking industry would be considered moderately competitive during the study period.

The Lerner index score tends to increase slightly in 2016 and 2017 (0.216 and 0.237, respectively) (see Table 4.4). A slight increase in the overall Lerner index implies a decline in competition. This decline in competition could be because of the decreasing number of Vietnam domestic banks, from 43 banks in

2008 to 32 banks in 2017 (Hoang et al., 2016). Fewer banks could lead to less competition and greater market power.

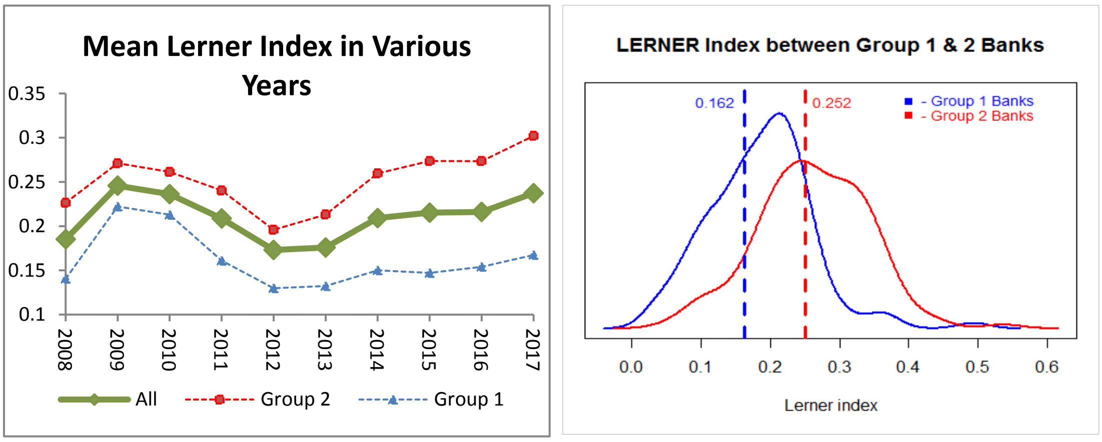
Table 4.4 The Lerner Index Values of Vietnam Domestic Banks from 2008 to 2017

Year	No. of Banks	All		Group 2 Banks		Group 1 Banks	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
PANEL A: PRICE							
2008	25	0.116	0.028	0.118	0.035	0.114	0.019
2009	25	0.077	0.013	0.076	0.013	0.078	0.015
2010	25	0.086	0.015	0.086	0.013	0.087	0.018
2011	25	0.125	0.025	0.115	0.016	0.141	0.030
2012	26	0.112	0.020	0.105	0.015	0.126	0.022
2013	26	0.084	0.016	0.085	0.014	0.082	0.019
2014	26	0.073	0.012	0.074	0.009	0.073	0.016
2015	26	0.070	0.013	0.070	0.013	0.069	0.014
2016	27	0.070	0.014	0.072	0.016	0.067	0.011
2017	27	0.073	0.015	0.076	0.020	0.070	0.006
Mean		**0.089**	0.017	**0.088**	0.016	**0.091**	0.017
PANEL B: MARGINAL COSTS							
2008	25	0.094	0.027	0.089	0.021	0.100	0.031
2009	25	0.058	0.013	0.055	0.014	0.061	0.012
2010	25	0.066	0.013	0.064	0.016	0.068	0.011
2011	25	0.099	0.023	0.088	0.015	0.117	0.021
2012	26	0.093	0.022	0.085	0.017	0.109	0.021
2013	26	0.069	0.015	0.067	0.014	0.071	0.016
2014	26	0.058	0.013	0.055	0.010	0.062	0.014
2015	26	0.055	0.012	0.051	0.010	0.059	0.013
2016	27	0.054	0.010	0.051	0.009	0.057	0.011
2017	27	0.055	0.010	0.052	0.010	0.059	0.008
Mean		**0.070**	0.016	**0.066**	0.014	**0.076**	0.016
PANEL C: LERNER INDEX							
2008	25	0.185	0.167	0.226	0.153	0.140	0.176
2009	25	0.246	0.090	0.271	0.101	0.222	0.075
2010	25	0.236	0.080	0.261	0.083	0.213	0.071
2011	25	0.209	0.083	0.240	0.063	0.161	0.090
2012	26	0.173	0.089	0.196	0.081	0.130	0.092
2013	26	0.176	0.100	0.213	0.106	0.132	0.074
2014	26	0.209	0.108	0.260	0.098	0.150	0.089
2015	26	0.215	0.106	0.274	0.101	0.147	0.064
2016	27	0.216	0.118	0.274	0.118	0.154	0.085
2017	27	0.237	0.124	0.302	0.125	0.167	0.079
Mean		**0.210**	0.107	**0.252**	0.103	**0.162**	0.090

Source: Author's computation

Figure 4.4 (see Panels A and B) show the Lerner index scores of both group 1 and 2 banks from 2008 to 2017. The mean Lerner scores of group 1 and 2 banks are different. The average of group 1 banks' Lerner score is always higher than the average group 2 banks' Lerner score and the average for all banks from 2008 to 2017 (see Figure 4.4 Panel A). The average Lerner index for group 2 banks is 0.252, which is higher than that for group 1 banks (0.162) by approximately 56% (see Figure 4.4 Panel B). This outcome implies that group 2 banks are stronger (or have greater market power) than group 1 banks. In addition, there is a less competitive environment among group 2 banks than among group 1 banks.

Figure 4.4 The Average Lerner Index for Group 1 and Group 2 Vietnam Banks from 2008 to 2017



(Panel A) **(Panel B)**

Source: Author's computation

The two-sample t-test was used to examine the statistical significance of the difference in the Lerner indexes between group 1 and 2 banks. The null hypothesis (H_{1_0}) is that the mean of difference in the Lerner indexes between groups 1 and 2 banks is zero. The t-test results are statistically significant at the 1% level (see Table 4.5). This result indicates a rejection of the null hypothesis. In other words, there is a statistically significant difference in market power between bank group 1 and 2 banks.

Table 4.5 A Two Sample T-Test of the Mean Lerner Indexes between Vietnam Bank Group 1 and 2 Banks

Two Sample t-test (Null hypothesis H_{1_0}: mean of difference = 0)	Lerner index
t-value	-6.83***
p-value	0.000

Note: *** = significance at 0.01.
Source: Author's computation

This outcome contradicts Nguyen and Nghiem's (2018) study, the only other study that compares the market power of state-owned and joint-stock Vietnam banks. Their results show that the difference in market power between commercial state and joint-stock banks is not statistically significant. The contradictory results imply that the grouping approach that categorises state and joint-stock Vietnam banks may be inappropriate when investigating differences in market power between these banks. In contrast, the SOM technique, with an unsupervised algorithm, can better capture differences in market power and categorise Vietnam domestic banks into strong banks (group 2 banks) and weak banks (group 1 banks). Group 2 banks include the four big commercial state-owned banks and several large commercial joint-stock banks. Hence, this study argues that there are two bank groups, with different levels of market power that exist separately in the Vietnam banking industry. Group 2 banks (with a mean of Lerner index value of 0.252) have greater market power than group 1 banks (with a value of 0.162). Group 2 banks also have greater total assets than group 1 banks (see Appendix Table B4).

4.3.2 Measuring the Efficiency of Group 1 and 2 Banks

4.3.2.1 Dataset

This study employs the non-parametric DEA technique under the VRS assumption to compute bank performance (or efficiency) including PTE and CE scores. Table 4.6 provides an overview of the statistical data used to compute the bank PTE and CE scores from 2008 to 2017.

Table 4.6 Summary Statistics of the Financial Data used to Compute Vietnam Bank Efficiency form 2008 to 2017

Variable			Obs	Mean	SD	Min	Max
Total Deposits	x1	Billion VND	258	128,325	189,744	1,172	1,011,314
Labour (Staff Costs)	x2	Billion VND	258	1,424	2,112	20	11,195
Fixed Assets	x3	Billion VND	258	1,823	2,272	46	11,437
Loans	y1	Billion VND	258	105,912	165,211	275	863,575
Other Earning Assets	y2	Billion VND	258	51,451	61,186	542	462,597
Price of staff cost	w1		258	0.0078	0.0028	0.0031	0.0184
Price of Deposits	w2		258	0.0851	0.0455	0.0316	0.3927
Price of Capital	w3		258	1.0500	1.3496	0.0792	17.3288

Note: Obs = observation; SD = Standard deviation. Source: Author's computation

4.3.2.2 Empirical Results and Discussion

The empirical results for the PTE and CE scores are provided in Table 4.7. The mean PTE score is 0.933 and the mean CE score is 0.855. The CE score indicates that Vietnam domestic banks could reduce their costs by 14.5% (from 100% - 85.5%) while maintaining the same outputs.

Table 4.7 PTE and CE Scores for Vietnam Banks (2008 - 2017)

year	No. of Banks	All Banks		Group 2		Group 1	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
PANEL A: PURE TECHNICAL EFFICIENCY							
2008	25	0.932	0.090	0.925	0.104	0.940	0.076
2009	25	0.963	0.056	0.957	0.070	0.970	0.043
2010	25	0.965	0.066	0.971	0.062	0.959	0.072
2011	25	0.870	0.164	0.911	0.144	0.826	0.179
2012	26	0.950	0.082	0.963	0.068	0.925	0.102
2013	26	0.945	0.089	0.943	0.088	0.948	0.095
2014	26	0.935	0.097	0.938	0.102	0.931	0.096
2015	26	0.940	0.095	0.922	0.113	0.960	0.068
2016	27	0.905	0.105	0.900	0.111	0.909	0.104
2017	27	0.928	0.093	0.917	0.107	0.940	0.076
Mean		**0.933**	0.094	**0.935**	0.097	**0.931**	0.091
PANEL B: COST EFFICIENCY							
2008	25	0.872	0.133	0.885	0.138	0.858	0.133
2009	25	0.923	0.083	0.934	0.087	0.913	0.082
2010	25	0.940	0.098	0.944	0.097	0.937	0.104
2011	25	0.732	0.253	0.786	0.254	0.675	0.249
2012	26	0.888	0.147	0.909	0.160	0.849	0.117
2013	26	0.838	0.143	0.847	0.147	0.827	0.145
2014	26	0.856	0.140	0.872	0.127	0.838	0.158
2015	26	0.882	0.130	0.866	0.134	0.901	0.127
2016	27	0.795	0.154	0.772	0.165	0.819	0.142
2017	27	0.824	0.160	0.770	0.176	0.882	0.123
Mean		**0.855**	0.144	**0.858**	0.148	**0.850**	0.138

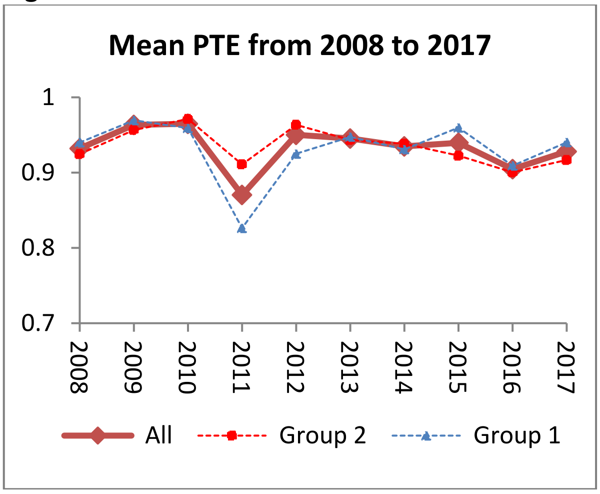
Note: SD = standard deviation.
Source: Author's computation

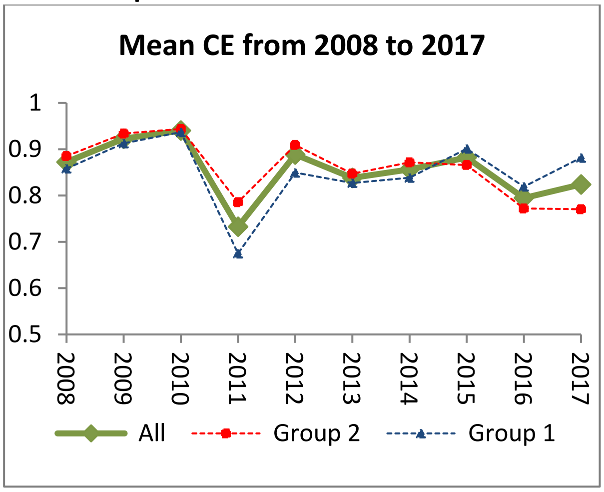
Table 4.8 Number of Efficient Vietnam Banks from 2008 to 2017

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total	%
Panel A: PURE TECHNICAL EFFICIENCY												
All banks	14	15	17	12	16	17	16	15	11	13	**146**	**57%**
Group 2	7	8	9	10	12	9	9	7	7	7	85	61%
Group 1	6	7	8	2	4	8	7	8	4	6	60	51%
Panel B: COST EFFICIENCY												
All Banks	10	10	16	10	11	8	6	10	8	9	**98**	**38%**
Group 2	6	5	8	8	10	4	3	4	4	4	56	40%
Group 1	4	5	8	2	1	4	3	6	4	5	42	36%
All Banks	25	25	25	25	26	26	26	26	27	27	**258**	
Group 2	13	12	12	16	17	14	14	14	14	14	140	
Group 1	12	13	13	9	9	12	12	12	13	13	118	

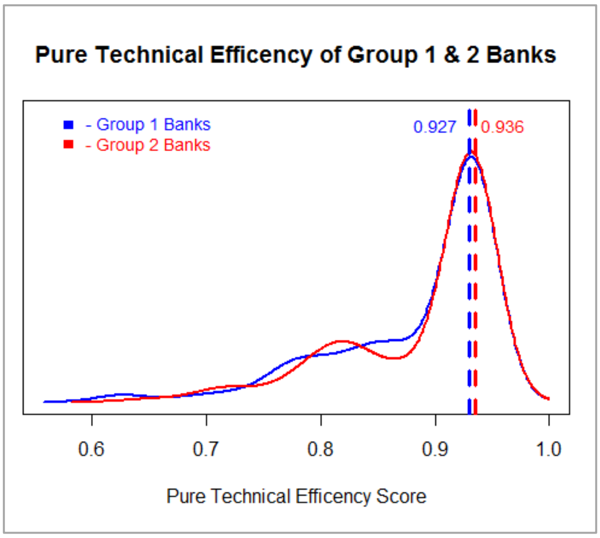
Source: Author's computation

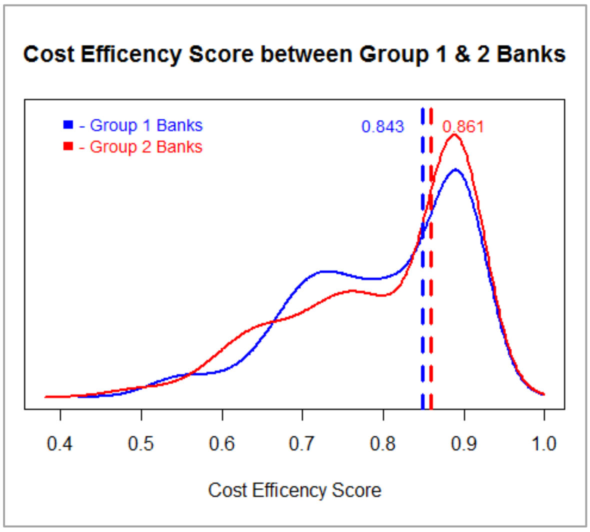
Figure 4.5 The Mean PTE and CE Scores of Vietnam's Group 1 & 2 Banks from 2008 to 2017





a) b)





c) PTE score frequency distribution d) CE score frequency distribution

Notes: (i) PTE = pure technical efficiency; CE = cost efficiency. (ii) There is a dip of PTE and CE scores in 2011. The reason is that the credit growth rate dropped from 30% to 10% from 2010 to 2011 (KPMG, 2013). This outcome indicates that domestic Vietnam banks reduced their loan amounts in 2011. Consequently, these banks decreased their cost efficiency. This is because outputs (loan amounts) decreased whereas inputs (deposit, staff cost) did not changed.
Source: Author's computation

Figure 4.5 shows the results for the PTE and CE scores for the two groups of banks. The mean PTE and CE scores for both groups of banks are quite similar. The mean of the PTE of group 2 banks is 0.935, higher than group 1 banks by only 1% (0.931) (see Figure 4.5c). In cost-efficiency, group 2 banks' CE score is 0.858; which is higher than group 1 banks' CE score (which is 1% lower, 0.850) (see Figure 4.5d). The two-sample t-test was used to examine the difference between the CE and PTE scores for the to bank groups. The null hypothesis (H_{2_0}) is that the efficiency means' (PTE and CE scores) difference between groups 1 and 2 banks equal zero. Table 4.9 shows the two-sample t-test results.

The t-value is insignificant at all conventional levels. The p-values of the paired PTE and CE tests are 0.6553 and 0.5817, respectively (see Table 4.9). These results confirm the null hypothesis. In other words, the differences in the PTE and CE scores for groups 1 and 2 banks are statistically *insignificant*. These findings echoe previous studies (Nguyen & Nghiem, 2018; Vu & Turnell, 2010). Those authors also find no significant difference between the CE scores of commercial state-owned and joint-stock banks (Nguyen & Nghiem, 2018; Vu & Turnell, 2010).

Table 4.9 The Two Sample t-Test for PTE and CE Scores between Vietnam Groups 1 & 2 Banks

Two Sample t-tests (Null hypothesis H_{2_0}: mean of difference = 0)	**Pure Technical efficiency score**	**Cost efficiency score**
t-value	-0.447(NS)	-0.551(NS)
p-value	0.6553	0.5817

Note: (NS) indicates not significant. Source: Author's computation

4.3.3 A Summary of the Vietnam Banking Industry

This uses the SOM technique to build the financial trajectories of Vietnam domestic banks. The SOM technique is used to categorise Vietnam domestic banks into two super-class bank groups (1 and 2). Group 2 banks include four commercial state-owned banks and joint-stock banks. Group 1 contains only commercial joint-stock banks. The results indicate that there are common trajectory patterns. Nineteen banks retained their positions in group 1 or 2 over the study period. In contrast, eight banks shifted from the small banks group (group 1) to bigger banks group (or group 2), and vice versa. A difference of market power (represented by the Lerner index) between the two groups of banks (1 and 2) is statistically significant at the 1% level, although their efficiency (represented by PTE and CE scores) is similar. The significantly different Lerner scores (between group 1 and 2 banks) confirms that there are two groups of banks, with different levels of market power, existing separately in the Vietnam banking industry. In other words, a group of strong banks (group 2) exists. This is a result of size concentration through mergers and acquisitions. This group of strong banks tends to be monopolists and damages the competitive atmosphere of Vietnam domestic banks (Hoang et al., 2016). The reason is that group 2 banks enjoy profitability because of higher market power whereas group 1 banks confront high competition because of lower market power.

Monopolies make it harder for smaller banks to survive. Group 2 banks (the bigger banks) will survive better in a competitive environment and will seek continued expansion (Tabak et al., 2012; Wang, 2015). In this environment, group 2 banks end up acquiring group 1 banks. This explains why the number of banks reduced from 43 to 32 between 2008 and 2017. In short, group 1 banks find it extremely difficult to survive. This group of banks struggles to cut costs because of a high level of competition and, ultimately, become a target for larger banks that are interested in M&As. This trend continues until only a small number of dominant banks with highly leveraged assets remains (Wang, 2015).

Prior studies report that bank market power tends to maintain its persistency (Adjei-Frimpong et al., 2016; Delis, 2012; Erler et al., 2017). This phenomenon leads to the existence of rigidity in the banking industry making it difficult for weak banks to compete in the banking industry. In addition, previous studies also show that bank characteristics (such as total assets, capitalisation and non-performing loans) impact directly on their market power and efficiency (Abel & le Roux, 2017; Adjei-Frimpong et al., 2014; Gardener et al., 2011; Phan et al., 2016). Therefore, this study examined the rigidity of the Vietnam banking industry and the determinants that influence bank market power and efficiency. These findings will help policy makers and regulators maintain an efficient competitive banking environment. These outcomes are also useful for bank executives who need to develop strategies to enhance their bank's power and efficiency. The investigations are discussed in the next section.

4.4 Objective 3: To Investigate The Determinants That Impact Vietnam Bank Market Power and Efficiency

4.4.1 Variable Selection and Data Description

This study investigates the determinants that influence bank market power and efficiency, including bank-specific characteristics and macroeconomic indicators. These determinants were selected based on the literature (Abel & le Roux, 2017; Anwar, 2018; Delis, 2012; Erler et al., 2017; Fred, 2019; Gardener et al., 2011; Phan et al., 2016; Soedarmono et al., 2011; Stewart et al., 2016). The relationships can be expressed by two empirical models as below:

The bank market power model is given as:

$$LER_{it} = f(BANKSIZE_{it}, NPL_{it}, CAP_{it}, GDP_t, INF_t) \qquad [4.1]$$

The bank cost efficiency model is given as:

$$CE_{it} = f(BANKSIZE_{it}, NPL_{it}, CAP_{it}, GDP_t, INF_t) \qquad [4.2]$$

Where:

i,t represent the i^{th} bank at time *t*;

CE_{it} , LER_{it},$BANKSIZE_{it}$, NPL_{it}, CAP_{it}, GDP_{it}, INF_{it} = responding and exploratory variables described in Table 4.10; and f = a regression function.

Table 4.10 The Definitions of the Responding and Exploratory Variables for the Bank Market Power and Bank Cost Efficiency Models

Variable		Definitions
Responding Variables		
LER_{it}	=	Lerner index score
CE_{it}	=	modified original cost efficiency $CE_{it}^{original}$ score
Exploratory Variables – Bank Specific Characteristics		
$BANKSIZE_{it}$	=	a proxy for bank scale (computed as natural logarithm of total assets of a bank)
NPL_{it}	=	a proxy for bank credit risk (counted as a ratio of bad loans per total outstanding loan value)
CAP_{it}	=	a proxy for capitalisation (quantified by a ratio of total owner equity per total assets)
Macroeconomic Conditions		
GDP_t	=	real annual GDP growth rate, a proxy for the Vietnam economy
INF_t	=	inflation rate, a proxy for the stability of the Vietnam economy

Source: Author's analysis with suggestions from past studies

4.4.1.1 Bank Market Power Model (Equation 4.1)

1. Responding Variable (LER_{it})

The Lerner index score (LER_{it}) represents the market power of Vietnam domestic banks, as discussed in the previous section (4.3). This Lerner index (LER_{it}) score is continuous and ranges from negative to positive values. The Lerner values are used as responding variables in equation (4.1).

2. Exploratory Variables for the Bank Market Power model

Five experimental variables are used in the bank market power model. They are described below.

Bank Size $(BANKSIZE_{it})$

The first explanatory variable, *BANKSIZE*, is calculated using a natural logarithm of total bank assets. During the study period (2008-2017), as has been discussed in section, 4.3.1, the mean Lerner score of group 2 banks was always higher than the mean for group 1 banks. In short, bigger banks are stronger (have more market power) than smaller banks. Therefore, this study argues that BANKSIZE is positively linked to the Lerner index score.

The positive relationship between BANKSIZE and LER_{it} is documented in previous studies (see for example, Delis, 2012; Fernández de Guevara et al., 2005). This is because big banks can access cheaper financial resources, deal better with moral hazards, and are able to limit competition policy to enhance their market power (Delis, 2012). In contrast, Fernández De Guevara and Maudos (2007) show a non-linear negative relationship between BANKSIZE and LER_{it}. The authors explain that medium-sized banks cannot compete with smaller banks because of the extensive branch networks of local banks acting as entry barriers (Fernández De Guevara & Maudos, 2007). As a result, medium-sized banks are weaker (lower market power) than smaller domestic banks. However, bigger banks have the highest market power because they hold the dominant position (Fernández De Guevara & Maudos, 2007).

Capitalisation Ratio (CAP_{it})

The second bank-specific characteristic explanatory variable, capitalisation ratio CAP_{it}, is measured as a ratio: total owner equity per total assets. Well-capitalised banks are safer (from a customer's perspective), because of their strength and reputation (Simpasa, 2013). These banks can attract more borrowers and have greater ability to negotiate better loan prices. Therefore, these banks can offer higher loan prices to their borrowers (Abel & le Roux, 2017). In addition, well-capitalised banks have greater financial resources that allow them to offer loans to higher risk borrowers. Thus, these banks can expand their lending, enjoy high interest margins and ultimately increase their market power (Abel & le Roux, 2017; Erler et al., 2017). As a result, a positive association between CAP_{it} and LER_{it} is expected and is supported by previous studies (see for example, Abel & le Roux, 2017; Delis, 2012; Erler et al., 2017; Simpasa, 2013). In contrast, a negative relationship between CAP_{it} and LER_{it} is documented by Adjei-Frimpong et al. (2016). The authors explain that banks must pay higher interest rates to depositors to compete with rates offered by government securities (Adjei-Frimpong et al., 2016).

Non-performance Loans Ratio (NPL_{it})

NPL is the major problem in the Vietnamese banking industry (KPMG, 2013). However, no investigation has reported on the relationship between NPL_{it} and LER_{it} for Vietnam domestic banks. Therefore, this study investigated the influence of NPL on Vietnam banks' market power. A higher interest rate charged can contribute to increased bank profitability. However, this may lead to a deterioration in a bank's loan portfolio (Abel & le Roux, 2017). In such circumstances, a bank can reduce the total number of loans and seek low-risk customers (Abel & le Roux, 2017). Banks cannot set higher loan prices because they must compete to attract quality borrowers. This situation implies a decrease in market power. In short, we anticipate a *negative* relationship between NPL and bank market power. Previous studies have reported that the NPL ratio has a negative relationship with bank market power (see for example, Abel & le Roux, 2017; Simpasa, 2013). Moreover, studies show a negative insignificant link between the NPL ratio and LER_{it} (Adjei-Frimpong et al., 2016; Fernández De Guevara & Maudos, 2007).

GDP Growth Rate (GDP_t)

The first macroeconomic explanatory variable, GDP_t rate per year, represents economic growth. Economic development increases customers' incomes (Erler et al., 2017). The bank may gain positively because of an increase in loan demand and banking services (Erler et al., 2017; Fernández de Guevara et al., 2005). This means that banks can set higher prices for their loan products (Abel & le Roux, 2017). In short, we expect to find a positive link between GDP_t and LER_{it}, as per previous studies (Abel & le Roux, 2017; Erler et al., 2017; Fernández de Guevara et al., 2005).

Inflation Rate (INF_t)

In contrast to GDP rate, the INF_t rate is projected to have a negative relationship with market power. Higher INF_t rates cause higher living costs (Simpasa, 2013). Banks may face negative effects because of lessening loan demand (Erler et al., 2017). As a result, banks may not be able to command a higher price on their financial products, which may also denote a decrease in bank market power (Erler et al., 2017). Though Angelini and Cetorelli (2003) find a significant negative link between INF_t and LER_{it}, Abel and le Roux (2017) and Erler et al. (2017) find an insignificant negative relationship. Panel A in Table 4.11 provides an overview of the relationships between the five explanatory variables $(BANKSIZE_{it}, CAP_{it}, NPL_{it}, GDP_t, INF_t)$ and bank market power (LER_{it}) as reported in the literature.

4.4.1.2 Bank Cost Efficiency Model (Equation 4.2)

1. Responding Variable (CE_{it})

The original cost efficiency $(CE_{it}^{Orginal})$ score discussed in section 4.3.2, ranges from zero to one. However, linear regression estimates (the right-hand side of equation 4.1) do not fall in the same range (from zero to one). Therefore, the $CE_{it}^{Orginal}$ score is modified so that it is now an unrestricted value based on the logit transformation. This is expressed mathematically as follows (Adjei-Frimpong et al., 2014; Cox, 1970; Kader, Adams, & Hardwick, 2010):

$$CE_{it} = \ln\left(\frac{CE_{it}^{Orginal}}{1 - CE_{it}^{Orginal}}\right) \qquad [4.3]$$

Where - $CE_{it}^{Orginal}$ = the original CE score, computed in the previous section (4.3.2).

The modified value of the CE_{it} score in equation (4.3) ranges from negative infinity to positive infinity. However, CE_{it} will be indefinite when the $CE_{it}^{Orginal}$ equals one. This will result in dropping observations and the loss of information. Hence, the value 1/2N (where N equals observations in the efficiency frontier) is added to both the numerator and denominator in equation (4.3) to avoid the loss of information (Adjei-Frimpong et al., 2014; Cox, 1970; Kader et al., 2010)

After adding the 1/2N value, there is no loss of observations when the $CE_{it}^{Orginal}$ score equals one. As a result, the modified CE_{it} can be used as a dependent variable in equation (4.2) to explore the impact of bank characteristics and the macroeconomic environment on bank efficiency. However, one shortcoming of inserting the 1/2N value is that we cannot compute the precise marginal effect of these exploratory variables on bank efficiency. As a result of this choice, this study reports only the sign and statistically significant explanatory determinants.

2. The Exploratory Variables of the Bank Cost Efficiency Model

Five predictor variables are used in the bank cost efficiency model (Equation 4.2)

Bank Size $(BANKSIZE_{it})$

The $BANKSIZE_{it}$ regressor is computed using the natural logarithm of total assets. This variable determines the asset value that identifies efficient and inefficient banks (Ataullah & Le, 2006; Delis & Papanikolaou, 2009). The literature shows mixed outcomes for the nexus between $BANKSIZE_{it}$ and CE_{it}. Several studies have found that $BANKSIZE_{it}$ is *positively* statistically significant with CE_{it} (Anwar, 2018; Phan et al., 2016; Stewart et al., 2016). A positive relationship signifies that the bank operates efficiently because of economy of scale and benefits in terms of increased profit when its size increases (Anwar, 2018). However, when banks become extremely large, the link between $BANKSIZE_{it}$ and CE_{it} may become *negative* (Delis & Papanikolaou, 2009). A negative relationship between these variables indicates that big banks are less efficient. This is because banks suffer from diseconomy of scale and are inefficient in spreading their operational costs because of bureaucracy (Al-Gasaymeh, 2016; Delis & Papanikolaou, 2009). In Vietnam domestic banks, the relationship between $BANKSIZE_{it}$ and CE_{it} is unclear. The positive connection has been documented in previous studies (Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016) but Gardener et al. (2011) report a negative relationship. Therefore, no expectation is suggested for this relationship.

Capitalisation Ratio (CAP_{it})

A well-capitalised (or high CAP_{it} ratio) bank is believed to be strong, have a good reputation and be safe from a customer's perspective (Simpasa, 2013). These banks can pay depositors a lower interest rate (Abel & le Roux, 2017). In short, the bank is supposed to be more efficient as a result of spending less on costs. For this reason, CAP_{it} is expected to have a positive impact on CE_{it}. This positive link is reinforced by previous studies (Anwar, 2018; Fiordelisi, Marques-Ibanez & Molyneux, 2011; Gardener et al., 2011). However, some studies show a negative relationship between CAP_{it} and CE_{it} (Adjei-Frimpong et al., 2014; Altunbas, Carbo, Gardener, & Molyneux, 2007; Kablan, 2010). A negative link indicates that the bank suffers higher costs when raising capital through issuing more shares than taking depositor funds (Ariff & Can, 2008). Gardener et al. (2011) and Nguyen and Nghiem (2018) report a positive association between CAP_{it} and CE_{it} in Vietnam banks.

Non-performance Loans Ratio (NPL_{it})

The NPL_{it} ratio reflects a bank's asset quality (Delis & Papanikolaou, 2009; Simpasa, 2013). A high NPL_{it} may be caused by weak or ineffective evaluating, monitoring and supervising systems (Nguyen, 2018). An efficient bank is anticipated to have an effective credit rating system to prevent failure and have a lower NPL_{it} ratio. Hence, a negative relationship between NPL_{it} and CE_{it} is predicted. Previous studies present a negative relationship between NPL_{it} and CE_{it} (Anwar, 2018; Kablan, 2010; Staub et al., 2010). This negative relationship indicates that banks had higher costs because of additional monitoring and loan repayment enforcement (Havrylchyk, 2006). Some authors note a *negatively insignificant* link between NPL_{it} and CE_{it} in Vietnam domestic banks (Phan et al., 2016; Stewart et al., 2016).

GDP Growth Rate (GDP_{it})

The GDP growth rate reflects national economic development (Gardener et al., 2011). Economic growth contributes to higher incomes and lower borrower default rates (Anwar, 2018; Erler et al., 2017). Hence, banks can reduce losses from bad debts and reserve lower provision for credit loss; these enhance banks' cost efficiency (Anwar, 2018). Therefore, we predict a positive relationship between GDP_{it} and CE_{it}. Previous literature has reported that GDP_{it} impacts positively on bank cost-efficiency in various countries such as Bangladesh, Indonesia, Malaysia, the Philippines and Vietnam (Anwar, 2018; Gardener et al., 2011).

Inflation Rate (INF_{it})

The inflation rate reveals the stability of an economy (Phan et al., 2016). A higher inflation rate might lead to higher financing costs, forcing banks to pay more interest to attract depositors, and an increase bad debts (Anwar, 2018; Chan & Karim, 2010). Banks also experience increased operating costs, such as administration, salaries, and employee benefits that result in lower efficiency (Anwar, 2018). In short, we predict a negative relationship. However, a positive relationship between INF_{it} and CE_{it} can occur if banks charge higher lending interest rates to compensate for the losses incurred as a result of bad borrowers (defaulting on payments) (Chan & Karim, 2010). A negative relationship between INF_{it} and CE_{it} has been documented for Indonesian banks (Anwar, 2018). Phan et al. (2016) note a mixed negative/positive link between INF_{it} and CE_{it}. The authors conclude there is an unclear relationship between INF_{it} and CE_{it} in various countries, such as Bangladesh, India, Indonesia, Malaysia, the Philippines and Vietnam. Table 4.11 summarises the expected signs of the exploratory variables on bank market power (Panel A) and cost efficiency (Panel B).

Table 4.11 The Expected Sign of the Independent Variables on Bank Efficiency and Market Power

Independent Variable	Coefficient sign	Source
PANEL A –Bank Market Power Model LER_{it}		
$BANKSIZE_{it}$	+	Denis (2012), Fernandez de Guevara et al. (2005)
	-/+	Fernandez de Guevara et al., (2007); Erler et al. (2017)
NPL_{it}	-	Simpasa (2013), Abel and le Roux (2017)
CAP_{it}	-	Adjei-Frimpong et al. (2014)
	+	Abel and le Roux (2017), Denis (2012), Erler et al. (2017), Simpasa (2013)
GDP_{it}	+	Erler et al. (2017), Fernandez de Guevara et al. (2005), Abel & le Roux (2017)
INF_{it}	-	Angelin and Cetorelli (2003)
PANEL B –Bank Cost Efficiency Model CE_{it}		
$BANKSIZE_{it}$	+	Ataullah and Le (2006), Anwar (2018), Delis and Papanikolaou (2009), Stewart et al. (2016), Phan et al. (2016), Nguyen & Nghiem (2018)
	-	Gardener et al. (2011), Al-Gasaymeh (2016)
NPL_{it}	-	Anwar (2018), Delis and Papanikolaou (2009), Kablan (2010), Staub et al. (2010),
CAP_{it}	-	Altunbas et al. (2007), Adjei-Frimpong et al. (2014), Kablan (2010)
	+	Anwar (2018), Fiordelisi et al. (2011), Gardener et al. (2011), Nguyen & Nghiem (2018)
GDP_{it}	+	Anwar (2018), Gardener et al. (2011), Phan et al. (2016)
INF_{it}	-/+	Anwar (2018), Phan et al. (2016)

Source: Author's summary of previous studies

4.4.1.3 Hypothesis developments

Prior studies indicate that big banks can access cheaper financial resources, deal better with moral hazards, and are able to limit competition to enhance their market power (Delis, 2012). In addition, bigger bank operates more efficiently because of economy of scale and benefits in terms of increased profit when its size increases (Anwar, 2018). Therefore, we hypothesise the following relationship:

H_{3_0}: Bank size positive and significantly impact bank market power

H_{4_0}: Bank size positive and significantly impact bank cost efficiency

Well-capitalised banks have greater financial resources that allow them to offer loans to higher risk borrowers. Thus, these banks can expand their lending, enjoy high interest margins and ultimately increase their market power (Abel & le Roux, 2017; Erler et al., 2017). These well-capitalised banks can pay depositors a lower interest rate because of their reputations (Abel & le Roux, 2017). In short, the banks are more efficient as a result of spending less on costs. Therefore, we hypothesise the following relationship:

H_{5_0}: There is a positive effect of capitalisation on bank market power.

H_{6_0}: There is a positive effect of capitalisation on bank efficiency.

When there is a deterioration in a bank's loan portfolio, the bank can reduce the total number of loans and seek low-risk customers (Abel & le Roux, 2017). Banks cannot set higher loan prices because they must compete to attract quality borrowers. Moreover, higher NPL_{it} ratio reflects lower bank's asset quality. Banks had higher costs because of additional monitoring and loan repayment enforcement (Havrylchyk, 2006). Therefore, we hypothesise the following relationship:

H_{7_0}: Non-performance loans ratio negatively significantly impact market power

H_{8_0}: Non-performance loans ratio negatively significant impact bank efficiency

Economic development increases loan demand and banking services (Erler et al., 2017; Fernández de Guevara et al., 2005). Banks have power to set higher prices for their loan products (Abel & le Roux, 2017). In addition, banks can reduce losses from bad debts and reserve lower provision for credit losses; these enhance banks' cost efficiency (Anwar, 2018). Hence, we hypothesise the following relationship:

H_{9_0}: There is a positive relationship between GDP growth rate and bank market power

H_{10_0}: There is a positive relationship between GDP growth rate and bank cost efficiency

Banks may face negative effects because of decrease in loan demand under inflationary environment (Erler et al., 2017). They cannot command a higher price on financial products and reduce market power (Erler et al., 2017). Moreover, these banks also experience increased operating costs, such as administration, salaries, and employee benefits that result in lower efficiency (Anwar, 2018). Therefore, we hypothesise the following relationship:

H_{11_0}: Inflation rate negative significantly impact bank market power.

H_{12_0}: Inflation rate negative significantly impact bank efficiency.

These hypothesises will be tested in the next section.

4.4.1.4 Data Information

Table 4.12 shows the financial data of Vietnam domestic banks from 2008 to 2017. The average capitalisation ratio is 10%. The non-performance loan average ratio is 2.25%, which is smaller than the 3% required by the state bank. Macroeconomic factors, such as the GPD growth rate, have an average value of 6%. The mean value of the inflation rate is 8.5% over the 10-year period. In summary, bank-specific and macroeconomic factors indicate that the Vietnam domestic banks have enjoyed stable development as a result of economic growth over the study period (from 2008 to 2017).

Table 4.12 A Summary of Vietnam Bank Financial Data 2008 – 2017

Variable	Observations	Mean	S.D.	Min	Max
CE_{it}	258	1.882	1.099	-0.442	3.497
LER_{it}	258	0.210	0.110	-0.041	0.616
$BANKSIZE_{it}$	258	11.302	1.268	7.791	14.000
CAP_{it}	258	10.12%	5.82%	3.50%	46.24%
NPL_{it}	258	2.25%	1.50%	0.00%	11.40%
GDP_{it}	258	6.01%	0.53%	5.25%	6.81%
INF_{it}	258	8.47%	6.71%	0.88%	23.11%

Note: S.D. = Standard Deviation.
Source: Author's computation.

4.4.2 Statistical Analysis

Analysis begins by testing non-stationarity and correlation in the dataset.

4.4.2.1 Stationarity Test

We use unbalanced panel data that contain 258 observations from 2008 to 2017. Hence, the stationarity test is used to avoid spurious regression when applying the pooled OLS regression estimation (Gujarati & Porter, 2009). If the dataset is non-stationarity, it may produce spurious significant results and be totally meaningless (Gujarati & Porter, 2009). A stationarity test for an unbalanced panel dataset (Fisher-type) is used (Choi, 2001). The null of the non-stationarity hypothesis (H_0) is assumed. The non-stationarity findings are shown in Table 4.13.

Table 4.13 The Results of Non-Stationarity Tests (Fisher-Type) of the Data

Fisher-type Unit-Root Test	Cost efficiency score CE_{it}	Lerner index score LER_{it}
Inverse chi-squared (48)	83.92***	70.93**
p-value	0.001	0.0173

Note: ***, ** denotes significance at 1% and 5% levels, respectively.
Source: Author's computation

The p-values of the non-stationarity tests are statistically significant at all levels (1% to 5%). Hence, the null hypothesis (H_0) is rejected. In other words, there is no unit root in our dataset or stationarity data. Therefore, it is possible to apply conventional regression models, such as pooled ordinary least squares (OLS) regression (Gujarati & Porter, 2009).

4.4.2.2 Correlation Test

The Pearson correlation test is used to assess the correlation between the dependent and independent variables. This test identifies variables that are highly correlated (above 0.8). Highly correlated variables may contain the same information and imply the presence of a serious multicollinearity problem (Gujarati & Porter, 2009). The variance-inflating factor (VIF) is calculated to confirm the existence of a multicollinearity problem. The Pearson pairwise correlation results are shown in Table 4.14. The results show that all correlation ratios are less than 0.8.

Table 4.14 The Pearson Correlation Test Results of the Vietnam Bank Data

	$\boldsymbol{CE_{it}}$	$\boldsymbol{LER_{it}}$	$\boldsymbol{BANKSIZE}$	$\boldsymbol{CAP_{it}}$	$\boldsymbol{NPL_{it}}$	$\boldsymbol{GDP_{it}}$	$\boldsymbol{INF_{it}}$
CE_{it}	1						
LER_{it}	0.26*	1					
$BANKSIZE_{it}$	0.12*	0.44*	1				
CAP_{it}	-0.08	-0.21*	0.01	1			
NPL_{it}	0.01	-0.07	-0.71*	0.01	1		
GDP_{it}	-0.05	0.13*	0.24*	-0.23*	-0.21*	1	
INF_{it}	0.01	-0.08	-0.33*	0.06	0.33*	-0.34*	1

Note: * Significance at 5% level. Source: Author's computation

Table 4.15 presents the VIF values for each independent variable. All VIF scores are less than 10. This confirms there is no multicollinearity problem in the dataset (Adjei-Frimpong et al., 2014; Phan et al., 2016).

Table 4.15 The VIF Test Results of Vietnam Bank Data

	VIF	1/VIF (Tolerance)
$BANKSIZE_{it}$	2.10	0.47
CAP_{it}	2.08	0.48
NPL_{it}	1.24	0.80
GDP_{it}	1.22	0.82
INF_{it}	1.06	0.94
Mean VIF	1.54	

Source: Author's computation

4.4.2.3 Pooled OLS Regression Model and Results

The results for both the non-stationarity (or unit root) and correlation tests in the previous section confirm the validity of conventional estimation techniques. Hence, pooled OLS regression is applied to investigate the factors that impact bank market power and efficiency. Equations (4.1) and (4.2) are rewritten as:

The bank market power model is:

$$LER_{it} = \alpha_0 + \beta_1 BANKSIZE_{it} + \beta_2 CAP_{it} + \beta_3 NPL_{it} + \beta_4 GDP_{it} + \beta_5 INF_{it} + \varepsilon_{it} \qquad [4.4]$$

The bank cost efficiency model is:

$$CE_{it} = \alpha_0 + \beta_1 BANKSIZE_{it} + \beta_2 CAP_{it} + \beta_3 NPL_{it} + \beta_4 GDP_{it} + \beta_5 INF_{it} + \varepsilon_{it} \qquad [4.5]$$

Where:

I represents the i_{th} bank at time *t*;

$\alpha_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 =$ coefficients to be estimated;

ε_{it}= error term;

$CE_{it}, LER_{it}, BANKSIZE_{it}, CAP_{it}, NPL_{it}, GDP_{it}, INF_{it}$ are the exploratory variables (see Table 4.10).

All independent variables are assumed to be strictly exogenous in equations (4.4) & (4.5). The OLS estimated parameter results are displayed in Table 4.16.

Table 4.16 The Pooled OLS Results of Determinants Impacting on LER and CE Scores

Independent Variable	**Market Power Model LER_{it}**		**Cost Efficiency Model CE_{it}**	
$BANKSIZE_{it}$	0.07	***	0.25	***
	(0.000)		(0.001)	
NPL_{it}	-1.68	***	-8.26	*
	(0.000)		(0.076)	
CAP_{it}	0.94	***	3.54	**
	(0.000)		(0.035)	
GDP_{it}	0.02		-20.36	
	(0.985)		(0.149)	
INF_{it}	0.05		0.30	
	(0.552)		(0.787)	
Constants	-0.64	***	0.07	
	(0.000)		(0.956)	
R^2value	0.3605		0.0518	
Adjusted R^2	0.3478		0.033	
p-value of F-Statistic test	0.0000		0.0193	
Total observations	258		258	
Total banks	27		27	

Note: The p-value is contained in the parentheses; * **, ** and * indicate significance at 1%, 5% and 10%, respectively.

Source: Author's computation

The $BANKSIZE_{it}$ variable has a positive significant relationship with bank market power and cost efficiency at the 1% level. The positive relationship between $BANKSIZE_{it}$ and LER_{it} indicates that large banks have higher market power. This is because large banks have greater opportunities to obtain cheaper financial resources through economies of scale, are better at dealing with moral hazard issues (borrowers), and can limit competition policy to improve bank market power (Delis, 2012). This outcome shows that Vietnam domestic banks are the same as banks in European countries as outlined in the literature (Delis, 2012; Fernández de Guevara et al., 2005). Additionally, the positive relationship between $BANKSIZE_{it}$ and CE_{it} reflects that Vietnam domestic banks enjoy economies of scale and profitability (Anwar, 2018). This finding echoes the literature (Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016). These outcomes suggest that Vietnam domestic banks with greater total assets will have higher market power and enjoy greater economies of scale (and/or be more efficient).

The capitalisation ratio (CAP_{it}) has a positive significant relationship with bank market power at the 1% level and cost efficiency at the 5% level. The positive relationship between CAP_{it} and LER_{it} indicates that high capitalisation enhances a bank's reputation (Simpasa, 2013). Banks can attract more borrowers and have more chances to offer better loan prices (Abel & le Roux, 2017). In short, capitalisation enhances bank market power. This positive relationship shows that Vietnam domestic banks are like European and Zambian banks (Abel & le Roux, 2017; Delis, 2012; Simpasa, 2013). The positive link between CAP_{it} and CE_{it} implies that using owner equity is cheaper than depositor funds (Adjei-Frimpong et al., 2014; Ariff & Can, 2008). This finding agrees with previous studies (Gardener et al., 2011; Nguyen & Nghiem, 2018). The findings suggest that CAP_{it} is an important determinant in upgrading both the market power and efficiency of Vietnam banks.

The non-performance loan (NPL_{it}) ratio has a negative significant relationship with LER_{it} at the 1% level and CE_{it} at the 10% level. The negative relationship between NPL_{it} and LER_{it} indicates that a higher NPL may harm a bank's strength and reputation and reduce its market power (Abel & le Roux, 2017). Banks are unable to set higher loan prices because they must compete to attract quality borrowers (Abel & le Roux, 2017). This result is similar to prior studies (Abel & le Roux, 2017; Simpasa, 2013) that report a negative relationship between NPL_{it} and LER_{it} for Zimbabwean commercial banks. A higher NPL_{it} ratio damages bank efficiency. This is because the higher the NPL ratio, the greater the likelihood of bad debts. This means that Vietnam domestic banks must pay additional monitoring and enforcement costs and thus they become less efficient. This finding echoes previous studies that show a negative relationship between NPL and *CE* (Anwar, 2018; Delis & Papanikolaou, 2009; Staub et al., 2010).

Macroeconomic conditions (GDP_{it}/INF_{it}) have no impact on the banking industry based on the insignificant LER_{it}/CE_{it} variable at all conventional levels. These findings are similar to prior studies that have found the relationship between GDP_{it}/INF_{it} and LER_{it}/CE_{it} are insignificant (Adjei-Frimpong et al., 2016; Gardener et al., 2011; Nguyen & Nghiem, 2018).

The major problem with using the pooled OLS estimator is that the regression cannot distinguish between banks or time. More specifically, the OLS estimator does not show if the determinants impacting on bank market power and efficiency are similar or dissimilar in terms of banks or time. In other words, the pooled OLS estimator may suffer from a heterogeneity problem because this method does not distinguish between various banks at different times (Gujarati & Porter, 2009). The fixed effect estimator is used to overcome this major problem because that regression allows individual banks to have their own intercepts. The next section discusses the results of the fixed effect estimation.

4.4.2.4 The Fixed-effect Regression Model and Results

Fixed-effect estimation is used to explore factors that impact bank market power and efficiency. This regression overcomes the problem of OLS estimation, where individual bank effects are lumped into the error term. Equations (4.1) and (4.2) are rewritten as:

The bank market power model:

$$LER_{it} = \alpha_0 + \beta_1 BANKSIZE_{it} + \beta_2 CAP_{it} + \beta_3 NPL_{it} + \beta_4 GDP_{it} + \beta_5 INF_{it} + \eta_i + \varepsilon_{it} \qquad [4.6]$$

The bank cost efficiency model:

$$CE_{it} = \alpha_0 + \beta_1 BANKSIZE_{it} + \beta_2 CAP_{it} + \beta_3 NPL_{it} + \beta_4 GDP_{it} + \beta_5 INF_{it} + \eta_i + \varepsilon_{it} \qquad [4.7]$$

Where:

i represents the i_{th} bank at time *t*;

$\alpha_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 =$ coefficients to be estimated;

ε_{it}= error term;

η_i = individual bank effect;

$CE_{it}, LER_{it}, BANKSIZE_{it}, CAP_{it}, NPL_{it}, GDP_{it}, INF_{it}$ are the exploratory variables (see Table 4.10).

All the independent variables are assumed to be strictly exogenous in equations (4.6) and (4.7). The fixed-effect estimated results are displayed in Table 4.17

Table 4.17 shows the factors that influence bank market power and cost-efficiency calculated using the fixed effect regression. The results estimated using fixed effect regression are different from the predicted results estimated by the pooled OLS regressor. $BANKSIZE_{it}$ is insignificant to LER_{it} and is negatively significant to CE_{it} at the 1% level. The negative relationship between $BANKSIZE_{it}$ and CE_{it} indicates that smaller banks are more efficient than larger banks. This outcome implies that Vietnam domestic banks experience diseconomies of scale (Al-Gasaymeh, 2016; Gardener et al., 2011). Large banks tend to have higher operating costs than small banks because of bureaucracy. This reduces their cost efficiency (Delis & Papanikolaou, 2009; Gardener et al., 2011). This outcome also reveals that small banks struggle to improve their efficiency when competing against large banks (Al-Gasaymeh, 2016). This finding echoes in previous studies on Vietnam domestic banks (Gardener et al., 2011). Large $BANKSIZE_{it}$ decreases cost efficiency but does not affect bank market power.

Table 4.17 The Fixed Effect Estimation Outcomes of the Determinants Impacting LER and CE

Independent Variable	Market Power Model LER_{it}		Cost Efficiency Model CE_{it}	
$BANKSIZE_{it}$	0.016		-0.501	***
	(0.164)		(0.001)	
CAP_{it}	0.483	**	-0.378	
	(0.000)		(0.811)	
NPL_{it}	-1.112	***	-5.096	
	(0.000)		(0.221)	
GDP_{it}	1.831	**	3.674	
	(0.044)		(0.763)	
INF_{it}	-0.163	**	-2.836	**
	(0.044)		(0.010)	
Constants	-0.087		7.721	***
	(0.491)		(0.000)	
R^2 value	0.178		0.076	
Adjusted R^2	0.065		0.050	
p-value of F-Statistic test	0.000		0.003	
Total observations	258		258	
Total banks	27		27	

Note: p-value is contained in the parentheses; * **, ** and * indicate significance at 1%, 5% and 10%, respectively.

Source: Author's computation

Well-capitalised (high CAP_{it} ratio) banks enhance bank market power and do not influence cost-efficiency. A high NPL_{it} has a negative significant relationship with LER_{it} and it does not impact on CE_{it}. These results suggest that CAP_{it} is a critical determinant in enhancing a bank's strength and reputation. A high NPL_{it} ratio indicates the erosion of bank market power. However, both the CAP_{it} and NPL_{it} ratios do not affect CE_{it}. These results echo previous studies that show a positive association between CAP_{it} and LER_{it} (Abel & le Roux, 2017; Delis, 2012; Simpasa, 2013) and a negative correlation between NPL_{it} and LER_{it}(Abel & le Roux, 2017; Simpasa, 2013).

GDP_{it} growth rate has a statistically significant association with LER_{it} at the 5% level. This indicates that economic growth contributes to increased bank market power. This is because customers' income increases during periods of economic development; Vietnam domestic banks experienced an associated increase in loan demand and banking services. As the number of loans increases, banks can offer better prices (Erler et al., 2017; Fernández de Guevara et al., 2005). This finding reveals that

Vietnam bank market power is influenced by economic growth like banks in the US, Europe and Zambia (Abel & le Roux, 2017; Erler et al., 2017; Fernández de Guevara et al., 2005).

The INF_{it} has a negatively significant relationship with LER_{it} and CE_{it} at the 5% level. These outcomes reveal that INF_{it} reduces both bank market power and efficiency. In times of higher inflation, people tend to reduce their living costs (Erler et al., 2017). As the number of loans decreases, banks cannot set higher loan prices and may also incur unanticipated losses (Abel & le Roux, 2017). In other words, this finding implies that the inflationary environment decreases Vietnam banks' market power. This is the same as Angelini and Cetorelli's (2003) finding for Italian banks. They find a negative relationship between *INF* and *LER* (Angelini & Cetorelli, 2003). High inflation costs mean that banks suffer higher financial costs because of increased depositor interest, more reserves for bad debts because of increased default rate and operating costs, including salaries and/or employee benefits (Anwar, 2018; Erler et al., 2017). These additional costs reduce a bank's efficiency. The negative association between INF_{it} and CE_{it} in Vietnam domestic banks is like Indonesian banks that shows Indonesian banks are less cost-efficient in an inflationary environment (Anwar, 2018).

The above findings show that fixed effect regression overcomes the heterogeneity problem and generates different results. There are two assumptions associated with conventional estimators (pooled OLS or fixed effect) to ensure they remain a Best Linear Unbiased Estimator. First, the error term must be constant over time. If the error term is not constant, the pooled OLS, or fixed effect estimators, have *heteroscedasticity* problems. Second, the disturbance term of one observation cannot be correlated with the disturbance term of others in time and space. If there is a correlation in the disturbance term, then *autocorrelation* exists. The presence of heteroscedasticity and autocorrelation means that the OLS estimate is no longer the best linear unbiased estimator (BLUE) (Baltagi, 2008; Gujarati & Porter, 2009). These two assumptions are checked in the next section.

4.4.2.5 Heteroscedasticity Test

The assumption of both the OLS and fixed effect estimators is that the error or disturbance term u_i is constant over time (Gujarati and Porter, 2009). This can be expressed as:

$$E\left(u_i^2\right) = \sigma^2 \quad where\ i = 1,2,\dots,n \tag{4.8}$$

In conditional variances, u_i is not constant and heteroscedasticity is present. There are many sources of heteroscedasticity. Varying bank size (small or big assets) could contribute to differences in the error term and cause heteroscedasticity (Gujarati & Porter, 2009). The modified Wald test is used to assess the presence of heteroscedasticity in the fixed effect estimation and the Breusch-Pagan/Cook-Weisberg test is used to check for heteroscedasticity in the pooled OLS estimator. The null hypothesis (H_0) is homoscedasticity (or constant variance) in the error term.. The results are shown in Table 4.18.

Except for the Breusch-Pagan/Cook-Weisberg test for the bank market power model, Table 4.18 shows that the p-value is statistically significant at all conventional significance levels. Hence, the null hypothesis can be rejected. This means that there is heteroscedasticity in our dataset. As a result, these estimators (the pooled OLS or fixed effect) are *inefficient* and not BLUE because their variance is no longer within the minimum range (Baltagi, 2008).

Table 4.18 The Heteroscedasticity Results (OLS And Fixed Effect Estimation) for Vietnam Banks

	Cost Efficiency Model CE_{it}	**Market Power Model** LER_{it}
Breusch-Pagan/Cook-Weisberg Test	2.82	0.03
p-value	0.093	0.863
Modified Wald Test	2.2E+31	753.7
p-value	0.000	0.000

Source: Author's computation

4.4.2.6 Autocorrelation Test

The second assumption of the OLS or fixed effect estimators is that a correlation in the disturbance term u_i between observations cannot exist in time and space (Gujarati & Porter, 2009). This can be expressed as:

$$cov(u_i, u_j | x_i, x_j) = E(u_i, u_j) = 0 \quad where\ i \neq j \qquad [4.9]$$

The Wooldridge autocorrelation test (2002) is used in this study because it can manage unbalanced panel data with gaps in the observations (Drukker, 2003). The test results are shown in Table 4.19.

Table 4.19 The Autocorrelation Results (OLS and Fixed Effect Estimators) for Vietnam Banks

	Cost Efficiency Model CE_{it}	**Market Power Model** LER_{it}
Wooldridge (2002) Test for Autocorrelation	8.150***	13.257***
p-values	0.009	0.001

Note: *** indicate significance at 1% level.
Source: Author's calculations

The p-value of the Wooldridge test is statistically significant at the 1% level (Table 4.19). Hence, the null hypothesis is rejected; autocorrelation is present in the dataset. Therefore, applying classical linear regression, such as OLS or fixed effect estimators, will result in inefficient results. These estimators are no longer BLUE and the t, F, and χ^2 are not valid (Gujarati & Porter, 2009).

4.4.2.7 Summary of the Statistical Analysis

The statistical tests identified the presence of heteroscedasticity and autocorrelation in the dataset. These situations violate the assumptions of conventional estimators such as constant and uncorrelated disturbance term in time and space. Therefore, the OLS and fixed effect estimators are considered inefficient if the violated assumptions are not eliminated (Gujarati & Porter, 2009). In other words, these estimators are no longer BLUE.

There is also another major problem, namely *endogeneity*, which may violate the assumptions of classical linear regression estimators. There is the potential for endogeneity to cause simultaneity between bank market power or efficiency and their past values, as well as the effect of the determinants on bank market power or efficiency and vice versa. For endogeneity, the OLS or fixed effects are considered biased, inconsistent and inefficient. An instrumented estimator, such as the generalised method of moments, is appropriate to handle endogeneity (Antoniou et al., 2008; Wintoki et al., 2012; Zeitun et al., 2017). *Endogeneity* is discussed in the next section.

4.4.3 Dynamic Analysis

4.4.3.1 Dynamic Relationships of Bank Market Power and Efficiency

Previous studies have reported that banks can continually maintain profits and market power over time (Adjei-Frimpong et al., 2016; Berger, Bonime, Covitz, & Hancock, 2000; Delis, 2012; Erler et al., 2017). Banks cannot continuously sustain high-performance in a market where their market power is non-existent (Berger et al., 2000). The reason is that competitors quickly enter the market, duplicate their strategies, and provide lower prices for similar financial products. These factors negatively affect the pre-existing bank's profit. Non-transparent information, lending relationships and impediments to competition are sources banks use to create their market power and generate constant profit (Berger et al., 2000; Delis, 2012). Therefore, banks often maintain their market power. This relationship can be expressed with one-year lagged bank market power as a regressor that is positively significant in relation to current market power. As a result, a dynamic relationship of market power may exist in Vietnam domestic banks.

Previous studies have reported that banks can maintain their cost efficiency over time (Adjei-Frimpong et al., 2014; Al-Gasaymeh, 2016; Anwar, 2018; Staub et al., 2010). This means that if a bank is cost-efficient in a specific year then it will be cost-efficient the following year. The bank has accumulated knowledge and technologies in previous years that enable them to decrease their costs and enhance productivity the following year (Adjei-Frimpong et al., 2014; Ataullah & Le, 2006). In short, one-year lagged bank cost efficiency acts like an exploratory variable and is positively significant in relation to current cost efficiency. As a result, a dynamic relationship of bank cost efficiency may exist in Vietnam domestic banks.

To capture these dynamic relationships, one-year lagged market power is included in the bank market power model (equation 4.1) and one-year lagged efficiency is included in the bank cost model (equation 4.2). These empirical models are rewritten as follows:

The dynamic bank market power:

$$LER_{it} = \beta_0 + \beta_1 LER_{it-1} + \beta_2 BANKSIZE_{it} + \beta_3 CAP_{it} + \beta_4 NPL_{it} + \beta_5 GDP_{it} + \beta_6 INF_{it} + \varepsilon_{it} \quad [4.10]$$

The dynamic bank cost efficiency model:

$$CE_{it} = \beta_0 + \beta_1 CE_{it-1} + \beta_2 BANKSIZE_{it} + \beta_3 CAP_{it} + \beta_4 NPL_{it} + \beta_5 GDP_{it} + \beta_6 INF_{it} + \varepsilon_{it} \quad [4.11]$$

Where:

i represents the individual i_{th} bank at time *t;*

$\alpha_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6 =$ coefficients to be estimated;

ε_{it}= error term;

$\eta_i =$ individual bank effect;

$CE_{it}, LER_{it}, BANKSIZE_{it}, CAP_{it}, NPL_{it}, GDP_{it}, INF_{it}$ are the exploratory variables (see Table 4.10).

CE_{it-1} = one-year lagged cost efficiency;

LER_{it-1}= one-year lagged market power;

The dynamic models for bank cost efficiency and market power (equations 4.10 and 4.11) are estimated using pooled OLS. Table 4.20 depicts the OLS findings for the dynamic models.

Table 4.20 The Dynamic OLS Results of the Factors Impacting LER and CE of Vietnam Banks

Independent Variables	Market Power Model LER_{it}		Cost Efficiency Model CE_{it}	
CE_{it-1}			0.595	***
			(0.000)	
LER_{it-1}	0.641	***		
	(0.000)			
$BANKSIZE_{it}$	0.019	***	0.078	
	(0.001)		(0.311)	
NPL_{it}	-1.006	***	-5.091	
	(0.000)		(0.224)	
CAP_{it}	0.243	*	2.502	
	(0.081)		(0.220)	
GDP_{it}	-0.737		-21.924	*
	(0.338)		(0.066)	
INF_{it}	-0.145		-4.934	***
	(0.092)	*	(0.000)	
Constants	-0.083		1.372	
	(0.341)		(0.270)	
Adjusted R^2	0.661		0.335	
Δ Adjusted R^2	0.313		0.302	
Total observations	230		230	
Total banks	27		27	

Note: Δ adjusted R^2 refers to an increase of the adjusted R2 from the static OLS models (equations 4.4 and 4.5) to the dynamic OLS models (equations 4.10 and 4.11). The p-value is in parentheses; *** and * indicate significance at 1% and 10%, respectively.
Source: Author's computation

The empirical results in Table 4.20 show that all one-year lagged market power and cost efficiency variables (LER_{it-1}, CE_{it-1}) are positively significant at the 1% level. Moreover, the change in adjusted R^2 ($or\ \Delta Adjusted\ R^2$) significantly increased from the static OLS models (equations 4.4 and 4.5) to the dynamic OLS models (equations 4.10 and 4.11). Specifically, the change in $adjusted\ R^2$ in cost efficiency from the static OLS model to the dynamic OLS model is 30.2% (0.302). The change in $adjusted\ R^2$ in market power from the static OLS model to the dynamic OLS model is 31.3% (0.313) (see Table 4.20). Wintoki et al. (2012) note that a dynamic relationship exists if the $adjusted\ R^2$ value increases from the static model to dynamic model. As a result, these are: i) positive statistically significant LER_{it-1} and CE_{it-1}; ii) a significant increase in the *adjusted* R^2, provides evidence of dynamic relationships in the dynamic models of bank cost efficiency and market power (equations 4.10 and 4.11). In summary, theoretical and empirical evidence confirms a dynamic relationship in the empirical models of both bank market power and cost efficiency.

4.4.3.2 Test for Strict Exogeneity

The basic assumption of OLS and fixed-effect estimators is that the exploratory variables and error term cannot be correlated (Hill et al., 2011). If any exploratory variables and the error term are correlated, the variables are considered *endogenous* variables (Hill et al., 2011), which leads to an endogeneity problem. Endogeneity happens in two ways: (i) the lagged dependent variable is useded as a regressor or (ii.) the future (or lead) value of the exploratory variables may be linked with the recent error term (Baltagi, 2008; Nadeem et al., 2017). Therefore, it is necessary to test for strictly exogenous variables.

The strict exogeneity test for the bank market power model is (Woolridge, 2010):

$$LER_{it} = \alpha_0 + \beta_1 BANKSIZE_{it} + \beta_2 CAP_{it} + \beta_3 NPL_{it} + \beta_4 GDP_{it} + \beta_5 INF_{it} + \beta_6 BANKSIZE_{it+1} + \beta_7 CAP_{it+1} + \beta_8 NPL_{it+1} + \beta_9 GDP_{it+1} + \beta_{10} INF_{it+1} + \eta_i + \varepsilon_{it} \quad [4.12]$$

The strict exogeneity test for the bank cost efficiency model is:

$$CE_{it} = \alpha_0 + \beta_1 BANKSIZE_{it} + \beta_2 CAP_{it} + \beta_3 NPL_{it} + \beta_4 GDP_{it} + \beta_5 INF_{it} + \beta_6 BANKSIZE_{it+1} + \beta_7 CAP_{it+1} + \beta_8 NPL_{it+1} + \beta_9 GDP_{it+1} + \beta_{10} INF_{it+1} + \eta_i + \varepsilon_{it} \quad [4.13]$$

Where:
i represents the individual i_{th} bank at time *t*;

$\alpha_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10} =$ coefficients to be estimated;

ε_{it}= error term;

$\eta_i =$ individual bank effect;

$CE_{it}, LER_{it}, BANKSIZE_{it}, CAP_{it}, NPL_{it}, GDP_{it}, INF_{it}$ are the exploratory variables (see Table 4.10).

$BANKSIZE_{it+1}$ = a one-year lead of the bank size exploratory variable;

CAP_{it+1} = a one-year lead of the capitalisation ratio exploratory variable;

NPL_{it+1}= a one-year lead of the NPL ratio exploratory variable;

GDP_{it+1} = a one-year lead of the GDP growth rate; and

INF_{it+1} = a one-year lead of the inflation rate.

The null hypothesis of strict exogeneity is that all parameters (β_6 to β_{10}) are equal to zero (Su, Zhang & Wei, 2016; Wooldridge, 2010). The parameters in equations (4.12) and (4.13) are estimated using fixed effect estimation (Wooldridge, 2010). The results are shown in Table 4.21.

Table 4.21 Outcomes of the Strict Exogeneity Test for Vietnam Banks

Independent Variable	Market Power Model LER_{it}		Cost Efficiency Model CE_{it}	
$BANKSIZE_{it}$	-0.024		-0.116	
	(0.401)		(0.759)	
NPL_{it}	-1.161	***	-5.350	
	(0.003)		(0.292)	
CAP_{it}	0.360	**	1.169	
	(0.012)		(0.531)	
GDP_{it}	-0.046		-12.908	
	(0.966)		(0.367)	
INF_{it}	0.078		-6.196	***
	(0.549)		(0.000)	
$BANKSIZE_{it+1}$	0.077	**	-0.245	
	(0.028)		(0.593)	
NPL_{it+1}	0.175		-3.404	
	(0.612)		(0.453)	
CAP_{it+1}	0.493	**	-0.465	
	(0.027)		(0.873)	
GDP_{it+1}	3.271	**	-28.911	
	(0.023)		(0.125)	
INF_{it+1}	0.247	**	7.124	***
	(0.019)		(0.000)	
Constants	-0.686	***	8.722	***
	(0.003)		(0.005)	
R^2 value	0.225		0.218	
p-value of F-Statistic test	0.000		0.000	
Total observations	230		230	
Total banks	27		27	

Note: the p-value is in parentheses; * **, ** and * indicate significance at 1%, 5% and 10%, respectively.
Source: Author's computation

The exploratory variables $BANKSIZE_{it+1}, CAP_{it+1}$, GDP_{it+1} and INF_{it+1} have a statistically significant relationship with LER_{it} at all conventional significance levels. The one-year lead of the exploratory variable INF_{it+1} has a statistically significant relationship with CE_{it} at the 1% level (see Table 4.21). These findings reject the null hypothesis of strictly exogenous variables. In other words, there is an endogeneity problem in the models. Endogeneity can result in biased and inconsistent outcomes when using OLS and fixed effect regressors (Antoniou et al., 2008; Gujarati & Porter, 2009; Wintoki et al., 2012).

- **The Issue of Endogenous Determinant in the Bank Market Power Model**

In prior studies, the NPL ratio is considered an endogenous variable of bank market power (Adjei-Frimpong et al., 2016; Erler et al., 2017). When high levels of bad debts occur in loan portfolios, banks may offer a lower interest rate to attract higher quality borrowers (Abel & le Roux, 2017). In short, bad debts lessen a bank's capacity to enhance their product prices. They may also decrease a bank's market power (Abel & le Roux, 2017; Angelini & Cetorelli, 2003; Simpasa, 2013). On the other hand, Agoraki, Delis and Pasiouras (2011) and Danisman and Demirel (2019) highlight that strong, reputable banks (high market power banks) choose to avoid risks and reduce NPL rates. This is because these banks want to maintain their franchise value (Danisman & Demirel, 2019). In other words, huge market power decreases the NPL ratio. This finding supports the competition-fragility standpoint in many studies (Agoraki et al., 2011; Danisman & Demirel, 2019). Therefore, the NPL ratio is considered an *endogenous* variable.

In prior studies, the capital ratio is also perceived as an endogenous variable in bank market power (Delis & Pagoulatos, 2009; Erler et al., 2017). Banks with higher capitalisation ratios have higher market power (Abel & le Roux, 2017; Delis, 2012; Simpasa, 2013). This is because well-capitalised banks have greater financial resources that allow them to offer competitive loans to borrowers. Therefore, banks can expand their lending, enjoy high interest margins and ultimately increase their market power (Abel & le Roux, 2017; Erler et al., 2017). High market power banks often choose to improve their capital ratios (Berger et al., 2009; Soedarmono et al., 2011). As a result, the CAP ratio is also an *endogenous* variable.

In a competitive environment, large banks have more competitive advantages and face lower levels of competition (or high market power) than small or medium ones (Fadloli & Chalid, 2017; Tabak et al., 2012). This is because large banks have dominant positions in the financial market that enable them to gain market power (Fernández De Guevara & Maudos, 2007). In short, the bigger the bank, the greater its market power. Bikker, Spierdijk and Finnie (2006) applied the Panza-Rosse model to a large sample of 18,000 banks in 101 countries. The authors conclude that banks with high market power increase in size. As a result, the BANKSIZE variable is considered an *endogenous* variable. This finding echoes previous studies (Delis & Pagoulatos, 2009; Erler et al., 2017).

In summary, this study considers BANKSIZE, the CAP ratio and the NPL ratio as endogenous variables in the bank market power model (Equation 4.1).

- **The Issue of Endogenous Determinants on the Bank Cost Efficiency Model**

In prior literature, the NPL ratio is assumed to be an endogenous variable of bank cost efficiency (Adjei-Frimpong et al., 2014; Delis, Iosifidi & Tsionas, 2017). An increase in the NPL ratio could cause additional repayment monitoring and enforcement costs and ultimately reduce a bank's cost efficiency (Anwar, 2018; Havrylchyk, 2006; Staub et al., 2010). In contrast, banks with low efficiency may have higher NPL levels because of failures in bank monitoring and supervising systems, and poor management decisions (Abel, 2018; Louzis, Vouldis & Metaxas, 2012; Nguyen, 2018; Podpiera & Weill, 2008). This is commonly known as the bad management hypothesis and has been previously discussed (Abel, 2018; Louzis et al., 2012; Podpiera & Weill, 2008).

Well-capitalised banks are considered strong, reputable and safe banks from a depositor's standpoint (Simpasa, 2013). The banks can pay lower interest rates for depositors. Thus, these banks have lower financial costs and are cost-efficient (Abel & le Roux, 2017). In short, a high CAP ratio boosts cost efficiency (Anwar, 2018; Gardener et al., 2011; Nguyen & Nghiem, 2018). In contrast, efficient banks contribute to an increase in the capitalisation ratio (Fiordelisi et al., 2011). This is because reimbursements from short-term cost efficiency (made by lessening the default rate and monitoring costs) can increase owner capitalisation (Fiordelisi et al., 2011; Williams, 2004).

In summary, this study considers both the CAP and the NPL ratios as endogenous variables in the bank cost efficiency model (Equation 4.2).

4.4.3.3 Dynamic Panel Estimation Model and Results

The previous sections have identified problems that violate the assumption of linear conventional estimators such as the pooled OLS and the fixed effect estimator. First, the dataset suffers from heteroscedasticity and autocorrelation problems. Second, dynamic relationships between bank market power and cost efficiency exist. Lastly, the results of the strict exogeneity test provide evidence of endogeneity. As a result, the pooled OLS and fixed effect estimators will generate biased, inconsistent and inefficient estimates (they are no longer BLUE) because of heteroscedasticity, autocorrelation and endogeneity in the dataset (Antoniou et al., 2008; Gujarati & Porter, 2009; Wintoki et al., 2012,).

Therefore, instead of using classic linear regression, this study uses the dynamic generalised method of moments (GMM) regressor (Antoniou et al., 2008; Wintoki et al., 2012; Zeitun et al., 2017). The dynamic panel models can be expressed mathematically as follows:

The dynamic bank market power model:

$$LER_{it} = \beta_0 + \beta_1 LER_{it-1} + \beta_2 BANKSIZE_{it} + \beta_3 CAP_{it} + \beta_4 NPL_{it} + \beta_5 GDP_{it} + \beta_6 INF_{it} + \eta_i + \varepsilon_{it} \quad [4.14]$$

The dynamic bank cost efficiency model:

$$CE_{it} = \beta_0 + \beta_1 CE_{it-1} + \beta_2 BANKSIZE_{it} + \beta_3 CAP_{it} + \beta_4 NPL_{it} + \beta_5 GDP_{it} + \beta_6 INF_{it} + \eta_i + \varepsilon_{it} \quad [4.15]$$

Where:

i denotes the individual i_{th} bank at time *t*;

$\alpha_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_6 =$ coefficients to be estimated;

ε_{it}= error term;

$\eta_i =$ individual bank effect;

$CE_{it}, LER_{it}, BANKSIZE_{it}, CAP_{it}, NPL_{it}, GDP_{it}, INF_{it}$ are the exploratory variables (see Table 4.10).

CE_{it-1} = a one-year lagged cost efficiency;

LER_{it-1}= a one-year lagged market power;

4.4.3.3.1 Reasons for Applying the Qualified GMM Estimator

This study uses the system GMM (sysGMM) estimator to determine the dynamic bank market power and cost efficiency models (equations 4.14 and 4.15). The reasons are:

- The GMM estimator can eliminate the individual bank effect (or the heterogeneity problem) by converting the first-differencing equation (Greene, 2018). Moreover, the GMM estimator can be used to investigate dynamic relationships because it uses a matrix of instruments to generate consistent, unbiased results (Antoniou et al., 2008; Wintoki et al., 2012; Zeitun et al., 2017). In other words, the GMM estimator is the best choice for estimating the dynamic panel models (Equations (4.14) and (4.15)).

- The Arellano and Bond (1991) test is used to report first-order AR(1) and second-order AR(2) correlation in the error term. This test is important because it is used to test the GMM estimator consistency (Baltagi, 2008, p. 153). The null of the non-existence serial correlation in the AR(2) test must be checked to confirm that there is no AR2 correlation in the error term (Baltagi, 2008). The dynamic GMM estimator is an appropriate technique for dealing with the presence of autocorrelation.

- The sysGMM regressor is as an extension of the first-differencing GMM estimator created by Arellano and Bond (1991). The system GMM uses supplementary instruments, including lagged levels and lagged differences of variables, to reduce biases and inconsistencies because of weak instruments associated with the first-differencing GMM estimator (Bond et al., 2001; Presbitero, 2005). This technique enhances accuracy and provides efficient estimating results when data series are persistent (α towards one) and have gaps in the panel data (Antoniou et al., 2008; Wintoki et al., 2012; Zeitun et al., 2017). As a result, the sysGMM is used in this study instead of first difference GMM. A two-step procedure is chosen to generate more efficient results because heteroscedasticity is present in the dataset (Blundell & Bond, 1998; Nguyen, 2018).

- The GMM estimator uses the lagged value of endogenous and dependent variables (BANKSIZE, the CAP ratio and the NPL ratio) as instruments to solve the endogeneity problem. This is the advantage of the GMM estimator; strictly exogenous instruments are difficult to find (Wintoki et al., 2012).

4.4.3.3.2 System GMM Estimation Empirical Results

The parameter estimation results from the sysGMM (2-step) are shown in Table 4.22. The P-values of the F-statistics test in both the dynamic bank market power and bank cost efficiency models are significant at the 1% level. These values confirm the joining of the independent variables. Several diagnostic tests linked with the sysGMM (2-step) technique are reported below.

- **Arellano-Bond Test**

The Arellano-Bond test presents the results of the first and second order correlation AR(1)/AR(2) tests of the error term. The null hypothesis states that non-existence of autocorrelation in the error terms. The p-value in Table 4.22 shows statistically significant levels for AR (1) (values of 0.003 and 0.012) and AR (2) (values of 0.498 and 0.183). Hence, the null of non-existence of autocorrelation is rejected in AR (1) and accepted in AR (2). In other words, second order correlation does not exist in the error term. These results indicate that the estimator is consistent (Baltagi, 2008).

- **Exogeneous Instruments: the Hansen Test**

The validity of the instruments used (or exogenous instruments) is critical in the sysGMM estimator. The Hansen Test is used to check for the exogeneity of the instruments (Roodman, 2009a). The null hypothesis of exogenous instruments is assumed. Table 4.22 shows that the p-values of the Hansen Test are above the conventional statistically significant levels (0.366 and 0.493). Hence, the null of exogenous instruments cannot be rejected. In short, the instruments are valid (Roodman, 2009a).

- **Exogeneous Subset Instruments Difference-in-Hansen Test**

The exogeneity of lagged differences in the subset instruments is assumed under the sysGMM estimators. The null of exogenous subset instruments is presumed. The p-values of the Difference-in-Hansen Test in Table 4.22 illustrate statistically significant levels (0.738 and 0.495). Hence, the null of exogenous subset instruments cannot be rejected. In other words, the subset instruments are strictly exogenous (Roodman, 2009a).

- **Total Instruments**

Total instruments must be less than the total group to avoid a weakening of the Hansen test (Roodman, 2009b). Table 4.22 shows the total number of instruments (16 and 17) is smaller than the total group (value of 27). This confirms that the sysGMM results are valid (Roodman, 2009b).

Table 4.22 The Empirical Parameters Estimation Results from the SysGMM (2-step) of the Determinants Impacting on LER and CE in Vietnam Banks

Independent Variable	Dynamic Market Power Model $\boldsymbol{LER_{it}}$		Dynamic Cost Efficiency Model $\boldsymbol{CE_{it}}$	
CE_{it-1}			0.692	**
			(0.022)	
LER_{it-1}	0.367	***		
	(0.007)			
$BANKSIZE_{it}$	0.042	**	-1.783	**
	(0.015)		(0.032)	
CAP_{it}	1.056	***	1.494	
	(0.007)		0.819	
NPL_{it}	-2.98	**	-55.083	*
	(0.023)		(0.097)	
GDP_{it}	-2.162		96.861	*
	(0.211)		(0.074)	
INF_{it}	-0.213	**	14.540	*
	(0.038)		(0.082)	
Constant	-0.235		15.117	
	(0.265)		(0.030)	
P-value of the F-Statistic test	0.000		0.004	
Total Observations	230		230	
Total Banks	27		27	
Total Instruments	17		16	
P-value of the Hansen Test	0.493		0.366	
P-value of the Arellano–Bond Test				
AR (1)	0.012		0.003	
AR (2)	0.183		0.498	
P-value the Difference-in-Hansen Test				
GMM Instruments for Levels	0.495		0.738	

Note: The p-value is in parentheses; * **, ** and * denote significance at 1%, 5% and 10%, respectively.
Source: Author’s computation

4.4.3.3.3 Analysing the Impact of Factors on the LER and CE of Vietnam Banks

The sysGMM empirical results in Table 4.22 show the one-year lagged CE_{it-1} has a positive relationship with cost efficiency at the 5% significance level. This finding shows that the Vietnam domestic banks have accumulated technological knowledge in the previous year that enables them to be more efficient the following year (Ataullah & Le, 2006). This outcome also provides evidence that the banking system has encouraged the Vietnam domestic banks to maintain cost efficiency over time (Adjei-Frimpong et al., 2014). This outcome reiterates previous studies that document a positive relationship between CE_{it} and CE_{it-1} (Adjei-Frimpong et al., 2014; Al-Gasaymeh, 2016; Ataullah & Le, 2006; Staub et al., 2010). The one year lagged LER_{it-1} has a positive relationship with the dependent variables at the 1% significance level. The Vietnam banking industry is characterised by its non-transparent information, network lending relationships and the limitation of competition regulations, to prevent a reduction in bank market power and produce constant profit (Berger et al., 2000; Delis, 2012). In other words, the Vietnam banking industry maintains its rigidity and stability. New entry banks are selected, well-managed and launched according to a schedule to maintain the Vietnam banking system's stability (Adjei-Frimpong et al., 2016). This finding explains the SOM result that 70% of Vietnam domestic banks (19 of 27 banks) have maintained their financial position (in either group 1 or 2) over the study period (2008 to 2017). This finding echo previous studies that show a positive relationship between LER_{it} and LER_{it-1} (Delis, 2012; Delis & Pagoulatos, 2009; Erler et al., 2017).

BANKSIZE has a positive relationship with bank market power at the 1% significance level. BANKSIZE has a negative relationship with bank cost efficiency at the 5% significance level. The positive relationship between $BANKSIZE_{it}$ and LER_{it} suggests that Vietnam domestic banks with more assets enjoy greater market power. Large banks have more opportunities to access cheaper financial resources through economies of scale in depositors' funds, are better at dealing with moral hazard issues (borrowers) and can limit competition regulation to enhance their market power (Delis, 2012). This finding reiterates previous studies (Delis, 2012; Fernández de Guevara et al., 2005). The negative relationship between BANKSIZE and CE_{it} reflects the fact that the bigger Vietnam domestic banks suffer from diseconomies of scale (Al-Gasaymeh, 2016). Bigger banks suffer from higher costs than smaller banks because of bureaucratic problems. This ultimately results in lower efficiency (Delis & Papanikolaou, 2009; Gardener et al., 2011).

Prior studies indicate that Vietnam domestic banks operate in response to the quiet life hypothesis (Nguyen & Nguyen, 2018; Nguyen et al., 2016b). This means that increased competition will improve bank cost-efficiency (Nguyen et al., 2016b). The statistically significant difference, in terms of market power, between the two super-class groups of banks (1 & 2) provides evidence that the two groups of banks, with different levels of market power, exist side-by-side in the Vietnam banking industry. Group 2 banks enjoy higher market power because they are less likely to be impacted by competition from other banks. As a result, bigger banks (group 2) are less efficient in response to the quiet life hypothesis. Hence, this study argues that BANKSIZE is negatively significant for bank cost efficiency, although this relationship (BANKSIZE and CE) is unclear in prior studies. A positive relationship is documented in previous studies (Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016). In contrast, Gardener et al. (2011) report a negative relationship.

The relationship between BANKSIZE and CE is unclear in previous studies, with some reporting a positive relationship and others, a negative relationship. However, this could be related to the chosen study period. For example, Gardener et al. (2011) collected Vietnam domestic bank data from 1998 to 2004, others gathered data from 2005 to 2012 (Phan et al., 2016); from 1999 to 2009 (Stewart et al., 2016); and from 2000 to 2014 (Nguyen & Nghiem, 2018). Big Vietnam domestic banks may be less efficient from 1998 to 2004 (as reported by Gardener et al., 2011). However, these banks may have improved their management systems and become more cost-efficient from 1999 to 2014 (Nguyen & Nghiem, 2018; Phan et al., 2016; Stewart et al., 2016). To briefly reiterate, this study collected bank data from 2008 to 2017. During that period, successful M&A deals may have impacted the efficiency of Vietnam domestic banks. For example, a merger between SCB, Tin Nghia Bank and First Commercial Bank occurred in 2011. This merger increased the assets (of SCB) 2.5 times (from 60,183 billion VND to 149,205 billion VND), and 2.4 times for the bank charter capital (from 4,710 billion VND to 11,370 billion VND). The new SCB almost doubled its staff numbers (from 2075 to 3959 employees). The NPL rate (of SCB) dramatically increased (1.28% to 7.25%) from 2009 to 2012 (Hoang et al., 2016; SCB, 2012). In short, M&A activity may quickly increase bank size, but impact negatively on bank efficiency because of increases in NPL and operating costs (because of increases in bank staff numbers). In other words, this study argues that BANKSIZE is negatively related to bank cost-efficient for the study period (2008 to 2017).

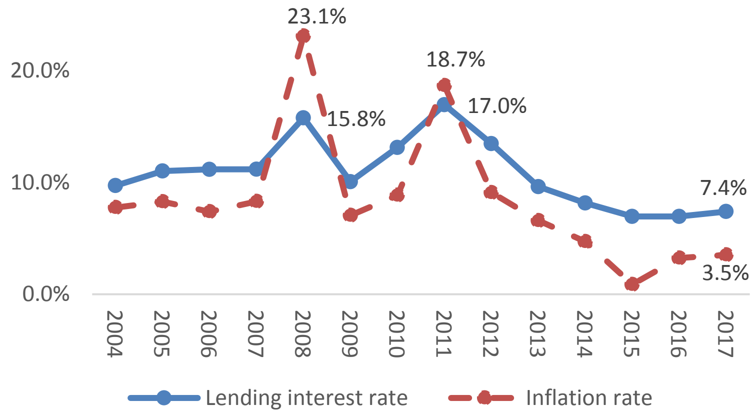
The CAP ratio is positively associated with bank market power at the 1% level, but it is insignificant in terms of bank cost efficiency. The positive relationship between CAP_{it} and LER_{it} highlights that high levels of capitalisation improve a bank's reputation (Simpasa, 2013). Banks can attract more borrowers and are able to offer better loan prices (Abel & le Roux, 2017). In addition, well-capitalised banks have greater financial resources that allow them to expand their lending market, enjoy higher interest margins and ultimately increase their market power (Abel & le Roux, 2017; Erler et al., 2017). This finding echoes previous studies that show a positive relationship between CAP_{it} and LER_{it} (Abel & le Roux, 2017; Delis, 2012; Erler et al., 2017; Simpasa, 2013). In contrast, the insignificant relationship between CAP_{it} and CE_{it} indicates that the funds' source (from equity owners or depositors) does not impact on bank efficiency (Adjei-Frimpong et al., 2016). In short, Vietnam domestic banks can increase capitalisation to increase market power, but not cost efficiency.

The NPL ratio negatively impacts on market power at the 5% level and on cost efficiency at the 10% level. The negative relationship between NPL_{it} and LER_{it} implies that a higher NPL rate harms a bank's strength and reputation and reduces its market power. The reason is that higher NPL rates lead to deterioration in a bank's loan portfolio (Abel & le Roux, 2017). In such circumstances, a bank can reduce its total number of loans and seek low-risk customers (Abel & le Roux, 2017). Banks are unable to set higher loan prices because they must compete to attract quality borrowers (Abel & le Roux, 2017). In other words, the NPL problem reduces Vietnam domestic banks' market power. This situation is similar to that of Zimbabwean banks that had higher NPLs, which ultimately impact their bank market power (Abel & le Roux, 2017; Simpasa, 2013). In addition, higher NPL_{it} rates damage bank efficiency. The reason is that the higher the NPL rate, the greater the possibility of bad debts. This means that Vietnam domestic banks must pay additional monitoring and enforcement costs meaning they are less efficient. In this respect, Vietnam domestic banks are like banks in Indonesia, Europe, and Brazil; scholars have noted a negative relationship between NPL and CE (Anwar, 2018; Delis & Papanikolaou, 2009; Staub et al., 2010). This finding confirms that the NPL_{it} rate impedes cost-efficiency in Vietnam domestic banks. In contrast, prior studies report a negative *insignificant* relationship (Phan et al., 2016; Stewart et al., 2016). In short, Vietnam domestic banks focus on diminishing the macro-prudential NPL factor to boost their market power and efficiency.

The GDP growth rate impacts positively on bank cost efficiency at the 10% significance level but does not influence market power. Economic growth contributes to higher incomes and lower borrower default rates (Anwar, 2018; Erler et al., 2017). Hence, banks can reduce losses from bad debts and reserve lower provisions for bad debts when default rates decrease (Anwar, 2018). In other words, Vietnam domestic banks can improve their efficiency under rapid GDP growth (Phan et al., 2016). This outcome echoes previous studies that show a positive relationship between GDP_t and CE_{it} for Vietnam domestic banks (Gardener et al., 2011; Phan et al., 2016). In contrast, the insignificant association between GDP_t and LER_{it} indicates that economic growth does not play an important role in moulding the competitive environment of Vietnam domestic banks.

The inflation rate has a positive relationship with cost efficiency at the 10% significance level. It is negatively significant with bank market power at the 5% level. Theory suggests that the inflation rate acts like a barrier preventing banks from achieving higher cost efficiency (Gardener et al., 2011). This is because a higher inflation rate leads to higher financial costs that are the result of an increased number of bad loans so they have to reserve more money to cover bad debts (because the default rate increases) and spend extra on operating costs, including salaries and/or employee benefits (Anwar, 2018; Chan & Karim, 2010). These additional costs reduce bank cost efficiency. However, the positive relationship between INF_t and CE_{it} indicates that Vietnam domestic banks use higher lending rates to cover the additional costs (Chan & Karim, 2010). Bank lending interest rates and inflation rates were gathered from the Worldbank database and analysed to verify this argument. The results are shown in Figure 4.6.

Figure 4.6 Inflation and Loan Interest Rates in Vietnam (2004-2017)



Source: Author's computation from Data of the World Bank

Figure 4.6 shows that when the Vietnamese inflation rate reached its highest rate of 23.1% in 2008 and a second peak at 18.7% in 2011, the bank lending interest rate also increased to its highest rate: 15.8% in 2008 and 17.0% in 2011. During the deflation period from 2011 to 2017, the inflation rate dropped from 18.7% to 3.52%; the bank lending interest rate also decreased, from 17.0% in 2011 to 7.4% in 2017 (see Figure 4.6). The inflation rate and bank lending interest rate are correlated at 91% level (have a correlation ratio of 0.91). The relationship between these two variables is statistically significant at the 1% level. The positive correlation of inflation rate and bank lending interest rate provides evidence that Vietnam domestic banks increased lending rates to compensate for losses associated with additional costs. In short, this shows the positive relationship between INF_t and CE_{it}.

The negative relationship between INF_t and LER_{it} indicates that higher inflation rates lower loan demand (Erler et al., 2017). Banks cannot set higher prices for their financial products because they are competing to attract more customers. The inflationary environment decreases Vietnam bank market power (Abel & le Roux, 2017). This agrees with Angelini and Cetorelli (2003) whose study showed a negative relationship between INF_t and LER_{it}.

4.4.4 Robustness Check

This study uses robustness checks to verify the robustness and credibility of the results. This provides evidence for the structural validity of the estimated coefficients (Lu & White, 2014). The robustness test results are shown in Tables 4.23 and 4.24.

Table 4.23 presents the results on the determinants that influence bank market power estimated using the pooled OLS, fixed effect, and sysGMM regressions. BANKSIZE and the CAP ratio are positively significant at all conventional levels (1% and 5%) in relation to bank market power when estimated by different regressions (Table 4.23). In short, BANKSIZE and the CAP ratio are robust and credible. In addition, the NPL ratio is negatively significant at conventional levels (1% and 5%) in impacting bank market power when predicted by different estimators (Table 4.23). These outcomes also provide evidence of the structural validity of the estimated coefficients (BANKSIZE, the CAP and NPL ratios). In other words, bank executives should be confident when deciding to enlarge bank size (or increase total assets), raise capital, develop IT infrastructure or improve bank risk systems in order to gain bank market power.

Table 4.24 shows the estimated coefficients of the determinants that influence bank efficiency, estimated using pooled OLS, the fixed effect, and the sysGMM estimators. The NPL ratio is negatively significant at the 10% level in relation to bank cost efficiency when estimated by different regressions (Table 4.24). This demonstrates a robust negative relationship between NPL_{it} and CE_{it}. In other words, bank executives should be confident developing IT infrastructure or improving bank risk systems to decrease NPLs (or bad debts) to gain cost-efficiency.

Table 4.23 The Robustness Check of the Determinants Impacting on the Bank Market Power Model

Market power model (LER_{it})	**Pool OLS**		**Fixed effect**		**System GMM**	
LER_{it-1}					0.367	***
					(0.007)	
$BANKSIZE_{it}$	0.07	***	0.016		0.042	**
	(0.000)		(0.164)		(0.015)	
CAP_{it}	0.94	***	0.483	**	1.056	***
	(0.000)		(0.000)		(0.007)	
NPL_{it}	-1.68	***	-1.112	***	-2.98	**
	(0.000)		(0.000)		(0.023)	
GDP_{it}	0.02		1.831	**	-2.162	
	(0.985)		(0.044)		(0.211)	
INF_{it}	0.05		-0.163	**	-0.213	**
	(0.552)		(0.044)		(0.038)	
Constants	-0.64	***	-0.087		-0.235	
	(0.000)		(0.491)		(0.265)	
R^2	0.3605		0.195			
Correlation 2 (p-value)					0.183	
Hansen (p-value)					0.493	
Difference-in-Hansen (p-values)					0.495	

Note: *p*-values are in parentheses; * **, ** and * denote significance at 1%, 5% and 10%, respectively.
Source: Author’s computation

Table 4.24 Robustness Check for Determinants Impacting on Bank Cost Efficiency Model

Cost Efficiency model (CE_{it})	Pool OLS		Fixed effect		System GMM	
CE_{it-1}					0.692	**
					(0.022)	
$BANKSIZE_{it}$	0.25	***	-0.501	***	-1.783	**
	(0.001)		(0.001)		(0.032)	
CAP_{it}	3.54	**	-0.378		1.494	
	(0.035)		(0.811)		0.819	
NPL_{it}	-8.26	*	-5.096		-55.083	*
	(0.076)		(0.221)		(0.097)	
GDP_{it}	-20.36		3.674		96.861	*
	(0.149)		(0.763)		(0.074)	
INF_{it}	0.30		-2.836	**	14.540	*
	(0.787)		(0.010)		(0.082)	
Constants	0.07		7.721	***	15.117	
	(0.956)		(0.000)		(0.030)	
R-squared	0.0518		0.127			
Correlation 2 (p-value)					0.498	
Hansen (p-value)					0.366	
Difference-in-Hansen (p-values)					0.738	

Note: *p*-values is contained in parentheses. * **, ** and * denote significance levels (1%, 5% and 10%).
Source: Author's computation

4.5 Chapter Summary

This study used the SOM technique to track the financial trajectory patterns and categorised Vietnam domestic banks into two super-class groups. This answered the study's first objective. The dataset used for this study is unbalanced panel data with 258 observations from 2008 to 2017. Two super-class groups of banks were identified: group 1 and group 2 banks. Whereas group 2 contains commercial state and joint-stock banks, group 1 features only joint-stock banks. Group 2 banks had a higher mean asset value than group 1 banks. The 2D SOM map provides a clear picture of the financial trajectory patterns in the Vietnam banking industry from 2008 to 2017. Some common trajectory patterns have been documented. Nineteen banks retained their positions in one of the groups (1 or 2) over the entire study period. In contrast, eight banks shifted from small bank group (group 1) to the bigger bank group (or group 2), or vice versa.

To answer objective two, this study has used a non-structural indicator, the Lerner index, to capture bank market power and used the non-parametric DEA technique (under the assumption of VRS) to measure the efficiency of Vietnam banks. Differences in the Lerner scores (representing market power) between the two super-class bank groups (1 and 2) are statistically significant at the 1% level. In contrast, the PTE and CE scores (that measure bank efficiency) are the same. The different market power between groups 1 and 2 banks contradicts Nguyen and Nghiem's (2018) results. The contradictory results might imply that dividing Vietnam domestic banks into state-owned and joint-stock may be inappropriate when investigating their market power. This is because the difference in market power between commercial state and joint-stock banks is not statistically significant (Nguyen & Nghiem, 2018). In contrast, the SOM technique (with an unsupervised algorithm) can better capture differences in bank market power and can be used to divide Vietnam domestic banks into two groups, consisting of weak banks (group 1) and strong banks (group 2). Hence, this study contends that there are two groups of banks with different levels of market power existing side-by-side in the Vietnam banking industry. This is a result of size concentration because of M&As. The existence of weak and strong banks damages the competitive environment of Vietnam domestic banks (Hoang et al., 2016). The reason is that group 2 banks are profitable because of greater market power whereas group 1 banks struggle to cut costs to remain viable.

To answer the third objective, the study identified the factors that impact bank market power and efficiency. The factors that influence bank market power and cost efficiency (performance) include bank characteristics (BANKSIZE, the CAP ratio, and the NPL ratio) and macro-economic indicators (GDP and the inflation rate). The dataset has tested stationarity (the Fisher-type option). The p-values are significant at all levels. Thus, the null of non-stationarity (unit root) can be confidently rejected. Thus, the dataset has stationarity or no unit root. The Pearson pairwise correlation results indicate that the correlation ratios among the exploratory variables are less than 0.8. Additionally, the VIF values do not exceed 10. As a result, the pooled OLS regressor is suitable for estimating the models. In addition, a fixed effect estimator is used to examine the relationship between bank characteristics and macroeconomic indicators of the bank market power and cost efficiency models. Unlike pooled OLS, a fixed effect estimator can deal with heterogeneity (Gujarati and Porter, 2009).

The p-values for the heteroscedasticity tests (including the Breusch-Pagan/Cook-Weisberg test and the Modified Wald Test) are all statistically significant. Moreover, the p-value for the autocorrelation test (Wooldridge test) is statistically significantly level at 1%. These results confirm that there is heteroscedasticity and autocorrelation in the dataset. In short, the pooled OLS and fixed effect are no longer BLUE (Baltagi, 2008; Gujarati and Porter, 2009). Furthermore, theoretical and empirical evidence has shown the existence of dynamic relationships in the bank market power and cost efficiency models. Therefore, the one-year lagged LER_{it} and CE_{it} were used as additional exploratory variables to capture the dynamic relationship. This study used the strict exogeneity test following Wooldridge (2010). The result confirms that the dataset has an endogeneity problem. BANKSIZE, the CAP ratio, and the NPL ratio are endogenous variables.

The existence of heteroscedasticity, autocorrelation, and endogeneity result in biased, inconsistent and inefficient parameters if estimated using pooled OLS and fixed effect estimators (Antoniou et al., 2008; Wintoki et al., 2012). Therefore, this study used the dynamic sysGMM (2-step) estimator to identify the determinants that impact on LER_{it} and CE_{it}. Finally, robustness checks were used to assess the results. The key empirical findings of this chapter are outlined below:

- One year lagged CE_{it-1} has a positive relationship with CE_{it} at the 5% significance level. This indicates that Vietnam domestic banks have accumulated technological know-how in the previous year; this enables banks to achieve higher efficiency the following year (Ataullah & Le, 2006). This outcome also provides evidence that the banking system encouraged Vietnam domestic banks to maintain efficiency over time (Adjei-Frimpong et al., 2014). The one year lagged LER_{it-1} also has a positive relationship with LER_{it} at the 1% significance level. This provides evidence that the Vietnam banking industry has specific features, such as non-transparent information, networking lending relationships and limitation of competition regulations, to prevent a decrease in bank market power and sustain constant profitability (Berger et al., 2000; Delis, 2012). In other words, the Vietnam banking industry maintains a rigid, stable system. New entry banks are designated, managed and established in a plan to maintain stability in the banking system and prevent a reduction in their franchise value (Adjei-Frimpong et al., 2016). This outcome explains the SOM result that over 70% of the local banks (19 of 27 banks) have kept their financial position as group 1 or 2 banks from 2008 to 2017.

- BANKSIZE has a significant positive relationship with LER_{it} at the 1% level, but it has a negative relationship with CE_{it} at the 5% significance level. The fact that BANKSIZE impacts positively on bank market power suggests that the larger asset value banks have greater power. The reason is that big banks have more opportunities to source cheaper financial resources through economies of scale related to depositors' funds. They have a greater ability to handle moral hazard issues (borrowers) and are able to maintain their market power through limiting competition (Delis, 2012). However, big banks are inefficient in spreading their operational costs compared with smaller banks because of bureaucratic problems that lead to less efficiency (Delis & Papanikolaou, 2009; Gardener et al., 2011). Bank size becomes an essential factor for banks to compete in the financial market. This study argues that BANKSIZE is positively significant to bank market power and is negatively significant to cost efficiency from2008 to 2017. This is because successful M&A activity may improve bank size quickly, but negatively impacts bank efficiency because of increases in NPL values and operational costs (because of an increased number of staff).

- Though the CAP ratio shows a significant positive relationship with LER_{it} at the 1% level, the CAP ratio is no significant relationship with CE_{it}. The fact that the CAP ratio positively influences LER_{it} indicates that well-capitalised banks have more available funds that allow them to seize opportunities that are considered riskier. This means they can lend to more borrowers and gain higher market power (Abel & le Roux, 2017; Erler et al., 2017).

- The NPL ratio significantly negatively influences LER_{it} and CE_{it} at 5% and 10 % levels, respectively. These findings indicate that higher NPL rates harm a bank's market power and reduce efficiency. Higher NPL rates lead to a deterioration of a bank's loan portfolio (Abel & le Roux, 2017). In such circumstances, a bank can reduce the number of loans and look for good borrowers (Abel & le Roux, 2017). Banks cannot set higher loan prices (or lower their market power) because they must compete to attract quality borrowers (Abel & le Roux, 2017). In addition, higher NPL rates lead to a higher probability of bad debts. Banks must pay additional monitoring and enforcement costs, thus they become less efficient.

- The macro-economic GDP rate has a significant positive impact on CE_{it} at the 10% level. This relationship suggests that the Vietnam domestic banks gained efficiency under rapid GDP growth. The INF rate variable has a significant positive relationship with bank cost efficiency at the 10% level, but is negatively significant with market power at the 5% level. The positive relationship between the INF and CE provide evidence that Vietnam domestic banks charge higher lending rates to compensate for additional costs (Chan and Karim, 2010).

- Lastly, the robustness results show that the NPL ratio is the macro-prudential determinant that bank executives must eliminate. To reduce the NPL ratio, the bank customer dataset needs to be analysed. The customer dataset is analysed in Chapter 5.

Chapter 5

Research Results and Findings on the Retail Banking Sector of Domestic Vietnam Banks

5.1 Introduction

This chapter explores the retail banking sector of domestic Vietnam banks to answer the study's objectives four to seven. The chapter investigates the critical factors that impact borrowers' creditworthiness, the study's method for decreasing borrowers' default rates (or reducing the total number of NPL), critical factors that affect interest charges and whether banks apply a risk-based pricing (RBP) strategy.

The chapter is organised as follows: Section 5.2 discusses the data collection. Section 5.3 explains the transformed empirical data and investigates the critical factors related to borrowers' creditworthiness to answer objective 4. Section 5.4 investigates the classificatory power of the hybrid model that integrates logit regression and ANN to respond about objective 5. Section 5.5 identifies the factors that determine the interest rate borrowers are charged to answer objective 6. Section 5.6 examines whether a RBP strategy is deployed within Vietnam banks to answer objective 7 and section 5.7 summarises the chapter.

5.2 Data Collection

In Chapter 4, it was argued that an NPL is a critical factor in reducing a bank's cost efficiency and market power. Hence, decreasing NPL is one of the main ways local Vietnam banks can improve their market power, efficiency and long-term development. To reduce the total number of non-performing loans, we analyse banks' customer datasets. However, collecting data for this study was difficult because:

1. Banks cannot release customers' personal information to a third party for any reason that is not related to the bank's business. They cannot publicize their customer's personal information because it is illegal.

2. Banks are apprehensive about their financial data being made public because it may negatively affect their business when they list on the stock market.

After contacting all banks, only one bank in group 1 agreed to provide its customer dataset for this analysis. The Viet Nam Thuong Tin Commercial Joint Stock Bank (Vietbank) has a total asset value of 41,533 billion VND, a total equity value of 3,329 billion VND and a charter capital value of 3,249 billion VND at the end of financial year 2016 (Vietbank, 2017). The bank wanted to investigate the critical factors that impact their customers' creditworthiness and so develop a method for decreasing the borrowers' default rate.

There are 15,865 observations in the customer dataset with a total value of outstanding loans at 12,177 billion VND. Table 5.1 summarises the borrowers' credit data.

Table 5.1 A Summary of the Variables in Vietnam Banks Borrowers' Credit Data 2006 to 2017

No.	Variable	Abbreviation	Description	Type	Note
1	Lending decision (Y1)	Loan_group_2	Loans are classified as 'good' or non-performance loans.	Category	
2	Registration date	Add_date_3	The date when a borrower is registered to the bank's system	Continuous number	2006 to 2017
3	Industry class	industry_4	Industry the borrowers	Category	
4	Supplements class	subclass_5	Subclass of the borrowers	Category	
5	Economic class	economic_6	Economic class of the borrowers	Category	
6	Loan value	Loan_value_9	Outstanding loan value	Continuous number	
7	Credit term	MIA_type_12	Credit terms for borrowing	Category	
8	Interest rate	interest_13	Interest rate charged	Continuous number	1% to 49%
9	Opening date	open_date_14	Date loan is released	Continuous number	2007 to 2017
10	Loan duration	duration_16	Loan duration	Continuous number	1 to 25 years
11	Loan purpose	purpose_20	Borrower's purpose	Category	
12	Purpose group	purpose_group_19	Purpose group	Category	
13	Lending programs	HTCV_21	Support program	Category	
14	Individual credit	TDCN_22	Support program	Category	
15	Business credit	TDDN_23	Support program	Category	
16	Sponsor	TOFL_24	Support program	Category	
17	Customer	Customer_8	Old or new borrower	Category	
		Number of customers	15,865		
		Total loan value	12 177 billion VND		
		Total overdue and bad loan value	381.8 billion VND		

Source: Author's summary

The original dataset contains continuous numerical (quantity) and categorical (quality) variables. The current variables cannot be used for the modelling process. These variables need to be transformed to *numerical form*. These transformations are discussed in the next section.

5.3 Objective 4: To Identify Critical Factors Influencing The Lending Decision and Relating to Borrowers' Creditworthiness in Vietnam retail banking

5.3.1 Empirical Lending Model

The empirical lending model used to investigate borrower creditworthiness is mathematically expressed as follows:

$$\boldsymbol{Lending\ decision}\ (\boldsymbol{Y_1}) = \boldsymbol{f}(Regdate_3, loanValue_9, interest_{13}, openDate_{14}, duration_{16}, industry_4, economic_6, customer_8, MIAtype_{12}, purposeGroup_{19}, purpose_{20}, HTCV_{21}, TDCN_{22}, TOFL_{24}, \boldsymbol{\varepsilon}) \qquad [5.1]$$

Where:

Lending decision ($\boldsymbol{Y_1}$) = 1 if loan is in 'default' (bad loan); 0 if loan is repaid (good loan)

$Regdate_3, loanValue_9, interest_{13}, openDate_{14}, duration_{16}, industry_4, economic_6, customer_8, MIAtype_{12}, purposeGroup_{19}, purpose_{20}, HTCV_{21}, TDCN_{22}, TOFL_{24}$= critical factors that impact lending as described in Table 5.1;

f = regression function; and

ε = error term.

5.3.2 Data Transformation

5.3.2.1 Weight of Evidence (WOE) and Information Value (IV) Techniques

The first step is to transform the raw data for the modelling. The WOE and IV indicators are used to transform the original dataset. WOE measures the relationship between an attribute (binary or category) of an independent variable and the target class of a dependent variable (good or bad) (Trinkle & Baldwin, 2007). The relationship can be expressed mathematically as follows (Abdou, 2009; Huysmans et al., 2006; Trinkle & Baldwin, 2007):

$$WoE = \left[ln\left(\frac{Proportion\_Goods\_in\_category}{Proportion\_Bad\_in\_category}\right)\right] \times 100 \qquad [5.2a]$$

Where:

$Proportion\_Goods\_in\_category$ is calculated by the number of good borrowers of an attribute divided by total number of good borrowers; and

$Proportion\_Bad\_in\_category$ is calculated by the number of bad borrowers of an attribute divided by total number of bad borrowers.

The WOE value ranges from negative to positive. Based on the WOE value, we can compute and arrange attributes following a logical distribution; e.g., from negative values (that indicate a high-risk borrower) to positive values (that indicate a low-risk borrower). This computation can identify the riskiest attribute (indicates the lowest negative WOE value) that contributes to 'default' borrowers.

The IV indicator is used to measure the variable's strength. The IV indicator is computed as follows:

$$IV = \sum_{i=1}^{n} [(Proportion\ Goods\ in\ category_i - Proportion\ Bads\ in\ category_i) \times WoE_i \quad [5.2b]$$

Variables with an IV value less than 0.02 are judged as having an unpredicted ability (Abdou, 2009; Siddiqi, 2006; Trinkle & Baldwin, 2007) and are eliminated from the study. Variables with an IV value greater than 0.02 have predictive power that ranges from very weak to strong and are used to develop the empirical lending model (see Table 5.2).

Table 5.2 Classification of the Predictive Power of IV Variables

IV value	Predictive Ability
< 0.02	Very weak (unpredictive)
0.02 to 0.1	Weak
0.1 to 0.3	Medium
> 0.3	Strong

Note: IV = information value.
Source: Abdou (2009); Siddiqi (2006)

5.3.2.2 WOE and IV Empirical Results

The critical factors that impact lending (see Table 5.1), are transformed for modelling using the WOE and IV techniques. The empirical results are shown below.

1. Variable 1: Lending Decision (Y1)

Article 10 of Circular 02/2013/TT-NHNN ("Circular 02") has been used to classify the quality of outstanding loans into five groups; group one to group five (see Table 5.3) (SCB, 2016)

Table 5.3 The Classification of Outstanding Loans

Group	Name	Description	Loan Class
1	Current	Current loans are assessed as fully recoverable for both principal and interest; loan repayments haven't been made for a period of less than 10 days	Good loans
2	Special mention	Loan repayments haven't been made for between 10 and 90 days	Overdue loans
3	Substandard	Loan repayments haven't been made for between 91 and 180 days	Non-performance loans
4	Doubtful	Loan repayments haven't been made for between 181 days and 360 days	Non-performance loans
5	Loss	Loan repayments haven't been made for more than 360 days	Non-performance loans

Source: Adjusted from the SCB bank's 2016 annual report (SCB, 2016)

Table 5.4 shows the values and number of good/bad borrowers in the dataset. There are 15118 (95.29%) borrowers considered good borrowers and 747 (4.71%) bad borrowers (see Table 5.4). Good borrowers are assigned a value of zero and bad borrowers are given a value of one. Therefore, the dependent variable lending decision (Y1) has value of zero (good borrower) and one (bad borrower) in the empirical lending model (Equation 5.1).

Table 5.4 The Classification of Loans in the Bank's Dataset

Group		Loan Value		Observations		Bad-Good Borrowers			
		(bil. VND)	%	Sample	%	Value		Sample	%
1	Current	11 796	96.9%	15 118	95.29%	Good	0	15 118	95.29%
2	Special mention	141	1.2%	251	1.58%	Bad	1	747	4.71%
3	Substandard	31	0.3%	72	0.45%				
4	Doubtful	59	0.5%	58	0.37%				
5	Loss	151	1.2%	366	2.31%				
Total		**12 178**	100%	**15 865**	100.0%			**15 865**	100%

Source: Author's computation

2. Variable 2: Registration Date

The registration date describes the date that a customer is registered in the bank's information system. This variable reflects the time when customers are on the highest rates (either default or non-default rates). The bank can identify and pay attention to risky periods. The variable has a continuous value of date ranges from the 3/1/2006 to 12/6/2017. In the empirical lending model, this variable is grouped by year from 2008 to 2017. Each year is defined in numerical form using 1 to 10 in the corresponding years from 2008 to 2017. Next, the WOE and IV value of each attribute of the registration date is computed. These values are shown in Table 5.5.

Table 5.5 The WOE and IV Values of the Loan Registration Date Variable

Value	**Year**	**1**	**0**	**B (%)**	**G (%)**	**WOE**	**100 x IV**
1	2008	3	202	0.40%	1.34%	12.02	1.123
2	2009	28	512	3.75%	3.39%	-1.01	0.036
3	2010	190	1 314	25.44%	8.69%	-10.74	17.975
4	2011	119	928	15.93%	6.14%	-9.54	9.336
5	2012	52	718	6.96%	4.75%	-3.82	0.845
6	2013	46	816	6.16%	5.40%	-1.32	0.100
7	2014	100	1 853	13.39%	12.26%	-0.88	0.099
8	2015	136	3 624	18.21%	23.97%	2.75	1.587
9	2016	61	3 199	8.17%	21.16%	9.52	12.376
10	2017	12	1 950	1.61%	12.90%	20.83	23.527
Total		747	15 116	100%	100%		**67.008**

Source: Author's computation

Table 5.5 shows that the WOE is negative between 2009 and 2014 and positive from 2015 to 2017 These figures reveal that risky borrowers (those with negative WOE values) were registered on the banking system from 2009 to 2014; conversely, low-risk borrowers (those with positive WOE values) were added to the bank system from 2015 to 2017.

Table 5.5 summarises the IV value, which equals 0.67. This IV value reflects the strong predictive power of this variable according to the classification table's predictive power (Table 5.2). Hence, the registration date is classified as a strong exploratory variable in the empirical lending model (Equation 5.1).

3. Variable 3: Industry Class

The industry class (industry_4) variable indicates in which industry the borrower is employed. There are 24 different industries; however, only five industries have default borrowers. They were 04 – processing; 06 – construction; 07 – commercial; 12 - asset management and consultancy services; and 18 - services for individual and society (see Table 5.6). These industries are assigned a numerical value from 1 to 5 in the empirical lending model.

Table 5.6 The WOE and IV Values of the Industry Class Variable

Value	Description	1	0	B (%)	G (%)	WOE	100 x IV
1	07 – Commercial	7	104	0.94%	0.69%	-3.09	0.077
2	18 - Services for individual and society	735	14 717	98.39%	97.35%	-0.11	0.011
3	04 - Processing	1	32	0.13%	0.21%	4.58	0.036
4	06 – Construction	3	104	0.40%	0.69%	5.38	0.154
5	Others	1	161	0.13%	1.06%	20.74	1.931
	Total	747	15 118	100%	100%		**2.209**

Source: Author's computation

Table 5.6 shows the WOE and IV value of the industry class variable. The WOE is negative (which indicates higher risk borrowers) in two industries, 07 – commercial and 18 - services for individuals and society. In contrast, it is positive (lower risk borrowers) for workers in 04 – Processing, 06 – Construction and others.

Table 5.6 summarises the IV value, which is equal to 0.022. This IV value reflects the weak predictive power of this variable according to the classification table's predictive power (Table 5.2). Hence, industry class is classified as a weak exploratory variable in the empirical lending model (Equation 5.1).

4. Variable 4: Supplement Class

The independent variable supplement class (Subclass_5) explains, in detail, the industry in which the borrowers work. There are 50 types of supplement classes; however, only eight supplement classes contain bad borrowers. In the empirical lending model, these supplement classes are assigned a numerical value from 1 to 8.

Table 5.7 The WOE and IV Values of the Supplement Class Variable

Value	Description	1	0	B (%)	G (%)	WOE	100xIV
1	Raw material wholesale	1	5	0.13%	0.03%	-13.98	0.141
2	Host operator	1	6	0.13%	0.04%	-12.16	0.115
3	Other retail	4	26	0.54%	0.17%	-11.36	0.413
4	Other products	1	12	0.13%	0.08%	-5.23	0.028
5	Construction materials wholesale	1	17	0.13%	0.11%	-1.74	0.004
6	Others	735	14 952	98.39%	98.90%	0.05	0.003
7	Construction	3	69	0.40%	0.46%	1.28	0.007
8	Commerce	1	31	0.13%	0.21%	4.26	0.030
	Total	747	15 118	100%	100%		**0.740**

Source: Author's computation

Table 5.7 presents the WOE and IV values of the supplement class variable. The WOE is negative for raw material wholesale, host operator, other retail, other products and construction materials wholesale in the supplement class variable. Conversely, it is positive for construction, commerce and others. These numbers can be used to identify which supplement class are risky (those with a negative WOE value) as well as safe borrowers (those with a positive WOE value).

Table 5.7 summarises the IV value which is equal to 0.0074. This IV value reflects the unpredictive power of the variable according to the classification table's predictive power (Table 5.2). Hence, the supplement class variable is classified having unpredictive power as an exploratory variable and is removed from the empirical lending model.

5. Variable 5: Economic Class

The economic class variable (economic_6) explains the different types of companies in which the borrowers work. Three of eight types of company have default borrowers. These are: 04 - company limited, 08 - private enterprise and others (see Table 5.8). In the empirical lending model, these three types of company are assigned a numerical value from 1 to 3.

Table 5.8 The WOE and IV Values of the Economic Class Variable

Value	Description	1	0	B (%)	G (%)	WOE	100xIV
1	04 - Company limited	11	211	1.47%	1.40%	-0.54	0.004
2	08 - Private enterprise	735	14 734	98.39%	97.46%	-0.10	0.008
3	Others	1	173	0.13%	1.14%	21.46	2.168
	Total	747	15 118	100%	100%		**2.181**

Source: Author's computation

Table 5.8 shows that the WOE is negative for company limited and private enterprise, and positive for other types of company. These results indicate which types of company are considered risky (those with a negative WOE value) as well as those which are considered less risky (those which have a positive WOE value)

Table 5.8 provides a total of IV value which equals 0.021. This total IV value reflects the variable's weak predictive power as outlined in the classification table's predictive power (Table 5.2). Hence, economic class is considered a weak exploratory variable in the empirical lending model (Equation 5.1).

6. Variable 6: Loan Value

The loan value (Loan_value_9) refers to the amount borrowed. This variable has a continuous number which ranges from 1946 VND to 420 million VND. In the empirical lending model, loan value is categorised into 11 groups based on the number of bad borrowers (see Table 5.9). The 11 groups are assigned a numerical value from 1 to 11 in the empirical lending model.

Table 5.9 The WOE and IV Values of the Loan Amount Variable

Value	Loan Value	1	0	B (%)	G (%)	WOE	100xIV
1	Loan value ≤ 10 mil. VND	50	291	6.69%	1.92%	-12.46	5.942
2	10 < Loan value < 20 mil. VND	32	461	4.28%	3.05%	-3.40	0.419
3	20 < Loan value ≤ 50 mil. VND	112	2 199	14.99%	14.55%	-0.30	0.013
4	50 < Loan value ≤ 100 mil. VND	133	4 405	17.80%	29.14%	4.93	5.582
5	100 < Loan value ≤ 200 mil. VND	93	1 802	12.45%	11.92%	-0.44	0.023
6	200 < Loan value ≤ 500 mil. VND	177	3 034	23.69%	20.07%	-1.66	0.602
7	500 mil VND < Loan value ≤ 1 bil VND	60	1 573	8.03%	10.40%	2.59	0.614
8	1 < Loan value ≤ 2 bil. VND	46	683	6.16%	4.52%	-3.10	0.508
9	2 < Loan value ≤ 5 bil. VND	34	427	4.55%	2.82%	-4.77	0.824
10	5 < Loan value ≤ 10 bil. VND	9	142	1.20%	0.94%	-2.49	0.066
11	> 10 bil. VND	1	101	0.13%	0.67%	16.08	0.858
	Total	747	15 118	100%	100%		**15.454**

Note: mil VND = million VND; bil = billion VND;
Source: Author's computation

Table 5.9 shows that the WOE is negative when the loan amount is less than 50 million VND; from 100 to 500 million VND; and from one billion to 10 billion VND. In other words, these loan amounts are considered risky. In contrast, safe borrowers (those with a positive WOE value) obtain loans from 50 to 100 billion VND; 500 million to one billion VND and larger than 10 billion VND.

Table 5.9 provides a total of IV value which equals 0.15. This total IV value reflects the variable's medium predictive power as outlined in the classification table of predictive power (Table 5.2). Therefore, loan amount is considered a medium exploratory variable in the empirical lending model (Equation 5.1).

7. Variable 7: Credit Terms

The loan credit term (MIA_type_12) variable refers to a borrower's credit terms. Thirteen of 37 credit terms have bad borrowers. These are: overdue loan 1, medium-term VND; overdue loan 1, interest first payment; overdue loan 1 , medium-term AMORT VND; overdue loan 2, long term and others;

medium-term loan AMORT VND; overdraft loan short-term VND; medium-term interest at first payment VND; long term interest at first payment VND; medium-term mortgage loan VND; short-term loan, interest first payment; unsecured staff loan VND; long term loan for bank staff VND and short-term mortgage loan VND (see Table 5.10). In the empirical lending model, these types of loan credit term are assigned a numerical value of 1 to 13.

Table 5.10 The WOE and IV Values of the Credit Terms Variable

Value	Credit Term	1	0	B (%)	G (%)	WOE	100 x IV
1	Overdue loan 1, medium term VND	63	1	8.43%	0.01%	-71.51	60.260
2	Overdue loan 1, interest first payment	249	19	33.33%	0.13%	-55.81	185.318
3	Overdue loan 1, medium term AMORT	12	1	1.61%	0.01%	-54.92	8.787
4	Overdue loan 2, long term and others	214	752	28.65%	4.97%	-17.51	41.448
5	Medium term loan AMORT VND	1	35	0.13%	0.23%	5.48	0.053
6	Overdraft loan short-term VND	2	73	0.27%	0.48%	5.90	0.127
7	Medium term interest at first payment	71	3 296	9.50%	21.80%	8.30	10.209
8	Long term interest at first payment VND	113	5 936	15.13%	39.26%	9.54	23.023
9	Medium-term mortgage loan VND	2	124	0.27%	0.82%	11.20	0.619
10	Short-term loan, interest first payment	10	1 457	1.34%	9.64%	19.74	16.382
11	Unsecured staff loan VND	3	432	0.40%	2.86%	19.62	4.819
12	Long term loan for bank staff VND	2	450	0.27%	2.98%	24.09	6.524
13	Short-term mortgage loan VND	5	2 542	0.67%	16.81%	32.24	52.047
	Total	747	15 118	100%	100%		**409.616**

Source: Author’s computation

Table 5.10 shows that the WOE is negative in almost every case of overdue loans for the first time/in the medium/long term. These figures indicate that overdue loans (first time loans) are risky loans. In contrast, other term loans such as short, medium- and long-term loans with interest first payments; overdraft short-term loans; unsecured and long-term loans for staff and short-term mortgage loans are considered less risky loans because the WOE value is positive.

Table 5.10 summarises the IV value which is equal to 4.09. This summary of IV value reflects the variable’s strong predictive power as outlined in the classification table of predictive power (Table 5.2). Hence, the credit terms variable is considered a strong exploratory variable in the empirical lending model (Equation 5.1).

8. Variable 8: Interest Rate

The interest rate (interest_13) variable refers to the interest rate charged by the bank. Interest rates are divided into 12 groups. Group 1 has an interest rate less than 9%. Each subsequent group represents a 1% increase on the previous group. Group 12 contains loans with interest rates over 19% (see Table 5.11). In the empirical lending model, groups 1 to 12 are assigned a numerical value from 1 to 12.

Table 5.11 shows that the WOE value is positive when the interest rate charged is less than 11%. In contrast, when the interest rate is higher than 11%, the WOE value is negative. This reflects an increase in the risk of default borrowers when the interest rate is higher than 11%.

Table 5.11 summarises the IV value which is equal to 1.92. This summary IV value reflects the variable's strong predictive power as outlined in the classification table of predictive power (Table 5.2). Hence, the interest rate charged is considered a strong exploratory variable in the empirical lending model (Equation 5.1).

Table 5.11 The WOE and IV Values of the Interest Rate Variable

Value	Interest Rate (%)	1	0	B (%)	G (%)	WOE	100 x IV
1	Interest rate ≤ 9%	56	5 038	7.50%	33.32%	14.92	38.531
2	9% < Interest rate ≤ 10%	44	2 243	5.89%	14.84%	9.24	8.265
3	10% < Interest rate ≤ 11%	136	3 657	18.21%	24.19%	2.84	1.700
4	11% < Interest rate ≤ 12%	171	3 330	22.89%	22.03%	-0.39	0.033
5	12% < Interest rate ≤ 13%	77	686	10.31%	4.54%	-8.21	4.735
6	13% < Interest rate ≤ 14%	37	89	4.95%	0.59%	-21.30	9.296
7	14% < Interest rate ≤ 15%	51	24	6.83%	0.16%	-37.61	25.083
8	15% < Interest rate ≤ 16%	36	18	4.82%	0.12%	-37.01	17.394
9	16% < Interest rate ≤ 17%	27	4	3.61%	0.03%	-49.17	17.643
10	17% < Interest rate ≤ 18%	59	6	7.90%	0.04%	-52.93	41.598
11	18% < Interest rate ≤ 19%	34	7	4.55%	0.05%	-45.88	20.670
12	Interest rate > 19 %	19	16	2.54%	0.11%	-31.79	7.750
	Total	747	15 118	100%	100%		**192.698**

Source: Author's computation

9. Variable 9: Opening Date

The opening date (open_date_14) variable refers to the date when the loan amount is released to borrowers. The opening date covers from 27 November 2007 to 30 June 2017. This variable is divided into seven groups: group 1 had their loans released in 2011. The number increases by one year until group 7, whose loans were released in 2017 (see Table 5.12). In the empirical lending model, these seven groups are given in a numerical value from 1 to 7.

Table 5.12 shows that the WOE value is negative from 2011 to 2016 and is positive in 2017. These outcomes indicate that current borrowers (in 2017) are safer than customers who received loans before this date (from 2011 to 2016) because of the positive value of the WOE.

Table 5.12 summarises the IV value which is equal to 2.84. This figure reflects the variable's strong predictive power, based on the classification table for predictive power (Table 5.2). Hence, opening date is considered a strong exploratory variable in the empirical lending model (see Equation 5.1).

Table 5.12 The WOE and IV Values of the Opening Date Variable

Value	Year	1	0	B (%)	G (%)	WOE	100 x IV
1	2011	94	10	12.58%	0.07%	-52.48	65.696
2	2012	118	4	15.80%	0.03%	-63.92	100.802
3	2013	73	34	9.77%	0.22%	-37.72	36.010
4	2014	33	177	4.42%	1.17%	-13.28	4.312
5	2015	90	1 127	12.05%	7.45%	-4.80	2.205
6	2016	199	2 826	26.64%	18.69%	-3.54	2.815
7	2017	140	10 940	18.74%	72.36%	13.51	72.442
	Total	747	15 118	100.00%	100.00%		**284.281**

Source: Author's calculation

10. Variable 10: Loan Duration

The loan duration (duration_16) variable displays the total amount of time that borrowers have to repay their loan. Loan durations are grouped into 12 groups: group 1 is loans that must be repaid in 1 year, and group 12 is loans with durations longer than 15 years (see Table 5.13). In the empirical lending model, these 12 groups are assigned a numerical value from 1 to 12.

Table 5.13 The WOE and IV Values of the Loan Duration Variable

Value	Time	1	0	B (%)	G (%)	WOE	100xIV
1	Duration ≤ 1 year	44	4 503	5.89%	29.79%	16.21	38.728
2	1 year < Duration ≤ 2 years	22	242	2.95%	1.60%	-6.10	0.820
3	2 years < Duration ≤ 3 years	123	451	16.47%	2.98%	-17.08	23.032
4	3 years < Duration ≤ 4 years	51	247	6.83%	1.63%	-14.30	7.427
5	4 years < Duration ≤ 5 years	233	3 152	31.19%	20.85%	-4.03	4.166
6	5 years < Duration ≤ 6 years	4	112	0.54%	0.74%	3.25	0.067
7	6 years < Duration ≤ 7 years	69	934	9.24%	6.18%	-4.02	1.230
8	7 years < Duration ≤ 8 years	4	74	0.54%	0.49%	-0.90	0.004
9	8 years < Duration ≤ 9 years	1	45	0.13%	0.30%	7.99	0.131
10	9 years < Duration ≤ 10 years	59	1 790	7.90%	11.84%	4.05	1.596
11	10 years < Duration ≤ 15 years	45	1 576	6.02%	10.42%	5.48	2.413
12	Duration > 15 years	92	1 992	12.32%	13.18%	0.68	0.058
	Total	747	15 118	100.00%	100.00%		**79.672**

Source: Author's calculation

Table 5.13 shows that the WOE value is negative when the loan duration is between 1 and 8 years. This indicates that risky borrowers have medium-term loans. In contrast, the WOE is positive when the duration is less than one year and more than eight years. This finding reflects that short-term (duration < 1 year) and long-term (duration > 5 years) loans are less risky.

Table 5.13 summarises the IV value which equals 0.79. This number reflects the strong predictive power of this variable based on the classification table for predictive power (Table 5.2). Hence, loan duration is a strong exploratory variable in the empirical lending model (Equation 5.1).

11. Variable 11: Loan Purpose

Table 5.14 The WOE and IV Values of the Loan Purpose Variable

Value	Purpose	1	0	B (%)	G (%)	WOE	100xIV
1	1229 - Buy and sell land	4	1	0.54%	0.01%	-43.94	2.324
2	0420 - Consumer goods, office equipment	24	47	3.21%	0.31%	-23.35	6.777
3	0423 - Infrastructure for business|0732 - Loan for buying house and business	3	6	0.40%	0.04%	-23.14	0.838
4	0410 - Loan in food processing	3	7	0.40%	0.05%	-21.60	0.768
5	1861 - Instalment loans for other bank staff	13	35	1.74%	0.23%	-20.17	3.043
6	0712 - Consumer goods business|0714 - Construction material business	39	127	5.22%	0.84%	-18.27	8.004
7	0716 - Raw material business	3	10	0.40%	0.07%	-18.04	0.605
8	0823 - Infrastructure for business	6	23	0.80%	0.15%	-16.64	1.083
9	0490 - Others producing|0820 - Restaurant operating|0910 - Transportation operations|1890 - Wedding loan	43	202	5.76%	1.34%	-14.61	6.456
10	0723 - Constructing infrastructure	4	23	0.54%	0.15%	-12.58	0.482
11	0810 - Hotel operating	1	7	0.13%	0.05%	-10.62	0.093
12	0719 - Other business	32	253	4.28%	1.67%	-9.40	2.453
13	0711 - Food business	9	73	1.20%	0.48%	-9.14	0.660
14	1833 - Loans for buying house and sale|1835 - Loans for preparing house and business	15	143	2.01%	0.95%	-7.53	0.800
15	0713 - Transportation equipment business	2	21	0.27%	0.14%	-6.56	0.085
16	1810 - Instalment loan in consumer goods	4	49	0.54%	0.32%	-5.02	0.106
17	1830 - Preparing house with repayment source by salary	100	1 471	13.39%	9.73%	-3.19	1.167
18	0715 - Medical equipment |1815 - Buying house with repayment source different from salary|1899 - Other purpose	105	1 637	14.06%	10.83%	-2.61	0.842
19	0710 - Commerce|1831 - Preparing house with repayment source different from salary|1870 - Instalment loans for consuming	257	4 666	34.40%	30.86%	-1.09	0.384
20	1834 - Loans for preparing house and sale	5	117	0.67%	0.77%	1.45	0.015
21	1832 - Loans for buying house and sale	1	24	0.13%	0.16%	1.70	0.004
22	1822 - Buying instalment house with repayment source different from salary	6	185	0.80%	1.22%	4.21	0.177
23	1820 - Buying instalment house with repayment source by salary|1813 - loan for buying a car	22	793	2.95%	5.25%	5.77	1.328
24	1860 - Loan for the bank staff|1814 - Buying house with repayment source by salary	37	1 909	4.95%	12.63%	9.36	7.182
25	1850 - Mortgage loans by saving passbook	9	3 289	1.20%	21.76%	28.94	59.464
	Total	747	15 118	100%	100%		**105.14**

Source: Author's computation

The loan purpose (purpore_20) variable explains the specific purpose of borrowing. Twenty-five of 77 loan purposes have bad borrowers (see Table 5.14). These loan purposes are given a numerical value from 1 to 25 in the empirical lending model

Table 5.14 shows that the WOE value is negative when the loan purpose is located in the first to nineteenth row. In contrast, the WOE value is positive when the loan purpose is from the twentieth to the last row of the table. Most of the less risky borrowers have loan purposes related to repairing or buying a house with repayments to be made from one's salary (see Table 5.14 from the tewntieth to the last row).

Table 5.14 provides a total IV value of 1.05. This figure reflects the strong predictive power of this variable based on the classification table for predictive power (Table 5.2). Therefore, loan purpose is considered a strong exploratory variable in the empirical lending model (Equation 5.1).

12. Variable 12: Loan Purpose Group

Loan purposes are categorised into business sectors; seven sectors are shown in Table 5.15. In the empirical model, these purpose sectors are assigned a numerical value from 1 to 7.

Table 5.15 The WOE and IV Values of the Loan Purpose Group Variable

Value	Purpose Group	1	0	B (%)	G (%)	WOE	100xIV
1	04. Technology processing	58	196	7.76%	1.30%	-17.90	11.577
2	09. Transportation, warehouse and communication	8	37	1.07%	0.24%	-14.76	1.220
3	08. Hotel and restaurant	12	56	1.61%	0.37%	-14.67	1.813
4	02. Fisheries	2	11	0.27%	0.07%	-13.03	0.254
5	12. Asset management, consultancy	5	27	0.67%	0.18%	-13.21	0.648
6	07. Trade, car repairing and consumption	139	1 481	18.61%	9.80%	-6.42	5.653
7	18. Serve for individual and society	523	13 310	70.01%	88.04%	2.29	4.130
	Total	747	15 118	100.%	100.%		**25.296**

Source: Author's computation

Table 5.15 provides a total IV value of 0.25. This indicates that the variable has medium predictive power as shown in the classification table for predictive power (Table 5.2). Therefore, loan purpose is considered a medium exploratory variable in the empirical lending model (Equation 5.1).

13. Variable 13: Lending Specific Programmes

Vietnam banks have some special programmes that work in tandem with real estate. A home buyer may buy a house or apartment with an advanced payment of 30% of the total value. The owner can borrow and finance the remaining 70% from commercial banks. The bank and developer gain mutual benefit. The bank knows the number of customers beforehand and the developer can attract more customers. Twelve of 30 specific programmes appear to be bad borrowers. These specific programmes are given numerical value from 1 to 12 in the empirical lending model.

Table 5.16 shows that the WOE value is positive in the lending programme for repairing a house, in cooperation with the Hung Thinh group, the diamond interest rate and acceleration programme for buying a car (ninth to twelfth row). Other lending programmes, such as the 1 trillion VND programme, incentive programmes for SME (first to ninth rows, Table 5.16) have negative WOE values.

Table 5.16 provides a total value of 0.25. This figure reflects the weak predictive power of this variable according to the classification table for predictive power (Table 5.2). Therefore, the lending programme (HTCV_21) is considered a weak exploratory variable in the empirical lending model (Equation 5.1).

Table 5.16 The WOE and IV Values of the Lending Programme Variable

Value	Program	1	0	B (%)	G (%)	WOE	100xIV
1	1 trillion VND programme	2	9	0.27%	0.06%	-15.03	0.313
2	Incentive programme for SME	5	25	0.67%	0.17%	-13.98	0.705
3	Incentive buying car purpose	4	49	0.54%	0.32%	-5.02	0.106
4	Mortgage loan for consumer lending	18	262	2.41%	1.73%	-3.30	0.223
5	Incentive program in 06 months	6	93	0.80%	0.62%	-2.67	0.050
6	Collaboration loan with Gia Hoa group	8	134	1.07%	0.89%	-1.89	0.035
7	Incentive programme for buying house	29	496	3.88%	3.28%	-1.68	0.101
8	Others	600	11 767	80.32%	77.83%	-0.31	0.078
9	Incentive home repaired programme	32	806	4.28%	5.33%	2.19	0.229
10	Collaboration loan with Hung Thinh group	6	153	0.80%	1.01%	2.31	0.048
11	Loan with diamond interest	34	1206	4.55%	7.98%	5.61	1.922
12	Acceleration programme for buying car	3	118	0.40%	0.78%	6.64	0.252
	Total	747	15 118	100%	100%		**4.063**

Source: Author's computation

14. Variable 14: Individual Credit

The individual credits (TDCN_22) variable refers to specific credits for individual lending. Fifteen of 36 individual creditors are bad borrowers (see Table 5.17). In our empirical model, these individual credits are assigned a numerical value from 1 to 15.

Table 5.17 shows the WOE of the individual credit variable. The riskiest is "consumer loans without a mortgage" because it has the highest negative WOE value (-20.44). In contrast, "mortgage loan by saving passbook" is the safest group with the highest positive WOE (32.2).

Table 5.17 provides a total IV value that is 1.345. This number reveals the strong predictive power of the variable according to the classification table for predictive power (Table 5.2). Therefore individual credit (TDCN_22) is a strong exploratory variable in the empirical lending model (Equation 5.1).

Table 5.17 The WOE and IV Values of the Individual Credit Variable

Value	Program	1	0	B (%)	G (%)	WOE	100 x IV
1	Consumer loans w/o mortgage	79	207	10.58%	1.37%	-20.44	18.821
2	Loan for real estate trading	14	47	1.87%	0.31%	-17.96	2.808
3	Instalment loan for business	142	637	19.01%	4.21%	-15.07	22.292
4	Others	43	224	5.76%	1.48%	-13.57	5.801
5	Overdraft loan	6	60	0.80%	0.40%	-7.05	0.286
6	Consumer loans with mortgage	167	2 667	22.36%	17.64%	-2.37	1.117
7	Loan for business	69	1 330	9.24%	8.80%	-0.49	0.021
8	Loan for buying house	79	1 898	10.58%	12.55%	1.72	0.339
9	Loan for repairing house	101	2 709	13.52%	17.92%	2.82	1.239
10	Loan for future house, mortgage for that house	28	779	3.75%	5.15%	3.18	0.447
11	Car loan	5	203	0.67%	1.34%	6.96	0.469
12	Instalment loan for car	2	152	0.27%	1.01%	13.23	0.976
13	Unsecured loan for staff	4	397	0.54%	2.63%	15.90	3.324
14	Certify financial capacity for abroad travelling by passing book	1	263	0.13%	1.74%	25.65	4.118
15	Mortgage loan by saving passbook	7	3 545	0.94%	23.45%	32.20	72.484
	Total	747	15 118	100.00%	100.00%		**134.543**

Source: Author's computation

15. Variable 15: Business Credit

The business credit (TDDN_23) variable explains specific business credit such as stock trading. There is only one of two business credit groups with bad borrowers: loans for business (see Table 5.18).

Table 5.18 provides a total IV value that is 0.0004. This number reveals the unpredictive power of the variable according to the classification table for predictive power (Table 5.2). Therefore, individual credit (TDCN_23) is considered an unpredictive exploratory variable and is removed from our empirical lending model (Equation 5.1).

Table 5.18 The WOE and IV Values of the Business Credit Variable

Value	Program	1	0	B (%)	G (%)	WOE	100xIV
1	Stock trading and others	1	10	0.13%	0.07%	-7.05	0.048
2	Loan for business	746	15 108	99.87%	99.93%	0.01	0.000
	Total	747	15 118	100.00%	100.00%		**0.048**

Source: Author's computation

16. Variable 16: Sponsor Programmes

The sponsor programmes variable (TDDN_24) mentions specific sponsor programmes offered by banks for local development. Three of the seven sponsor programmes have bad borrowers. In the empirical lending model, these three sponsor programmes are assigned a numerical value from 1 to 3.

Table 5.19 The WOE and IV Values of the Sponsor Programme Variable

Value	Sponsor Program	1	0	B (%)	G (%)	WOE	100xIV
1	a. Local trading sponsor	242	1 867	32.40%	12.35%	-9.64	19.334
2	h. Loan for investment fixed asset	5	123	0.67%	0.81%	1.95	0.028
3	Others	500	13 128	66.93%	86.84%	2.60	5.181
	Total	747	15 118	100.00%	100.00%		**24.543**

Source: Author's computation

Table 5.19 shows the riskiest group is "local trading sponsor" because it has the greatest negative WOE value (-9.64). The total IV value is 0.24. This reflects medium predictive power of the variable according to the classification table for predictive power (Table 5.2). Therefore, the sponsor programme variable is a medium exploratory variable in the empirical lending model (Equation 5.1).

17. Variable 17: Customers

There is a difference between the registration date value and opening date value for each borrower. The difference between these dates (register date and opening date) indicates that some customers have been added to the bank's information system (e.g., in 2009) even though the loan was not released until 2016. This difference in dates may indicate old and new customers. If the difference in dates (between registration and opening date) is more than 15 days, a customer is classified as an existing customer. In contrast, if the difference in dates is fewer than 15 days, the customer is classified as a new customer. In the empirical lending model, these customers are assigned a numerical value (1 for a new customer and 2 for an old/existing customer).

Table 5.20 The WOE and IV Values of the Customer Variable

Value	Customer	1	0	B (%)	G (%)	WOE	100 x IV
1	New	409	4 946	54.75%	32.72%	-5.15	11.343
2	Old	338	10 170	45.25%	67.28%	3.97	8.740
	Total	747	15 116	100.00%	100.00%		**20.083**

Source: Author's computation

Table 5.20 shows that new customers are riskier than old customers. The WOE value of new customers is negative (-5.15), whereas the WOE of an old customer is positive (3.97).

Table 5.20 provides a total of IV value of 0.2. This reflects medium predictive power of this variable according to classification table for predictive power (Table 5.2). Therefore, the customer variable (Customer_8) is a medium exploratory variable in the empirical lending model (Equation 5.1).

5.3.2.3 A Summary of the Transformed Variables

The IV value indicates the predictive power of transformed variables. If the IV value of a variable is less than 0.02, the variable is classified as having unpredictive ability (Siddiqi, 2006) and is removed from the empirical lending model (Equation 5.1). Table 5.21 summarises all IV values for the transformed variables. It shows that there are two unpredictive power exploratory variables (IV < 0.02): supplements class (subclass_5) and business credit (TDDN_23). They are removed from our dataset because they are not useful for prediction (Anderson, 2007; Kocenda & Vojtek, 2011; Siddiqi, 2006; Upadhyay, 2013). Therefore, we include 14 variables in the empirical lending model (Equation 5.1). Appendix Table D1 summarises the statistical information of the transformed variables.

Table 5.21 A Summary the IV Values of the Independent Variables

No.	Variable		IV Value	Strength	Classify
1	Credit term	MIA_type_12	409.62	Strong	useable
2	Opening date	Open_date_14	284.28	Strong	useable
3	Interest rate	Interest_13	192.70	Strong	useable
4	Individual credit	TDCN_22	134.54	Strong	useable
5	Loan purpose	purpose_20	105.14	Strong	useable
6	Loan duration	duration_16	79.67	Strong	useable
7	Registration date	Reg_date_3	67.01	Strong	useable
8	Purpose group	purpose_group_19	25.30	Medium	useable
9	Sponsor	TOFL_24	24.54	Medium	useable
10	Customer	Customer_8	20.08	Medium	useable
11	Loan value	Loan_value_9	15.45	Medium	useable
12	Lending programs	HTCV_21	4.06	Week	useable
13	Industry class	industry_4	2.21	Week	useable
14	Economic class	economic_6	2.18	Week	useable
15	Supplements class	subclass_5	*0.74*	*Very week*	*Removed*
16	Business credit	TDDN_23	*0.05*	*Very week*	*Removed*

Source: Author's computation

5.3.3 Expected Sign of the Transformed Variables

Several of the variables used in this study have been discussed in prior literature. The following section briefly reviews these other studies.

1. Loan Purpose

The loan purpose variable explains borrowers' purpose for borrowing money. This is a qualitative variable. In other words, every attribute of the variable exhibits a different sign. For example, Dinh and Kleimeier (2007) find that the riskiest loan purpose is business financing whereas the safest is the credit card. Pham and Lensink (2008) find that loans for housing/apartments are less risky than business and consumer loans. Kocenda and Vojtek (2011) note that house purchase loans are safer than loans for house renovation or maintenance. The authors explain that most customers who decide to buy a house have a plan for repaying their financial obligations.

2. Loan Duration

The loan duration variable represents the loan maturity. A longer borrowing period may be a signal of unaffordable short-term repayment capacity and therefore expected higher uncertainty in the future, increasing the probability of default (Van Gool et al., 2012). In short, a positive relationship between loan duration and default borrowers is hypothesised. The positive link between loan duration and default borrowers is supported by prior studies (see for example, Bekhet & Eletter, 2014; Limsombunchai et al., 2005; Nguyen, 2015). However, some authors have found contradictory evidence to the above. Van Gool et al. (2012) argue that long-term (more than 19 months) loans are the safest loans and that medium term (13 – 18 months) loans are the riskiest. Dinh and Kleimeier (2007) find loan duration is negative and significant with respect to the probability of borrower default to Vietnamese commercial banks.

3. Loan Value

The loan value (Loan_value_9) variable refers to the total loan amount requested. As the loan amount increases, an increased probability of borrower default is expected (Van Gool et al., 2012). In other words, there is a positive, significant relationship between loan size and default risk. This result is similar to previous studies: Limsombunchai et al. (2005), for agricultural lending in Thailand; Jagric et al. (2011) in Slovenian banking; Nguyen's work (2015) on banks in China and Van Gool et al. (2012) for Bosnia–Herzegovinian lenders.

Table 5.22 Expected Signs of the Transformed Variables

No.	Variable	Signs	Countries	Sources
1	Loan purpose	Negative/Positive	Vietnam Vietnam Czech Jordan	Dinh & Kleimeier (2007); Pham & Lensink (2008); Kocenda & Vojtek (2011); Bekhet & Eletter (2014)
2	Loan duration	Negative	Vietnam Bonia	Dinh & Kleimeier (2007); Van Gool et al. (2011)
		Positive	Thailand Jordan China	Limsombunchai et al. (2005); Bekhet & Eletter (2014); Nguyen (2015)
3	Loan value	Positive	Thailand Slovenia China	Limsombunchai et al. (2005); Jagric et al. (2011); Nguyen (2015)
4	Customer	Negative	Vietnam Thailand Czech	Dinh & Kleimeier (2007); Limsombunchai et al. (2005); Kocenda & Vojtek (2011)

Source: Author's summary from previous literature

4. Customer

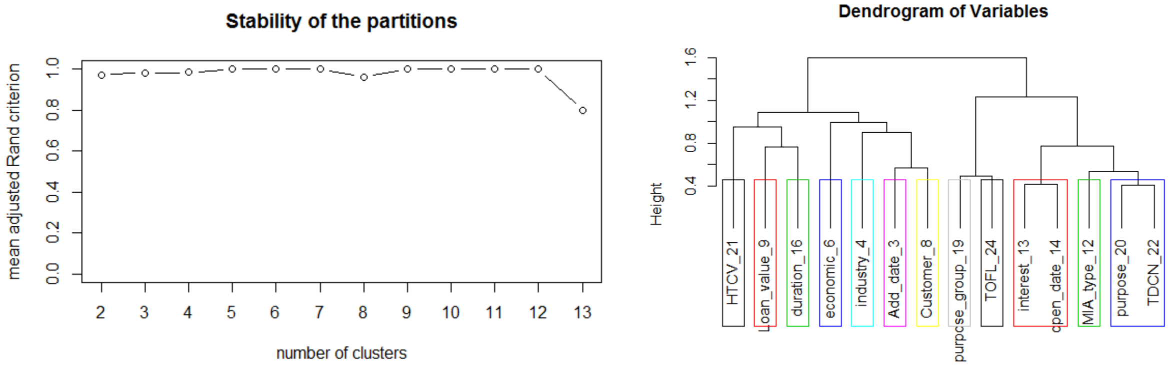
The old or new customer variable indicates a customer's relationship with the bank. The longer a customer stays with a bank, the more the bank identifies him/her as a good borrower (Limsombunchai et al., 2005). In other words, an old customer is a safer borrower. This expectation is similar to results found by Dinh and Kleimer (2007) for Vietnamese commercial banks; Limsombunchai et al. (2005) for agricultural lending; and Kocenda and Vojtek (2011) for Czech commercial banks. Table 5.22 summarises the expected sign for loan purpose, loan duration, loan value, and customer variables:

5.3.4 Diagnostic Tests

5.3.4.1 Homogeneous Clustering Analysis

The dataset is analysed to find homogenous clusters. The software package "Clust0fVAR" was chosen to arrange all the variables into homogeneous clusters (Chavent, Kuentz, Liquet, & Saracco, 2011). This method is beneficial for reducing data dimension and variable selection (Chavent et al., 2011). First, the stability of the partitions is computed using bootstrap 10 times. This step evaluates the number of clusters selected (see Figure 5.1a). The stability of the partition shows that there is a drop in the mean adjusted R value when the number of clusters increases from 12 to 13. Thus, 12 clusters are selected to draw the dendrogram (Figure 5.1b). The dendrogram shows that the *interest rate* (interest_13) and *opening date* (open_date_14) variables in one cluster. The *loan purpose* (purpose_20) and the *individual credits* (TDCN_22) are in another cluster. These outcomes imply that these variables may contain similar information (Chavent et al., 2011).

Figure 5.1 The Stability of the Partitions and a Dendrogram of the Independent Variables of the Transformed Dataset

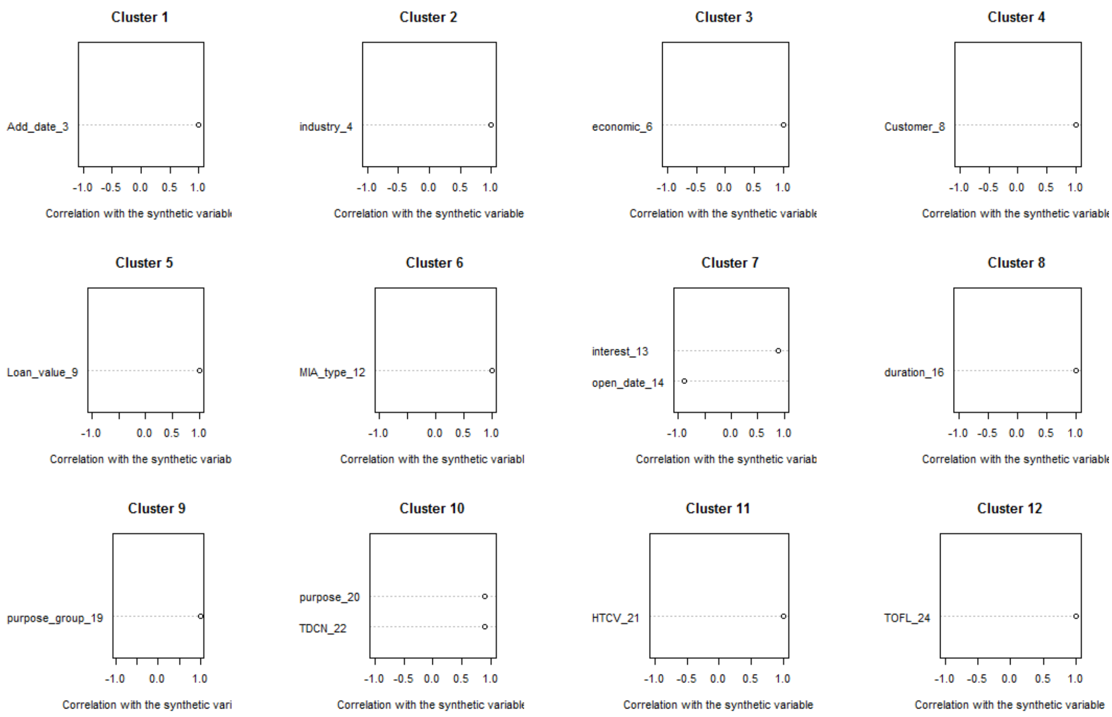


a) b)

Source: Author's calculation

Figure 5.2 shows the correlations between the independent variables within clusters 1 to 12. The results show that the "interest_13" and "open_date_14" variables are grouped into cluster 7 and the "purpose_20" and "TDCN_22" variables are in cluster 10 (see Figure 5.2). In addition, cluster 7 variables "interest_13" and "open_date_14" are negatively correlated. Cluster 10 variables "purpose_20" and "TDCN_22" are positively correlated. These variables can be removed from the dataset because they provide the same information. However, these variables will be re-checked using the coefficient correlation matrix to select uncorrelated variables.

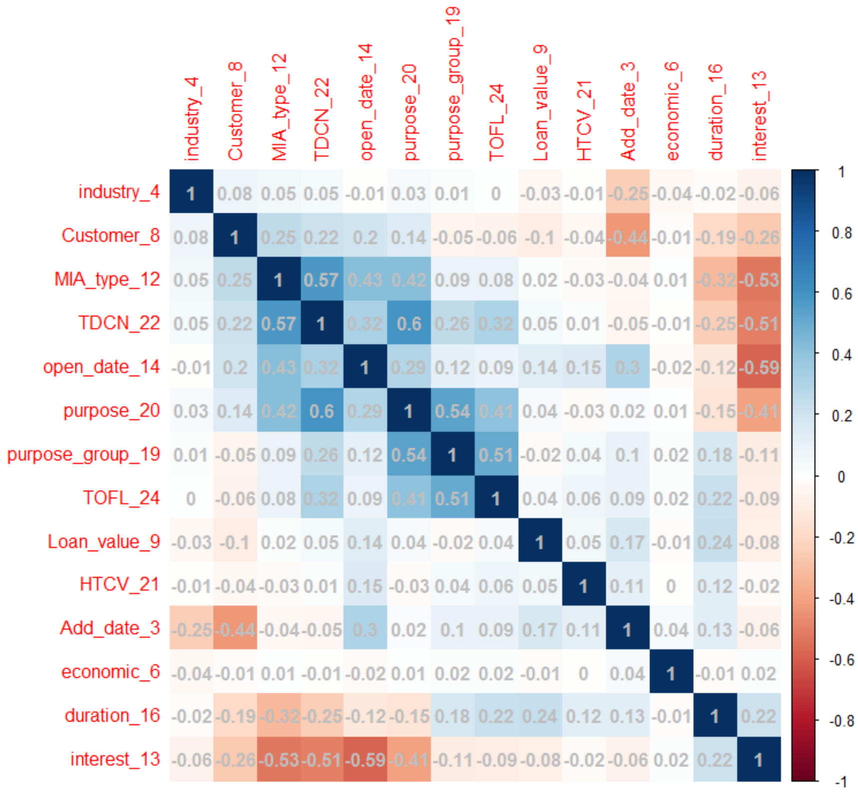
Figure 5.2 The Correlations of the Independent Variables within the Clusters



Source: Author's calculation

5.3.4.2 The Correlation Matrix of the Independent Variables

Figure 5.3 A Correlation Matrix of the Independent Transformed Variables



Source: Author's computation

Figure 5.3 shows the coefficient correlation matrix of independent variables from the transformed dataset. The figure shows that the “interest_13” and “open_date_14” variables are negatively correlated with a value of -0.59. In contrast, the variables “purpose_20” and “TDCN_22” are positively correlated with a value of 0.6. The absolute values of the correlating coefficients are less than 0.7. Du Jardin and Séverin (2012) note that correlation values are acceptable for model development. Therefore, these variables will not be removed from our dataset. A total of 14 independent variables and one dependent variable are used in the empirical lending model (Equation 5.1).

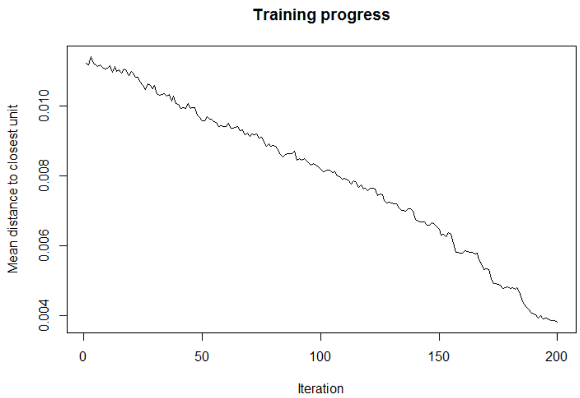
5.3.5 SOM Analysis

The transformed numerical dataset is screened and analysed using the SOM technique. Neuron grids such as 15 x 15, 20 x 20, 25 x 25, 30 x 30, 35 x 35 and 40 x 40 are trained to obtain a minimum mean distance, show the cluster structure clearly and provide the best prediction of ‘bad’ borrowers (Waidyarathne & Samarasinghe, 2014). The topology grid as a “hexagon” is selected because it gives more neighbouring connections. The SOM map is trained in “batch mode” with 200 epochs and a learning rate of 0.05 to 0.1. Appendix Figure D2 gives displays the SOM neighbour distances and Appendix Figure D3 shows the predictions of ‘bad’ borrowers in different SOM grids (from 15 x 15 to 40 x 40 grids). These outcomes show that the 20 x 20 SOM grid can generate the most accurate predictions of ‘bad’ borrowers and depict clearly ordinary clusters. Consequently, the 20x20 SOM grid is used to screen the dataset.

5.3.5.1. The SOM Results with a 20 x 20 Grid

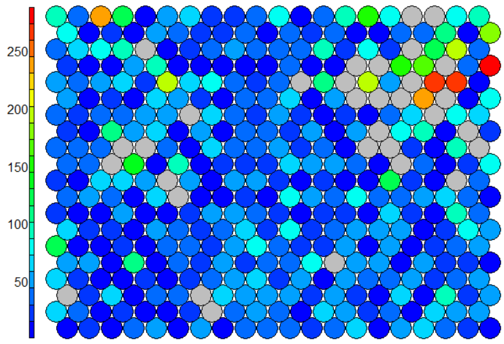
Figure 5.4 shows the training process of the SOM with a 20 x 20 grid in “batch mode” from iteration 1 to 200. The mean distance of neurons to the closest observation drops from 0.012 at iteration 0 to a minimum value of 0.04634 at iteration 200. This outcome indicates that the SOM grid is trained to have the shortest Euclidean distance for the dataset. In other words, the SOM map’s results will reveal any clusters that exist inside (Samarasinghe, 2006).

Figure 5.4 The SOM Training Process

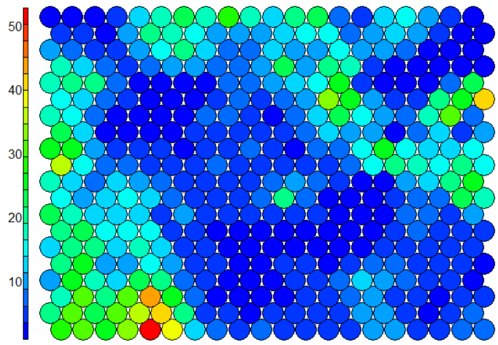


Source: Author’s computation

Figure 5.5 The SOM Results: Observations Hit by Neurons and Neighbour Distances



a) Observations Hit by Neurons



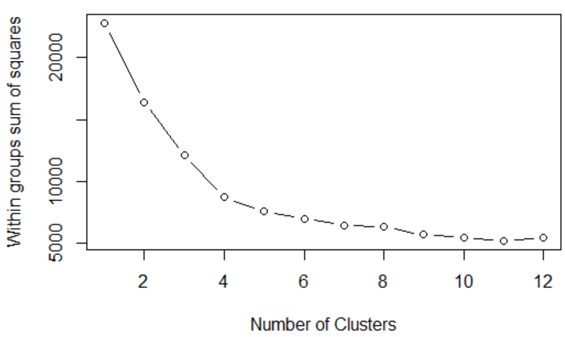
b) SOM Neighbour Euclidean Distance

Source: Author's computation

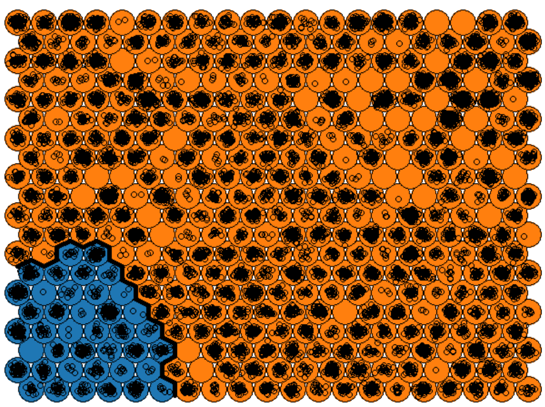
Figure 5.5 also shows the SOM map results that include a map of observations hit by neurons and a map of neighbour distances. The 'observation hit by neurons' (Figure 5.5 a) displays the number of data units captured by neurons. Neurons that contain over 250 units are in *red*; otherwise, almost all neurons hold approximately 50 units. The 'neighbour distances' (Figure 5.5b) demonstrate the distance between these neurons. It has two separate clusters; one is in the top-right corner, which constitutes a quarter of the SOM maps and other is the three-quarter-bottom left of the SOM map.

To find the optimal number of clusters, the within group sum of square (WSS) is calculated depending on the number of clusters (from 1 to 12 clusters). The results are shown in Figure 5.6. The WSS value drops as the number of clusters increases. The WSS value drops sharply when the number of clusters increases from 1 to 2 (Figure 5.6a). Therefore, 2 clusters are selected to compute the U-matrix for the self-organisation maps (Figure 5.6b)

Figure 5.6 The WSS Value Based on the Number of Clusters and U-Matrix Corresponding to Two Clusters



a) WSS value based on Number of Clusters



b) U-matrix Corresponding to 2 Clusters

Source: Author's computation

Table 5.23 lists the number of neurons and the number of observations in groups 1 and 2 in the U-matrix map. Group 1 includes 352 neurons (orange) and group 2 includes 48 neurons (blue) (Figure 5.6b). Group 1 contains 14,708 observations with 554 'bad' borrowers. Group 2 has 193 bad borrowers in 1,157 observations. In other words, though this SOM methodology can predict 'bad' borrowers, it has low accuracy (it estimates approximately 74% of the actual number of bad cases (74% = 554/747) (Table 5.23)).

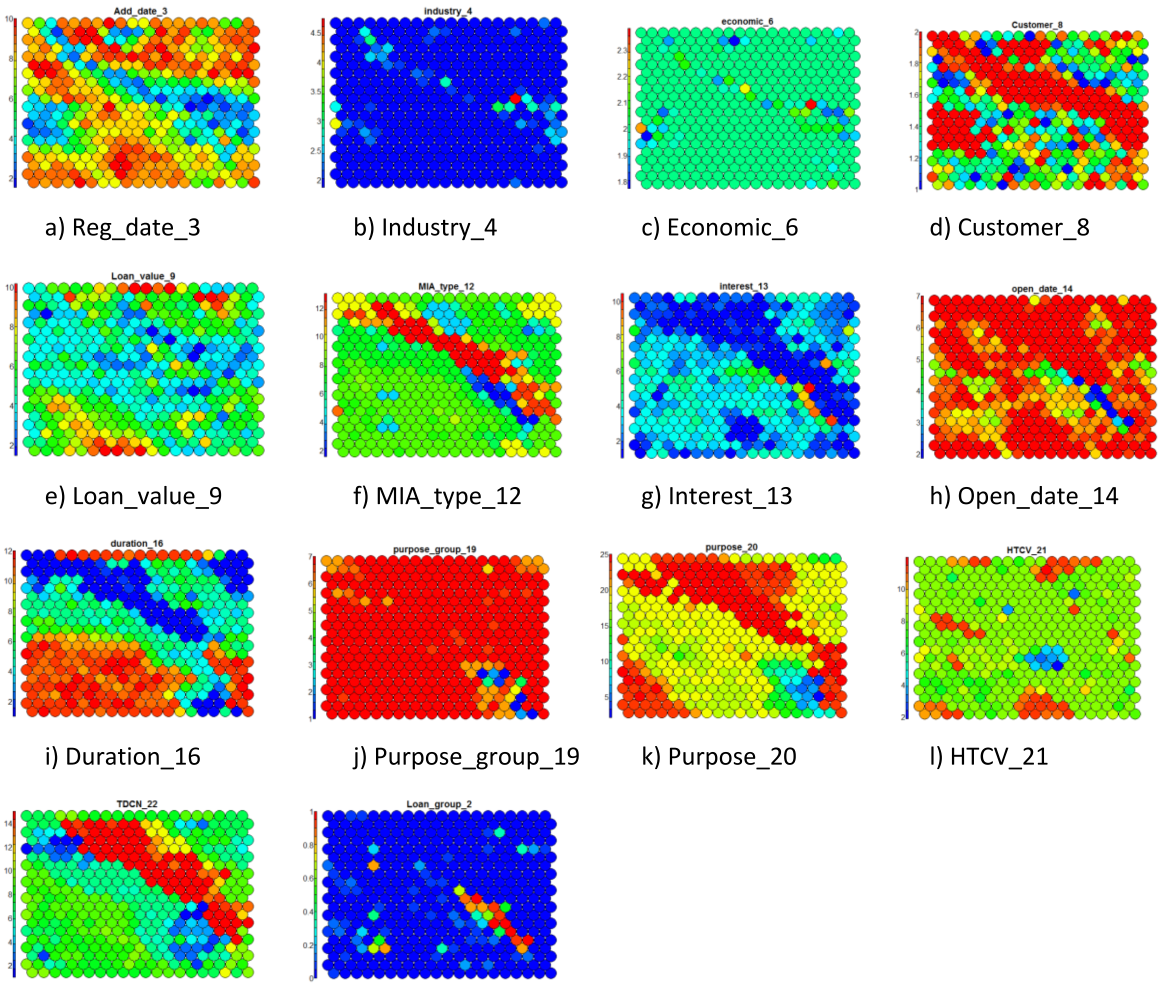
Table 5.23 The Number of Neurons, Observations and Bad Borrowers in Groups 1 and 2

	Neuron Numbers	Observations	Bad Borrowers	Accuracy
Group 1	V7 to V20, V27 to V40, V47 to V60, V67 to 80, V86 to V100, V125 to V140, V141, V142, V145 to V400	14 708	554	74%
Group 2	V1 to V6, V21 to V26, V41 to V46, V61 to V66, V81 to V85, V101 to V105, V121 to V124, V143, V144	1 157	193	
Total	400	15 865	747	

Note: V = neuron number
Source: Author's calculation

5.3.5.2. Component Plane Analysis

Figure 5.7 Component Planes Showing Neural Code Values of the Variables



a) Reg_date_3 b) Industry_4 c) Economic_6 d) Customer_8

e) Loan_value_9 f) MIA_type_12 g) Interest_13 h) Open_date_14

i) Duration_16 j) Purpose_group_19 k) Purpose_20 l) HTCV_21

m) TDCN_22 n) Loan_group_2

Source: Author's computation

Figure 5.7 demonstrates the component planes in each variable. This plot shows the neural weight values that match each variable. These component planes reflect which variables will have a strong or weak effect on the final classified group 1 or group 2. For example, variables MIA_type_12 (see Figure 5.7f.), interest_13 (see Figure 5.7g.) and open_date_14 (see Figure 5.7h.) have patterns similar to the variable Loan_group_2 (see Figure 5.7n). In addition, the correlation coefficients (the SOM codes) between these variables (MIA_type_12, Interest_13 and Open_date_14) and good/or bad borrowers (Loan_group_2) equal -0.5363, 0.6001 and -0.7728, respectively (see Table 5.24). These findings provide evidence that these variables have a significant impact on 'good' and 'bad' borrowers.

Table 5.24 Correlation Coefficients (the SOM Code) between Independent Variables (Good" and "Bad" Variable)

Variable	Correlation Coefficient	Significance at 5% Level
Reg_date_3	-0.2333	*
Industry_4	-0.0645	
Economic_6	0.0269	
Customer_8	-0.1757	*
Loan_value_9	-0.054	
MIA_type_12	-0.5363	*
Interest_13	0.6001	*
Open_date_14	-0.7728	*
Duration_16	-0.0307	
Purpose_group_19	-0.2735	*
Purpose_20	-0.3204	*
HTCV_21	-0.0943	
TDCN_22	-0.3194	*
TOFL_24	-0.2614	*

Source: Author's computation

In conclusion, the SOM technique is an unsupervised learning algorithm that has a strong capacity to present data visually. It assists researchers in understanding the motivation behind the classification by presenting the correlations between the component planes in a visual form (Huysmans et al., 2006). However, the prediction accuracy slightly lower. Therefore, this approach is generally used with other techniques such as supervised learning or statistical approaches (Huysmans et al., 2006). This issue will be investigated in the next section.

5.3.6 Model Development and Results

5.3.6.1 Logit Regression

This study employs logit regression to estimate critical factors that impact borrowers' creditworthiness. This is because logit regression is broadly applied in financial practice and is easy to calculate and interpret (Chi & Hsu, 2012; Crook et al., 2007; Kruppa, Schwarz, Arminger & Ziegler, 2013). Moreover, the technique can help financial institutions reduce the number of 'default' borrowers (Baklouti & Bouri, 2014; Dinh & Kleimeier, 2007; Nguyen, 2015). Finally, West (2000) concludes that logistic regression is the most accurate technique among conventional methods to build a credit scoring model. The empirical lending model is rewritten mathematically as follows:

$$
\begin{aligned}
logit_i = \ln\left(\frac{p_i}{1-p_i}\right) &= \beta_0 + \beta_1 Regdate_3 + \beta_2 industry_4 + \beta_3 economic_6 + \beta_4 customer_8 \\
&+\beta_5 loanValue_9 + \beta_6 MIAtype_{12} + \beta_7 interest_{13} + \beta_8 openDate_{14} + \beta_9 duration_{16} \\
&+\beta_{10} purposeGroup_{19} + \beta_{11} purpose_{20} + \beta_{12} HTCV_{21} + \beta_{13} TDCN_{22} + \beta_{14} TOFL_{24} + \varepsilon
\end{aligned} \quad [5.3]
$$

Where:

p_i = the probability of borrower default;

$\beta_0\ to\ \beta_{14}$ = coefficients to be estimated;

$Regdate_3,\ loanValue_9, interest_{13},\ openDate_{14},\ duration_{16}, industry_4,$ $economic_6,\ customer_8,\ MIAtype_{12},\ purposeGroup_{19}\ , purpose_{20},\ HTCV_{21},$ $TDCN_{22}$, and $TOFL_{24}$ = independent variables (as described in Table 5.1); and

ε = the error term.

5.3.6.2 Dummy Variables

Several variables, $industry_4, economic_6,\ customer_8,\ MIAtype_{12}\ \ purpose_{20},\ purposeGroup_{19},$ $HTCV_{21}, TDCN_{22}, TOFL_{24}$ are *qualitative* variables. To capture the effect of each attribute of the qualitative variables, a binary variable or a dummy variable needs to be generated (Wooldridge, 2012). Equation (5.3) is rewritten mathematically with dummy variables as follows:

$$
\begin{aligned}
& logit_i = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Regdate_3 + \beta_2 loanValue_9 + \beta_3 interest_{13} + \beta_4 openDate_{14} \quad [5.4] \\
& +\beta_5 duration_{16} + \alpha_6 industry_{4.1} + \cdots + \alpha_9 industry_{4.4} + \alpha_{10} economic_{6.1} + \alpha_{11} economic_{6.2} \\
& +\alpha_{12} customer_8 + \alpha_{13} MIAtype_{12.1} + \cdots + \alpha_{24} MIAtype_{12.12} + \alpha_{25} purposeGroup_{19.1} + \cdots \\
& \alpha_{30} purposeGroup_{19.6} + \alpha_{31} purpose_{20.1} + \cdots + \alpha_{54} purpose_{20.24} + +\alpha_{55} HTCV_{21.1} + \cdots \\
& \alpha_{65} HTCV_{21.11} + \alpha_{66} TDCN_{22.1} + \cdots + \alpha_{79} TDCN_{22.14} + \alpha_{80} TOFL_{24.1} + \alpha_{81} TOFL_{24.2} + \varepsilon
\end{aligned}
$$

Where:

p_i = probability of borrower default;

$\beta_0\ to\ \beta_5$ = coefficients to be estimated for the continuous variables;

α_6 to α_{81} = coefficients to be estimated for dummy variables;

$Regdate_3, loanValue_9, interest_{13}, openDate_{14}, duration_{16}$ = independent variables (described in Table 5.1); and

$industry_{4.1...4}, economic_{6.1,6.2}, customer_8, MIAtype_{12.1...12}, purposeGroup_{19.1..6}, purpose_{20.1...24}, HTCV_{21.1...11}, TDCN_{22.1..14}, TOFL_{24.1,2}$ = dummy variables generated from the independent variables in Table 5.1.

5.3.6.3 Empirical Results

Stepwise logit regression method was used to select the critical experimental factors that influence default borrowers. Beginning with an empty model, a variable was added at each step and checked whether it is significant (through the likelihood ratio test and the Wald test) below the 5% level. Table 5.25 shows the critical experimental factors that influence default borrowers through logit regression and stepwise logit regression. Forty-three of 81 variables are statistically significant at the 1% to 5% levels (Chi-square = 3518.3, P-value= 0.000, degree of freedom= 81). The McFadden R^2 of 0.584, 58.4%, also confirms that the model fits well. The marginal effect is computed to examine the most, or least, important variables that impact default borrowers. The marginal effect reflects a change in the probability of default borrowers (either an increase or decrease) given a unit change of one independent variable while holding other variables constant (Wooldridge, 2012).

Table 5.25 The Critical Factors Influencing Default Borrowers

No.	Variable	Estimate	Standard Error	P-value		Marginal Effect
1	Constant	-9.02	2.63	0.001	***	-0.1696
2	Registration date	-0.14	0.04	0.001	***	-0.0026
3	Loan value	0.13	0.04	0.001	***	0.0025
4	Interest rate	0.12	0.05	0.011	**	0.0023
5	Opening date	-0.49	0.08	0.000	***	-0.0092
6	Loan duration	0.10	0.03	0.002	***	0.0018
7	07 - Commercial	4.57	1.93	0.018	**	0.0859
8	18 - Service for individual and society	5.02	2.16	0.020	**	0.0944
9	04 - Company limited	4.37	1.94	0.025	**	0.0821
10	New	0.38	0.16	0.014	**	0.0072
11	Overdue loan 1, medium term VND	4.29	1.46	0.003	***	0.0807
12	Overdue loan 1, medium term AMORT VND	3.47	1.56	0.026	**	0.0653
13	Overdraft loan short-term VND	-3.81	1.55	0.014	**	-0.0717
14	Medium term interest at first payment VND	-4.27	0.95	0.000	***	-0.0802
15	Long term interest at first payment VND	-4.68	0.96	0.000	***	-0.0881
16	Medium-term mortgage loan VND	2.34	0.90	0.010	**	0.0440
17	Short-term loan, interest first payment	-5.37	1.05	0.000	***	-0.1010
18	Long term loan for bank's staff VND	-4.47	1.19	0.000	***	-0.0841
19	1229 - Buy and sell land	10.43	3.60	0.004	***	0.1962
20	0410 - Loan in food processing	4.89	2.38	0.040	**	0.0919
21	0712 - Consumer goods business | 0714 - Construction material business	3.04	0.94	0.001	***	0.0571
22	0719 - Other business	3.49	0.93	0.000	***	0.0657

23	1833 - Loans for buying house and sale|1835 - Loans for preparing house and business	2.13	0.93	0.022	**	0.0401
24	1830 - Preparing house with repayment source by salary	2.42	0.85	0.005	***	0.0454
25	0715 – Medical equipment|1815 - Buying house with repayment source different from salary|1899 - other purpose	3.23	0.83	0.000	***	0.0607
26	0710 - Commerce|1831 - Preparing house with repayment source different from salary|1870 - Instalment loans for consuming	2.79	0.83	0.001	***	0.0525
27	1834 - Loans for preparing house and sale	2.45	0.98	0.013	**	0.0461
28	1832 - Loans for buying house and business	3.06	1.34	0.022	**	0.0576
29	1822 - Buying instalment house with repayment source different from salary	2.46	0.98	0.012	**	0.0463
30	1820 - Buying instalment house with repayment source by salary|1813 - Loan for buying a car	2.83	0.88	0.001	***	0.0531
31	1860 - Loan for bank staff|1814 - Buying house with repayment source by salary	2.42	0.85	0.005	***	0.0454
32	Incentive home repairs program	-2.02	0.86	0.019	**	-0.0380
33	Loan for real estate trading	5.28	0.74	0.000	***	0.0993
34	Instalment loan for business	2.78	0.72	0.000	***	0.0524
35	No credit	2.70	0.66	0.000	***	0.0508
36	Overdraft loan	3.97	1.19	0.001	***	0.0748
37	Consumer loans with mortgage	3.28	0.59	0.000	***	0.0617
38	Loan for business	3.38	0.70	0.000	***	0.0637
39	Loan for buying house	2.49	0.61	0.000	***	0.0469
40	Loan for repairing house	2.97	0.61	0.000	***	0.0558
41	Loan for future house, mortgage by that house	2.91	0.66	0.000	***	0.0547
42	Car loan	2.25	0.87	0.009	***	0.0424
43	Instalment loan for car	2.25	1.07	0.036	**	0.0424
44	a> local trading sponsor	1.06	0.25	0.000	***	0.0199
	Number of Observations:	15865				
	Log likelihood function:	-1252.7	(df=81)			
	Chi-Squared Statistics:	3518.3				
	Degrees of Freedom:	81				
	Prob [ChiSqd > value]:	0				
	McFadden R2:	0.584				
	AIC:	2669.4				
	Hosmer-Lemeshow ChiSqd(8)	54.59				
	Prob > ChiSqd:	0.00				
Note	1/ Dependent variable is bank lending decision (Loan_group_2). 2/ To avoid the singularity problem, a dummy variable is dropped from each group. 3/ ***, **, * denotes 0.1%, 1% and 5% significance levels, respectively.					

Source: Author's calculations

A *negative* sign of the estimators indicates a decrease in the probability of loan default. In other words, the independent variables associated with negative coefficients imply low risk borrowers.

The *register_date* and *opening_date* variables (items 2 and 5 in Table 5.25) are negative and statistically significant in terms of the default rate at the 1% level. Their marginal effects have a value of -0.0026 and -0.0092, respectively. These outcomes indicate that adding a year to the register date and the opening date variable (for example, a one-year increase from 2008 to 2009) would reduce the probability of default by 0.26% and 0.92%, respectively. In other words, borrowers are safer recently. After the 2008 GFC, the Vietnam economy grew stability. This led to increased individual incomes. Borrowers do not want to lose their creditworthiness and are therefore more willing to meet their financial obligations.

Long-term loan for bank staff (item 18) is negative and statistically significant with the probability of default borrowers at the 1% level. Its marginal effect is -0.0841 (see Table 5.25). This result indicates that long-term loans (for bank staff) can decrease the possibility of default at 8.4%. This is a special loan because bank staff members are always screened and closely monitored to ensure they meet their repayment obligations.

Loan for home repairs (item 32) is negatively statistically significant with the probability of default borrowers at the 1% level. This loan has a marginal effect of -0.038. This result implies borrowers who use the money for home repairs are 3.8% less likely to default on their repayment.

An *overdraft short-term loan* (item 13) is considered a 'safe' loan because it is negative and statistically significant with the default rate at 1%. Those who borrow an overdraft short-term loan are 7.17% less likely to default on their repayments because its marginal effect has value of -0.0717. This finding suggests that an individual may need emergency money and s/he may take an overdraft loan and repay later.

The *positive* sign of the estimators indicates an increase in the probability of loan default. In other words, positive statistically significant variables imply risky borrowers.

The *loan value* variable (item 2) is positive and has a statistically significant relationship with the default rate at the 1% level. Its marginal effect has a value of 0.0025. This result suggests that an increase in loan value would increase by 0.25% the chance of borrowers defaulting. This finding echoes previous studies: Limsombunchai et al.'s (2005) study on agricultural lending in Thailand; Jagric et al. (2011) for Slovenian banks; Abdou et al. (2008) for Egyptian banks; Nguyen (2015) and Wang (2010) for Chinese banks. These authors note that larger loan amounts increase default risk.

The *loan duration* (item 6) is positive and has a statistically significant relationship with the default rate at the 1% level. A one-year increase in loan duration would increase the probability of default by 0.18%. This is because its marginal effect has a value of 0.0018. A longer loan duration may be a signal of unaffordable short-term repayment capacity and expected higher uncertainty in future, thus, increasing the probability of the borrower defaulting on repayment (Van Gool et al., 2012). This finding reiterates previous studies (Bekhet & Eletter, 2014; Limsombunchai et al., 2005; Nguyen, 2015)

The *interest rate* variable is also a risk factor. The marginal effect has a value of 0.0023. A 1% increase in the interest rate will increase the default rate by 0.23%. A higher interest rate will increase the bank's profitability. However, increased interest rates may lead to less credit demand and an increase in bad debts. Hence, banks must weigh profitability against the potential losses they may incur through bad debts.

New customers are classified as risky borrowers. This is because the new customer variable (item 10) is positive and statistically significant at the 5% level. Its marginal effect is 0.0072. This result highlights that a new customer has a 0.72% greater chance of defaulting. This finding is not surprising because a customer who has a long-time relationship with the bank is considered a better borrower (Dinh & Kleimeier, 2007; Kocenda & Vojtek, 2011).

Loan for real estate trading (item 33) is considered an unsafe loan. This variable is positive and statistically significant at the 1% level. Its marginal effect value is 0.0993. The value indicates that this loan increases the probability of default by 9.93%. Vietnam's real estate market experienced a bubble in 2008 because of speculation. This led to real estate loans being riskier. Banks must be careful when lending for real estate purchases.

Lastly, *car loan* (item 42) is positive and statistically significant at the 1% level. This result suggests that car loans are unsafe. Recent studies (see Kocenda & Vojtek, 2011; Pham & Lensink, 2008), indicate that car loans are safer than other types of loans. These authors explain that customers who purchase cars, carefully plan how they will repay their financial obligation (Kocenda & Vojtek, 2011). However, our finding suggests that customers who borrow money to purchase a car failed to repay their loans. One possibility is that these borrowers bought a car to lease it. However, the market for rental cars may have crashed and they are thus unable to make repayments.

In summary, the results have identified several low-risk factors (or negative and statistically significant) and high-risk factors (positive and statistically significant) that lead to non-performing loans. Banks needs to consider these factors when creating lending policies. Using this information, banks can identify risky borrowers and charge these customers higher interest rates to compensate for potential future losses.

5.3.6.4 Evaluating the Critical Factors that Impact Default Borrowers using SOM and Logit Regression

Table 5.26 compares the critical factors that are associated with default borrowers as identified by logit regression and SOM. Some critical factors, industry_4, economic_6, Customer_8, MIA_type_12, purpose_group_19, purpose_20, HTCV_21, TDCN_22 and TOFL_24, have been transformed into dummy variables. Therefore, these variables cannot be matched.

Other variables (registration date, loan value, interest rate, opening date and loan duration) are compared. The registration date, interest rate and opening date present the same results for both logit regression and SOM; they are negative and statistically significant (see Table 5.26). However, when compared by these two different techniques, the loan value and loan duration variables are not consistent. These findings reveal that component planes enable researchers to present results visually and this allows them to understand the motivation behind particular classifications. However, SOM is less accurate than logistic regression in identifying risk factors.

Table 5.26 Comparing the Critical Factors using Logit Regression and SOM

No.	Variable	Logit Regression		SOM	
		Coefficient	Significance Level	Correlation Coefficient	Significance Level
1	Registration date	-0.14	***	-0.23	*
2	Loan value	0.13	***	-0.05	
3	Interest rate	0.12	**	0.60	*
4	Opening date	-0.49	***	-0.77	*
5	Loan duration	0.10	***	-0.03	

Note: ***, **, and * denote 0.1%, 1% and 5% significance levels, respectively.
Source: Author's computation

The major shortcoming of logit regression is that its overall classificatory power is less than advanced computing and various hybrid approaches (Lessmann et al., 2015; Trinkle & Baldwin, 2016; Yu et al., 2015). Therefore, after examining the classificatory power of logit regression, in the next section I propose a different approach that reduces the default rate.

5.4 Objective 5: To Investigate The Classificatory Power of The CSM Based on a Hybrid Technique Using a Combination of Artificial Neural Network and Logit Regression.

5.4.1. The Hybrid Technique

5.4.1.1 Model Development

This study uses a hybrid technique that integrates logit regression and ANNs to improve classificatory power. The empirical lending model (Equation 5.1) is rewritten below in the hybrid technique form:

$$\boldsymbol{Lending\ decision} = \boldsymbol{H}(Adddate_3, loanValue_9, interest_{13}, openDate_{14}, duration_{16}, industry_{4.1\cdots4.4}, economic_{6.1,6.2}, customer_8, MIAtype_{12.1\cdots12.12}, purposeGroup_{19.1\cdots19.6}, purpose_{20.1\cdots20.24}, HTCV_{21.1\cdots21.11}, TDCN_{22.1\cdots22.14}, TOFL_{24.1,24.2}, \boldsymbol{\varepsilon}) \quad [5.5]$$

Where:

Lending decision = 1 if loan is in 'default' (bad loan); 0 if loan is repaid (good loan);

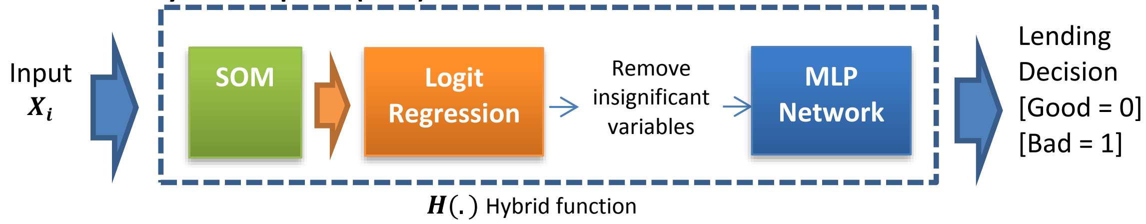
$Regdate_3, loanValue_9, interest_{13}, openDate_{14}, duration_{16}$ = independent variables (described in Table 5.1);

$industry_{4.1...4}, economic_{6.1,6.2}, customer_8, MIAtype_{12.1...12}, purposeGroup_{19.1..6}, purpose_{20.1...24}, HTCV_{21.1...11}, TDCN_{22.1..14}, TOFL_{24.1,2}$ = dummy variables generated from the independent variables (see Table 5.1);

ε = error term and

$\boldsymbol{H}(.)$ = a hybrid function (see Figure 5.8).

Figure 5.8 Hybrid Technique Integrated using the SOM technique, Logit Regression and Multi-Layer Perceptron (MLP)



Source: Author's computation

The hybrid technique processes are as follows:

The first phase involves screening the dataset using SOM. The final 2D map of the SOM neurons delivers a clearer view of the group of cluster input vectors. These results show ordinary clusters that determine outputs (default or non-default borrowers). In other words, this step helps understand the dataset and presents the relationships between inputs and outputs.

Next, the critical characteristics of default borrowers are estimated using logit regression. We expect to find significant variables that determine the probability of default borrowers. The Wald test is used to decide whether to add variables that impact the classification results (Lu et al., 2013).

Finally, insignificant variables are removed from the inputs. Removing insignificant variables helps reduce the noise in the dataset so that it generates better results. The remainder (the significant variables) are used as input variables in the multilayer perceptron (MLP) technique. This technique is a non-linear supervised learning algorithm that improves discriminant power.

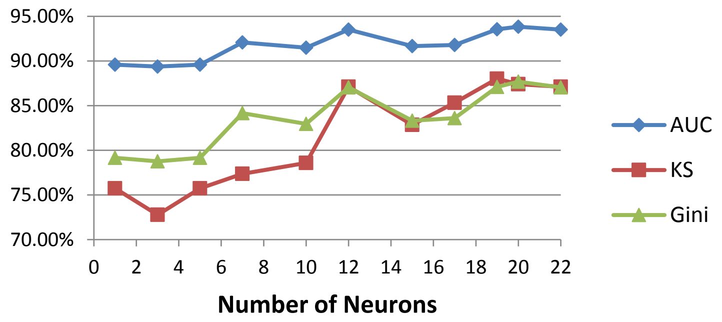
5.4.1.2 Hybrid Empirical Results

Section 5.3.4 provides the 2D SOM map results. Logit regression identified 43 statistically significant variables with levels ranging from 1% to 5% (section 5.3.5). All 43 variables are used as data inputs in MLP.

The neural network has three layers: input, hidden, and output layers. The number of neurons in the input and output layers are known. However, the optimal number of neurons in the hidden layer is ambiguous and needs further research (Samarasinghe, 2016). There are no formal guidelines about how to choose the optimal number of hidden layers nor the number of neurons in the hidden layers.

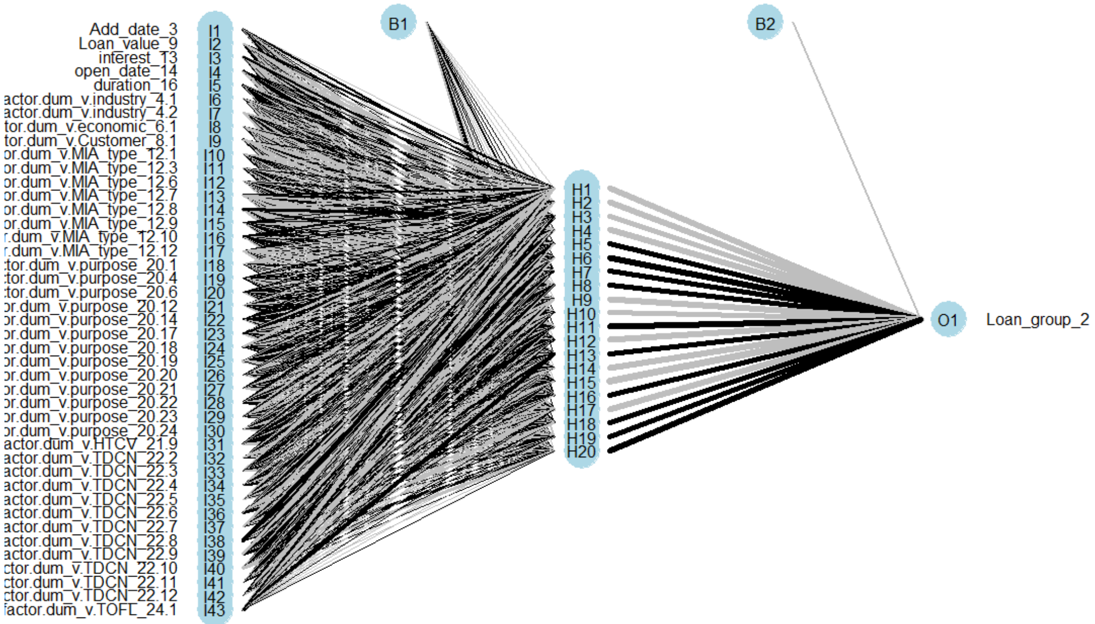
This study selects one hidden layer with a random number of neurons. A random weight value is initially given and the MLP is trained repeatedly with a maximum of 10000 epochs and a learning rate of 0.01.

Figure 5.9 The AUC, KS and Gini Coefficients Generated using MLP with Different Neurons in the Hidden Layer



Source: Author's calculation

Figure 5.10 The MPL with 20 Neurons in the Hidden Layer



Source: Author's computation

Figure 5.9 exhibits the discriminant power of the MLP technique based on the AUC, KS, and Gini coefficients corresponding to the number of neurons in one hidden layer. The AUC, KS and Gini values are optimised when the number of neurons is 20. Hence, we use 20 neurons in the hidden layer to build the MLP networks (see Figure 5.10).

5.4.2. Evaluating the Classification Power of Logit Regression and the Hybrid Technique

The empirical lending model's classificatory power (Equation 5.1) is examined using six criteria: the confusion matrix, the area under the receiver operating characteristic curve (AUC), the Kolmogorov–Smirnov (KS) coefficient, the Gini coefficient, the misclassification cost; and goodness of fit test

(Hosmer-Lemeshow test). These criteria are assessed separately using both the hybrid technique and logit regression as explained below.

5.4.2.1 Confusion Matrix

The 'confusion matrix' shows the actual and predicted values computed using the two techniques (Abdou, 2009; Tsai, 2014). Based on number of factors (overall accuracy (AC rate), sensitivity, specificity, Type I and II errors), the confusion matrix quickly identifies errors and evaluates which technique is better. Table 5.27 shows the confusion matrix for both logit regression (Panel A), the hybrid technique (Panel B), and compares the two techniques (Panel C).

Table 5.27, Panel A shows that the AC rate of the logit technique is 98.02%. The overall incorrect classification rate is 1.98% (314 observations). The sensitivity value indicates the proportion of correct actual goods to total predicted goods has a value of 99.6%. This result indicates that logit regression has identified 99.6% good borrowers correctly and 0.4% incorrectly (Type I error). In contrast, the specificity value is 65.3%. This result indicates that logit regression has correctly identified only 65.3% of bad borrowers, i.e., it made 34.7% errors (Type II). Table 5.27, Panel B shows that the AC rate for the hybrid technique is 99.21%. The overall incorrect classification rate is 0.79% (126 observations). The sensitivity value is 99.9%, and the specificity value is 84.7%. Type I and Type II errors are 0.1% and 15.3%, respectively.

Table 5.27 The Confusion Matrix for the Empirical Lending Model Generated Using Logit Regression and the Hybrid Technique

	Good	Bad	Overall
Panel A: Logit Regression			
Correct	15063	488	15551
	99.6%	65.3%	98.02%
	Sensitivity	Specificity	
Incorrect	55	259	314
	0.4%	34.7%	1.98%
	Type I error	Type II error	
Panel B: The Hybrid Technique			
Correct	15106	633	15739
	99.9%	84.7%	99.21%
	Sensitivity	Specificity	
Incorrect	12	114	126
	0.1%	15.3%	0.79%
	Type I error	Type II error	
Panel C: Comparison			
Correct	0.3%	19.4%	1.2%
	Sensitivity	Specificity	
Incorrect	-0.3%	-19.4%	-1.2%
	Type I error	Type II error	
Number of Observations	15118	747	15865

Note 1/ Sensitivity measures the proportion of correct actual goods to total predicted goods.
2/ Specificity measures the proportion of correct actual bads to total predicted bads.
3/ Type I error measures proportion of incorrect actual goods to total predicted goods.
4/ Type II error measures proportion of incorrect actual bads to total predicted bads.

Source: Author's computation

The hybrid technique has higher overall accuracy (by 1.2%) compared with logit regression (Table 5.27 Panel C). The hybrid technique improves specificity to 84.7% from 65.3% (using logit regression). This technique also has significantly fewer Type II errors (errors of the wrong classification of bad borrowers): 15.3% vs. 34.7%. In short, the hybrid technique outperforms logit regression, particularly in identifying bad borrowers. This outcome is the same as in Lu et al.'s (2013) study.

5.4.2.2 AUC, K-S and Gini Coefficient

Three other criteria, the AUC, KS coefficient and the Gini coefficient, were calculated to measure the classificatory power of both logit regression and the hybrid technique. Table 5.28 shows the results for the logit technique (Panel A) and the hybrid technique (Panel B).

Table 5.28 The AUC, Gini Coefficient and the KS Coefficient Values Generated Using Logit Regression and the Hybrid Technique.

Criterion	Value	Classification
Panel A: Logit Regression		
AUC value	92.64%	Good
Gini coefficient	85.28%	Very good
K-S coefficient	74.10%	Acceptable
Panel B: The Hybrid Technique		
AUC value	93.84%	Good
Gini coefficient	87.41%	Very good
K-S coefficient	87.68%	Acceptable
Number of observations	15865	

Source: Author's computation

Logit regression's and the hybrid technique's AC values are 92.64% (Panel A) and 93.84% (Panel B), respectively. These numbers are greater than 70%. These outcomes indicate that both techniques have good discriminatory power as suggested by Van Gool et al.'s (2012) classification system. The K-S coefficient measures the maximum difference between the cumulative distribution of 'good' and 'bad' (Chi & Hsu, 2012). In short, higher K-S values reflect better discriminatory power. The K-S coefficients for logit regression and the hybrid technique are both acceptable, 74.1% and 87.68%, respectively (Panels A and B, Table 5.28). The Gini coefficients for the techniques are 85.28% (logit regression) and 87.41% (the hybrid technique). These values indicate that both techniques have very good discriminatory power as suggested by Abdou et al.'s (2016) classification system.

In short, both the logit regression and the hybrid technique have good classificatory power. However, the hybrid technique outperforms logit regression. This is because all evaluation criteria produced by the hybrid technique, AUC (93.84%), the KS coefficient (87.41%) and the Gini coefficient (87.68%), are higher than those generated by logit regression (AUC = 92.64%, KS = 85.28% and Gini = 74.10%).

5.4.2.3 Misclassification Cost

The misclassification cost considers costs generated by wrong classifications. Type I and Type II errors produce misclassification cost. Banks want to use the technique that offers the smallest expected losses. The misclassification costs generated by logit regression and the hybrid technique are shown in Table 5.29. This table shows that the misclassification cost increases when the cost ratio (good: bad loan) decreases (from 1:1 to 1:10). A cost ratio of 1:1 indicates that the cost of losing one good borrower equals 100% of the loss from one bad borrower. When the ratio is 1:10, this indicates that the cost of losing one good borrower equals 10% of the loss of incurred by a bad borrower (the lending interest rate has a value of 10%). Therefore, when the cost ratio decreases from 1:1 to 1:10, the bank will lose a compensation fee because the lending interest rate has decreased.

Table 5.29 The Misclassification Costs for Logit Regression and the Hybrid Technique

Misclassification	**Cost Ratio (Good Loan: Bad Loan)**									
Cost	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
Logit Regression	0.020	0.036	0.052	0.069	0.085	0.101	0.118	0.134	0.150	0.167
Hybrid Technique	0.008	0.015	0.022	0.029	0.037	0.044	0.051	0.058	0.065	0.073
Difference	-0.01	-0.02	-0.03	-0.04	-0.05	-0.06	-0.07	-0.08	-0.08	-0.09
%	-60%	- 58%	-57%	-57%	-57%	-57%	-57%	-57%	-56%	-56%

Source: Author's computation

Table 5.29 shows the different misclassification costs produced by logit regression and the hybrid technique. The hybrid technique's misclassification cost is 56% to 60% compared with logit regression's misclassification cost (see Table 5.29).

5.4.2.4 The Goodness of Fit Test

The Hosmer-Lemeshow test the goodness of fit test of both models (logit and hybrid). The null hypothesis is that the model fits. The p-value (of the Hosmer-Lemeshow test) (Table 5.29) for the logit regression is statistically significant. In contrast, the Hosmer-Lemeshow test p-value for the hybrid model is insignificant. Hence, the null of our model fit is rejected for logit regression but is accepted for the hybrid model. In other words, the hybrid model is a better fit than logit regression.

Table 5.30 The Hosmer and Lemeshow Goodness of Fit (GOF) Test Results

Model	**df**	χ^2	**Significance Level**
Logit Regression	8	54.1	0.0000 (***)
Hybrid Technique	8	7.51	0.4826 (NS)

Note: NS = non-significant; *** = denotes 1% significance level.
Source: Author's computation

5.4.3 A Summary of the Hybrid Technique

SOM's contribution to the hybrid technique is unclear. The reason is that the SOM technique estimated approximately 74% of the totally bad cases recorded by the bank. If this result (74% of actual bad cases) is used as outputs in the MLP technique, which was suggested in Waidyarathne and Samarasinghe's (2014) study, bad borrowers' information is lost. In short, the hybrid classification's results are incorrect. However, the component planes generated by the SOM technique have a strong capacity to illustrate the data. The component planes can assist researchers to understand results in a visual way and realise the motivation behind the classification.

In contrast, the merger of logit regression and MLP outperforms logit regression alone. In the hybrid model, the overall classification accuracy increases by 1.2%. Type II errors decreased significantly by 19.4%. This leads to a significant reduction in misclassification costs of 56% (the cost ratio is 1:10). Other criteria that evaluate the technique's discriminatory power, AUC, Gini and KS coefficient value, provide evidence that the hybrid technique has greater classificatory power than conventional logit regression. The Hosmer and Lemeshow test's results also indicate that the hybrid technique fits better than logit regression. In short, the bank executives should use the hybrid technique to build a more efficient CSM. The hybrid technique provides a visual picture of the data and has greater classificatory power.

Banks can also compensate for losses associated with bad debts by offering a different interest rate based on a customer's risk level. Hence, we need to investigate the critical factors that influence the interest rate charged and examine if the interest rate offered is based on risk factors associated with a borrower's probability of defaulting on the loan repayments. This will be discussed in the sections that follow.

5.5 Objective 6: To Investigate Factors That Determine The Interest Rate Charged on Retail Loans.

5.5.1 Research Model

We use the interest rate as a dependent variable to investigate the critical factors that influence the interest rate charged. The updated interest dataset includes 80 independent variables and one dependent variable (the interest rate charged). OLS regression is used to investigate the critical factors in the interest rate empirical Equation [5.6] is expressed mathematically as follows:

$$Interest\ rate = \beta_0 + \beta_1 Adddate_3 + \beta_2 loanValue_9 + \beta_3 openDate_{14} + \beta_4 duration_{16} + \alpha_5 industry_{4.1} + \cdots + \alpha_8 industry_{4.4} + \alpha_9 economic_{6.1} + \alpha_{10} economic_{6.2} + \alpha_{11} customer_8 + \alpha_{12} MIAtype_{12.1} + \cdots + \alpha_{22} MIAtype_{12.12} + \alpha_{23} purposeGroup_{19.1} + \cdots \alpha_{28} purposeGroup_{19.6} + \alpha_{29} purpose_{20.1} + \cdots + \alpha_{53} purpose_{20.24} + + \alpha_{54} HTCV_{21.1} + \cdots \alpha_{64} HTCV_{21.11} + \alpha_{65} TDCN_{22.1} + \cdots + \alpha_{79} TDCN_{22.14} + \alpha_{79} TOFL_{24.1} + \alpha_{80} TOFL_{24.2} + \varepsilon \quad [5.6]$$

Where:

Interest rate = the interest rate charged by the bank;

$\beta_0\ to\ \beta_4$ = coefficients to be estimated for the continuous variables;

$\alpha_5\ to\ \alpha_{80}$ = coefficients to be estimated for the dummy variables;

$Adddate_3, loanValue_9, interest_{13}, openDate_{14}, duration_{16}$ = independent variables (described in Table 5.1);

$industry_{4.1\ldots4}, economic_{6.1,6.2}, customer_8, MIAtype_{12.1\ldots12}, purposeGroup_{19.1..6}, purpose_{20.1\ldots24}, HTCV_{21.1\ldots11}, TDCN_{22.1..14}, TOFL_{24.1,2}$ = dummy variables generated from the independent variables (see Table 5.1); and

ε = error term.

5.5.2 Empirical Results and Discussion

Stepwise regression is used to select the critical factors that influence the lending interest rate. Beginning with empty model, a variable was added each step and checked whether it was significant (through the likelihood ratio test and the Wald test) below the 5% level. The results are shown in Table 5.31. The R-squared value is 61.79%. This result implies that all explanatory variables explain 61.79% of the total variation in the interest rate charged. The F-statistic is statistically significant at 1%, confirming the joint significance of the coefficients of the explanatory variables. The results reveal that the model fits the data well. Fifty-eight of the 80 independent variables are statistically significant at between 1% and 5% (see Table 5.31).

Table 5.31 The Critical Factors that Influence Loan Price

No.	Variable	Estimate	P-value	
1	Constant	4.97	0.000	***
2	Registration date	-0.02	0.000	***
3	Loan value	-0.02	0.000	***
4	Opening date	-0.66	0.000	***
5	Loan duration	0.01	0.019	**
6	18 - Service for individual and society	0.67	0.000	***
7	06 - Construction	0.27	0.032	**
8	New customer	0.17	0.000	***
9	Overdue loan 1, medium term VND	-0.48	0.003	***
10	Overdue loan 1, interest first payment	2.07	0.000	***
11	Overdue loan 1, medium term AMORT	7.15	0.000	***
12	Medium term loan AMORT VND	5.61	0.000	***
13	Overdraft loan short-term VND	1.67	0.000	***
14	Medium term interest at first payment VND	0.37	0.000	***

15	Long term interest at first payment VND	0.24	0.000	***
16	Short-term loan, interest first payment	-0.18	0.005	**
17	Unsecured staff loan VND	-0.79	0.000	***
18	Long term loan for bank's staff VND	-0.95	0.000	***
19	08. Hotel and Restaurant	-0.94	0.003	***
20	02. Fishery	0.67	0.018	**
21	0420 - Consumer goods, office equipment	0.96	0.002	***
22	1861 - Instalment loans for other bank staff	1.24	0.000	***
23	0712 - Consumer goods business|0714 - Construction material business	0.27	0.007	***
24	0723 - Constructing infrastructure	0.86	0.000	***
25	0810 - Hotel operating	1.64	0.001	***
26	0719 - Other business	0.25	0.005	***
27	0711 - Food business	0.55	0.000	***
28	1833 - Loan for buying house and sale|1835 - Loan for preparing house and business	0.81	0.000	***
29	1810 - Instalment loan for consumer goods	0.52	0.000	***
30	1830 - Preparing house with repayment source by salary	0.29	0.000	***
31	0715 – Medical equipment |1815 - Buying house with repayment source different from salary|1899 - other purpose	0.36	0.000	***
32	0710 - Commerce|1831 - Preparing house with repayment source different from salary|1870 - Instalment loans for consumption	0.43	0.000	***
33	1834 - Loan for preparing house for sale	0.59	0.000	***
34	1822 - Buying instalment house with repayment source different from salary	0.28	0.003	***
35	The one trillion incentive programme	0.82	0.011	**
36	Incentives program for SME	0.73	0.001	***
37	Incentive buying car purpose programme	1.04	0.000	***
38	Mortgage loan for consumer lending	0.45	0.000	***
39	Incentives program in 06 months	0.56	0.000	***
40	Collaboration loan with Gia Hoa	0.68	0.000	***
41	Incentive buying house purpose programme	0.76	0.000	***
42	Others	0.76	0.000	***
43	Incentive home repaired programme	0.60	0.000	***
44	Collaboration loan with Hung Thinh	0.59	0.000	***
45	Loan with diamond interest	0.92	0.000	***
46	Loan for real estate trading	1.09	0.000	***
47	Instalment loan for business	0.79	0.000	***
48	Others	0.61	0.000	***
49	Overdraft loan	2.35	0.000	***
50	Consumer loans with mortgage	0.75	0.000	***
51	loan for business	0.41	0.000	***
52	Loan for buying house	0.79	0.000	***
53	Loan for repairing house	0.78	0.000	***
54	Loan for future house, mortgage by that house	0.19	0.003	***
55	Car loan	0.72	0.000	***
56	instalment loan for car	-0.27	0.009	***
57	A> local trading sponsor	0.10	0.016	**
58	H> Loan for investment fixed asset	-0.40	0.000	***
	R-squared:	0.6179		
	Adjusted R-squared:	0.616		
	F-statistic:	319.1		
	p-value:	0		
	Number of observations	15865		
Note	1/ Dependent variable is the interest rate. 2/ ***, ** represent 1% and 5% significance, respectively.			

Source: Author's calculation

The *negative* sign of the estimators indicates a decreased interest rate. In other words, the independent variables associated with negative coefficients receive a lower interest rate.

The registration date, opening date and loan value variables are negative and statistically significant at the 1% level. Those negative estimators indicate that the interest rate has reduced from 2011 to 2016. More specifically, the loan price is reduced by 66 basis points when adding a year to the register date and opening date variable (for example, a one-year increase from 2009 to 2010) (see Item 4 in Table 5.31). Moreover, an increase in loan value could reduce the loan price by two basis points (see Item 3 in Table 5.31). These results reveal that the bank wants to expand credit services into the retail market because the lending interest rate has significantly reduced in recent years.

Unsecured and long-term staff loans are negative and statistically significant at the 1% level. Long-term loans are offered at an interest rate lower than 95 basis points. Staff are offered unsecured loans at 79 basis points (see Items 17 and 18 in Table 5.31). The results suggest banks encourage their staff to borrow because they give them a lower lending interest rate than for other loans.

The hotel and restaurant sector and loans for buying a house are negative and statistically significant at the 1% level. Hotel and restaurant sector loans are offered at a lower rate: 94 basis points. Home loans also receive a lower interest rate (40 basis points). These findings indicate that the bank considers these loans 'safer' than others. The bank believes these borrowers have a good financial plan to repay their financial obligations.

The *positive* sign of estimators indicates an increased interest rate. In order words, positively statistically significant variables indicate a higher interest rate.

Loan duration variable is positive and statistically significant at the 5% level. An increase of one-year in loan duration (e.g., from 9 to 10 years) increases the basis point by one basis point of the interest rate charged (see Item 5 in Table 5.31). This outcome indicates that the bank promotes or favours short-term/shorter term loans. Increasing the total number of short-term loans can reduce the liquidity risk because borrowers must repay the bank faster.

New customers are offered a rate 17 basis points higher than old customers. This result confirms that the longer a customer stays with a bank, the more the bank identifies him/her as a good borrower. In short, old borrowers are charged lower interest rates as a result of their prior relationship with the bank; they can command lower interest rates (Limsombunchai, 2005; Wang, 2010).

Loans for sectors such as construction, fishery, consumer goods, office equipment, food businesses and medical health are positive and have a statistically significant relationship with the interest rate, at the 1% to 5% level. Loans for these sectors are offered at between 27 to 164 basis points higher interest rate. These findings suggest that the bank considers these business sectors risky and therefore charges them a higher interest rate on any loans.

In conclusion, the interest rate charged fluctuates and depends on the duration, purpose, old vs new borrowers, the business sector and some incentive or collaboration programmes. One question remains: Do banks consider risk factors when determining their interest rate policies? Thisis investigated in the next section.

5.6 Objective 7: To Examine Bank Ranked Customers According to Their Risk Level to Offer a Loan Price

Magri (2018) argues that bank lending interest rates should reflect a borrowers' default risk. Each potential credit customer should be charged an interest rate that reflects his or her specific credit risk level. Higher frequency default borrowers might be charged a higher interest rate to compensate for potential losses; borrowers with the highest credit risk may be refused loans (Getter, 2006). In the beginning, the CSM was used to categorise borrowers and determine whether to approve a loan (Scroggins et al., 2004). Recently, credit scoring has improved, which means that banks can determine the loan price (Thomas, 2010). This approach is called risk-based pricing (RBP) and is used by many banks (Magri, 2018).

5.6.1 Evaluating the RBP Strategy

To examine whether a bank uses the RBP strategy, risk factors associated with default borrowers (outlined in Table 5.25) are compared with the factors that influence interest rates (Table 5.31). The comparisons are shown in Table 5.32.

Table 5.32 Comparing the Risk and Interest Rate Factors for Evaluating the RBP Strategy

No	Variable	Lending Model [Equation 5.1] (Column 1)		Interest Rate Model [Equation 5.6] (Column 2)		Classified (Column 3)
		Coef.	Sign	Coef.	Sign	
1	Register date	-0.14	***	-0.02	***	RBP
2	Loan value	0.13	***	-0.02	***	Contradict
3	Interest rate	0.12	**			
4	Opening date	-0.49	***	-0.66	***	RBP
5	Loan duration	0.10	***	0.01	**	RBP
6	07 - Commercial	4.57	**	0.10		Missed
7	18 - Service for society	5.02	**	0.67	***	RBP
8	06 - Construction	3.15		0.27	**	N.A.
9	04 - Company limited	4.37	**	0.06		Missed
10	New customer	0.38	**	0.17	***	RBP
11	Overdue loan 1, medium term	4.29	***	-0.48	***	Contradict
12	Overdue loan 1, first payment	1.03		2.07	***	N.A.
13	Overdue loan 1, medium AMORT	3.47	**	7.15	***	RBP
14	Medium term loan AMORT VND	-2.56		5.61	***	N.A.
15	Overdraft loan short-term VND	-3.81	**	1.67	***	Contradict

16	Medium term, first payment VND	-4.27	***	0.37	***	Contradict
17	Long term, first payment VND	-4.68	***	0.24	***	Contradict
18	Medium-term mortgage loan	2.34	**	0.17		Missed
19	Short-term loan, first payment	-5.37	***	-0.18	**	RBP
20	Unsecured staff loan VND	-2.34		-0.79	***	N.A.
21	Long term loan for bank's staff VND	-4.47	***	-0.95	***	RBP
22	08. Hotel and Restaurant	-0.12		-0.94	***	N.A.
23	02. Fishery	-0.75		0.67	**	N.A.
24	1229 - Buy and sell land	10.43	***	0.57		Missed
25	0420 - Consumer goods, office equipment	3.27		0.96	***	N.A.
26	0410 - Loan in food processing	4.89	**	0.43		Missed
27	1861 - Instalment loans for other staff	2.83		1.24	***	N.A.
28	0712 - Consumer goods business | 0714 - Construction material business	3.04	***	0.27	***	RBP
29	0723 - Constructing infrastructure	3.17		0.86	***	N.A.
30	0810 - Hotel operating	-0.28		1.64	***	N.A.
31	0719 - Other business	3.49	***	0.25	***	RBP
32	0711 - Food business	2.33		0.55	***	N.A.
33	1833 - Loans for buying house and sale	2.13	**	0.81	***	RBP
34	1810 - Instalment loan consumer	1.30		0.52	***	N.A.
35	1830 - Preparing house with repayment source by salary	2.42	***	0.29	***	RBP
36	0715 – Medical equipment | 1815 - Buying house with repayment source different from salary|1899 - other purpose	3.23	***	0.36	***	RBP
37	0710 - Commerce|1831 - Preparing house with repayment source different from salary|1870 - Instalment loans for consuming	2.79	***	0.43	***	RBP
38	1834 - Loan for preparing house and sale	2.45	**	0.59	***	RBP
39	1832 - Loan buying house and business	3.06	**	-0.14		Missed
40	1822 - Buying instalment house with repayment source different from salary	2.46	**	0.28	***	RBP
41	1820 - Buying instalment house with repayment source by salary|1813 - loan for buying a car	2.83	***	0.06		Missed
42	1860 - Loan for bank staff|1814 - Buying house with repayment source by salary	2.42	***	0.02		Missed
43	The 1 trillion government programme	-0.06		0.82	**	N.A.
44	Incentives program for SME	-0.98		0.73	***	N.A.
45	Incentive buying car purpose programme	-0.73		1.04	***	N.A.
46	Mortgage loan for consumer lending	-0.76		0.45	***	N.A.
47	Incentives program in 06 months	-1.43		0.56	***	N.A.
48	Collaboration loan with Gia Hoa	0.51		0.68	***	N.A.
49	Incentive buying house purpose programme	-1.33		0.76	***	N.A.
50	Others	-1.53		0.76	***	N.A.
51	Incentive home repaired programme	-2.02	**	0.60	***	Contradict
52	Collaboration loan with Hung Thinh	-0.21		0.59	***	N.A.
53	Loan with diamond interest	-0.76		0.92	***	N.A.
54	Loan for real estate trading	5.28	***	1.09	***	RBP
55	Instalment loan for business	2.78	***	0.79	***	RBP
56	No credit	2.70	***	0.61	***	RBP
57	Overdraft loan	3.97	***	2.35	***	RBP
58	Consumer loans with mortgage	3.28	***	0.75	***	RBP
59	Loan for business	3.38	***	0.41	***	RBP
60	Loan for buying house	2.49	***	0.79	***	RBP
61	Loan for repairing house	2.97	***	0.78	***	RBP
62	Loan for future house, mortgage by that house	2.91	***	0.19	***	RBP
63	Car loan without mortgage	2.25	***	0.72	***	RBP

64	Instalment loan for car	2.25	**	-0.27	***	Contradict
65	A> local trading sponsor	1.06	***	0.10	**	RBP
66	H> Loan for investment fixed asset	-1.47		-0.40	***	Missed

Note 1/ ***, ** indicate significance at 1% and 5%, respectively.
2/ RBP = the RBP strategy; N.A = non-application of RBP strategy; Contradict = contrast with RBP strategy; Missed = missing a risk factor.

Source: Author's computation

Each variable is classified in terms of the RBP strategy as follows:

If a variable has a positive coefficient that is positive and statistically significant at 1% to 5% (Column 1 in Table 5.32), this implies that an increase in this variable will lead to an increased risk of a borrower defaulting on a loan. Therefore, the bank needs to increase its interest rate to compensate for potential future losses. In other words, if this variable is positive and statistically significant at 1% to 5% it influences the interest rate charged (see Column 2 in Table 5. 31). If a variable has a coefficient that is positive and significant (1%-5%) in both Columns 1 and 2 in Table 5.32, it indicates that the bank has successfully applied the RBP strategy. In contrast, if a variable has a coefficient that is negative and statistically significant (1% - 5%) in both Columns 1 and 2 in Table 5.29, it indicates that an increase in this variable will decrease the risk of a single borrower defaulting on a loan and borrowers will be offered a lower interest rate. It also indicates that the bank has successfully applied the RBP strategy to this variable. The abbreviation "RBP" is used in Column 3 in Table 5. 31 to indicate that the bank has successfully applied the strategy.

If a variable has an insignificant coefficient in Column 1 in Table 5. 31, but is statistically significant level between 1% and 5% in Column 2, it implies that the bank has offered an interest rate that is not determined by a risk factor. This is because the variable has been identified as not impacting a borrower's probability of defaulting on loan repayments. Column 3 in Table 5. 31 uses the abbreviation "N.A." (not applicable) to indicate such a situation.

In contrast, if a variable has coefficient that is statistically significant at any level from 1 to 5% in Column 1, but is insignificant in Column 2, it indicates that the bank has ignored a critical risk factor. As a result, the bank could experience financial losses because the borrower defaults on the loan. Column 3 in Table 5.32 uses the term "missed" to describe this situation.

Lastly, in some cases, a variable has a coefficient that is statistically significant between 1% and 5% in both Columns 1 and 2 in Table 5.32, but with opposite signs; i.e., a variable has a positive sign in Column 1 but a negative one in Column 2 and vice versa. This indicates the bank has wrongly used the RBP strategy. In other words, the bank has offered a lower interest rate to a high-risk borrower or given a high loan price to a safe borrower. Banks need to identify such mistakes and quickly revise their interest policy. The term "contradict" is used in Column 3 in Table 5.32 to indicate this scenario.

5.6.2 Analysing the RBP Empirical Results and Discussion

In some cases, the strategy is applied correctly in relation to specific risk factors; that is, individuals with low risk variables are charged lower interest rates and conversely. These variables are marked with the abbreviation “RBP”. For example, recently released loans are considered safer; these customers they are thus offered a 66 basis points lower interest rate (see Item 4 in Table 5.32). In addition, short-term loans and loans for bank staff are recognised as being safer loans. Hence, customers who fit these categories are given lower interest rates, of 18 and 95 basis points, respectively (Items 19 to 21 in Table 5.32).

In contrast, high-risk factors are charged a higher interest rate. For instance, loans with a longer maturity date are seen as riskier because they are associated with a higher default rate. Thus, an increase of one year to the maturity date increases the interest rate by 1 basis point (see Item 5 in Table 5.32). New customers are recognised as being higher risks than old borrowers; thus, it is not surprising that new customers are given a higher interest rate of 17 basis points (Item 10 in Table 5.32). Overdue loans and overdraft loans are also high-risk loans and are charged a higher loan price of 715 and 235 basis points, respectively. Loans for real estate purposes and the construction sector have higher default rates. They are also charged a higher interest rate, between 27 to 109 basis points. Consumer loans are considered high risk and charged a higher interest rate (75 basis points). Loans (local sponsor), car loans without mortgages and loans for services are charged higher interest rates, 10, 72 and 67 basis points, respectively, because those loans have a higher default rate.

The bank has ignored several risk factors. These are identified by the term “missed”. These factors include high-risk borrowers, such as the commercial business sector, a borrower working for a limited company, medium-term mortgage loans, buying and trading land, and buying a house using an instalment plan with one’s salary. These factors could potentially lead to future losses because of the possibility of default. However, the interest rate charged by the banks doesn’t reflect the increased risk that these factors indicate.

Some situations demonstrate that the bank has decided an interest rate that is not based on risk factors. These situations are indicated by the term “N.A.”. For example, loans for unsecured staff and loans for the hotel and restaurant sector are given a lower interest rate (-79 and -94 basis points, respectively). These loans are not seen as being risky because they have a statistically insignificant probability of default.

In a few cases, the risk-based pricing rule has been wrongly applied. These situations are marked with the word “contradict”. Larger loans are associated with a higher probability of default, but customers in this category are given a lower interest rate (they are charged a value of -2 basis points) (see Item 2 in Table 5.32). In short, the bank has given customers a lower interest rate for a high-risk loan and, as a result, exposed itself to the possibility of future losses through default. Additionally, medium-term overdue loans and instalment loans for a car are verified as high-risk loans, but they are offered lower interest rates (-48 and -27 basis points, respectively). The bank needs to recognise these cases and revise its interest policies quickly to avoid future losses.

5.6.3 Summary of the RBP Empirical Results

Table 5.33 summarises all the variables that correctly follow the RBP strategy ("RBP" is used to indicate such cases; those where the strategy has not been applied are marked as "missed"). Situations where the strategies have been wrongly used, are marked by the word "contradict." The abbreviation "N.A" indicates where the rules have been ignored or not applied. Forty-two variables are verified as risk factors and 23 are not. The bank applies the RBP strategy correctly in relation to 27 of the 42 risk variables (64%). The bank has ignored nine risk factors and incorrectly uses RBP in the case of six variables. Those 15 (9+6) factors have the potential to cost the bank significant amounts of money because of the potential for customers to default on their loan.

Table 5.33 A Summary of the RBP, Missed, N.A. and Contradictory Cases of Bank Loans

	RBP	Missed	Contradict	Total Risk Factors	N.A.	Total
Number of variables	27	9	6	42	23	65
Percentage	64%	22%	14%	100%		

Note 1/ RBP = the RBP strategy; N.A = non-application of RBP strategy; Contradict = contrast with RBP strategy; Missed = missing a risk factor.

Source: Author's computation

5.7 Chapter Summary

This section summarises the empirical results for the Vietnamese retail banking industry.

First, the behaviour of borrowers has changed. When a loan is issued after one year, this loan decreases by 0.92% the probability of default. After the Vietnam economy survived the 2008 GFC, borrowers' income increased in stability. In other words, borrowers did not want to lose their creditworthiness and were willing to repay their financial obligations. Moreover, interest rates had also recently reduced. The lending interest rate dropped to 66 basis points when a loan is issued after one year. These results reveal that the bank wants to expand its market into the retail sector as indicated by its reduced loan prices.

This chapter has identified some primary risk factors to answer the study's fourth objective. Loan value is a risk factor. An increase in loan value increases the default rate by 0.25%. Loan duration also has a positive influence on the default rate. An increase of one year in the duration could increase the probability of default by 0.18%. This outcome confirms that a longer maturity date might be a signal of unaffordable short-term repayment capacity and higher future uncertainty, thus increasing the probability of default (Van Gool et al., 2012). The interest rate variable is also a risk factor. An increase of 1% in the interest rate will increase the default rate by 0.23%. Higher interest rates produce greater profitability for the bank. However, the bank may be confronted with a decline in credit demand and an increased level of bad debts. New customers are seen as risky borrowers. A new customer increases the probability of default by 0.72%. This finding is not surprising because an old customer, who has a long-term relationship with the bank, is considered a good borrower (Dinh & Kleimeier, 2007; Kocenda & Vojtek, 2011).

This chapter has found that component planes generated by the SOM technique have a strong capacity to illustrate the data. These component planes can assist researchers to understand results visually and realise the motivation behind the classification process by seeing and evaluating the correlation coefficients between the planes. The integrated model (a combination of logit regression and MLP) outperforms logit regression alone. Overall classification accuracy is improved to 1.2% in the hybrid technique. Type II errors decrease significantly from 34.7% to 15.3% (a reduction of over 50 per cent) with the hybrid technique. This leads to misclassification costs generated by the hybrid technique being reduced to 56% (at a cost ratio is 1:10). Other evaluating discriminatory power criteria such as AUC, Gini and KS coefficient value provide robust evidence that the hybrid technique has better classificatory power than logit regression. The goodness of fit test (the Hosmer and Lemeshow test) results also indicate that the hybrid technique fits better than logit regression. In short, the hybrid technique is a better option for bank executives wanting to build a more efficient CSM. This is because this hybrid technique has a great capacity to visualise data exploration and produce more accurate results.

Lastly, this chapter examined whether the bank uses the RBP strategy to determine borrowing interest rates. The result shows that 27 of 42 cases (64%) follow the RBP strategy correctly. However, the bank failed to notice some critical risk factors and issued an interest rate policy that could cause future financial losses through bad debts.

The next chapter summarises all the findings of domestic Vietnam banks and the retail banking sector as well as the results' implications.

Chapter 6

Conclusions and Policy Implications

6.1 Introduction

This chapter summarises the study's findings, implications, limitations and recommendation for future research. The study's findings are summarised in section 6.2. Section 6.3 presents the policy implications. Section 6.4 discusses the study's limitations and provides recommendations for future study.

6.2 A Summary of the Study's Findings

6.2.1 Vietnam Banking Industry

The first objective of the study was to determine whether a group of high market power banks exist in Vietnam's banking industry. The study first identified changes in the financial conditions of Vietnamese banks using the SOM trajectory technique. This technique is useful because it identifies financial trajectory patterns of domestic Vietnam banks that bank executives can use to track their financial condition over time. Bank executives can understand their financial circumstances, potential risks and take corrective action (Chen, 2012; Du Jardin & Séverin, 2011; 2012). In this study, the SOM trajectory technique was also used to categorise domestic Vietnam banks into two super-class groups for further analysis in terms of benchmarking market power and efficiency between these two groups

The SOM technique was used to categorise domestic Vietnam banks into two super-class groups, group 1 and group 2 banks (see Table 4.2). These two groups differ in several respects. Group 1 banks have smaller asset values than group 2 banks (see Appendix Table B4). Group 2 banks consist of the four commercial state banks and some big joint-stock banks; group 1 has only joint-stock banks. The 2D self-organisation map provides a clear picture of the financial trajectory patterns of domestic banks from 2008 to 2017 (see Figure 4.3). Several financial trajectory patterns were identified. First, 19 banks maintained their position in group 1 or group 2 over the study period. In contrast, eight banks moved group, either from the small group banks (group 1) to the large group banks (group 2) or vice versa.

This study used a non-structural indicator, the Lerner index, to measure domestic Vietnam banks' market power. It used the non-parametric DEA technique with the VRS assumption to capture bank efficiency (represented by PTE and CE scores) to answer research objective two. Though the difference in market power (represented by the Lerner scores) between the two groups of banks (1 and 2) is statistically significant at the 1% level, their efficiency (represented by the PTE and CE scores) was the same (see Tables 4.5 and 4.9).

Our results contradict Nguyen and Nghiem's (2018) results. They find no significant difference in market power between state-owned and joint-stock banks. These contradictory results might imply that the grouping approach, dividing domestic Vietnam banks into state-owned and joint-stock banks, may be inappropriate for identifying differences in market power within domestic Vietnam banks. This study argues that the SOM technique, based on an unsupervised algorithm, can better capture differences in market power and can categorise the domestic Vietnam banks into strong banks (group 2) and weak banks (group 1). This study argues that there are two groups of banks, with different levels of market power that exist side-by-side in the Vietnam banking industry. This situation was caused by M&As. The existence of different levels of market power in the two groups of banks negatively affects the competitive environment of the Vietnam banking industry. The results show group 2 banks are more profitable because of higher market power, whereas group 1 banks struggle to cut costs to maintain their competitiveness.

The third objective of this study was to explore the factors that impact bank market power and efficiency. The study used two bank market power and cost efficiency models to determine bank market power and cost efficiency. Bank-specific characteristics (*BANKSIZE*, the *CAP* ratio and the *NPL* ratio) and macroeconomic indicators (GDP growth rate and the inflation rate) were used as independent variables. The Lerner index and cost-efficiency scores were used as response variables in the bank market power and cost efficiency models (Equations 4.1 and 4.2).

The panel dataset was examined for stationarity using the Fisher-Type option. The p-values were statistically significant in both empirical models (see Table 4.13). Thus, the null of non-stationarity (unit root) was confidently rejected. This means that the dataset had stationarity and no unit root problem. The Pearson pairwise correlation results show that the correlation ratios among the exploratory variables are less than 0.8 (see Table 4.14). The VIF values did not exceed 10 (Table 4.15). Thus, there is no multicollinearity in the dataset. These outcomes confirm that it is appropriate to use pooled OLS regression, a conventional estimation technique (Gujarati & Porter, 2009). The fixed effect estimator was used to estimate the factors that impact bank market power and efficiency. Fixed effect regression was used because it can handle the heterogeneity associated with pooled OLS (Gujarati and Porter, 2009).

The p-value of the Breusch-Pagan/Cook-Weisberg test and the modified Wald Test for groupwise heteroscedasticity was statistically significant at all conventional significance levels (Table 4.18). The p-value of the Wooldridge Test for autocorrelation was statistically significant at the 1% level. These results confirm that heteroscedasticity and autocorrelation were present in the dataset. Consequently, the pooled OLS and fixed-effect estimators were no longer BLUE (Gujarati and Porter, 2009). Furthermore, theoretical and empirical evidence identified the existence of dynamic relationships in the empirical models for bank market power and cost efficiency. Therefore, one-year lagged market power and cost efficiency (LER_{it} and CE_{it} scores) were used as additional exploratory variables to capture these dynamic relationships. This study used Wooldridge's (2010) strict exogeneity test; the result indicated an endogeneity problem in the dataset. *BANKSIZE, CAP* ratio, and *NPL* ratio were identified as endogenous variables.

Pooled OLS and fixed effects estimation results may be biased, inconsistent and inefficient if heteroscedasticity, autocorrelation and endogeneity are not eliminated (Gujarati and Porter, 2009; Wintoki et al., 2012). A dynamic estimator, such as the dynamic panel GMM, is appropriate to handle these types of problem so it was used (Antoniou et al., 2008; Wintoki et al., 2012; Zeitun et al., 2017). In this study, robustness tests were used to examine the robust estimated coefficients. Table 6.1 summarises the empirical findings.

Table 6.1 A Summary of the Empirical Findings on the Factors that Impact Models of a Bank's Market Power Using Various Methods

	Empirical Model of a Bank's Market Power (LER_{it})			**Empirical Model of a Bank's Cost Efficiency (CE_{it})**		
Estimator	Pooled OLS	Fixed effect	SysGMM	Pooled OLS	Fixed effect	SysGMM
LER_{it-1}			(+) ***			
CE_{it-1}						(+) **
$BANKSIZE_{it}$	(+) ***	(+)	(+) **	(+) ***	(-) ***	(-) **
CAP_{it}	(+) ***	(+) **	(+) ***	(+) **	(-)	(+)
NPL_{it}	(-) ***	(-) ***	(-) **	(-) *	(-)	(-) *
GDP_{it}	(+)	(+) **	(-)	(-)	(+)	(+) *
INF_{it}	(+)	(-) **	(-) **	(+)	(-) **	(+) *

Note: (+) and (-) denote a positive or negative relationship; *, ** and *** are significant at the 10%, 5% and 1% levels, respectively; sysGMM symbolises the 2-step system GMM estimator.
Source: Author's computation

My results suggest that domestic Vietnam banks have accumulated technological knowledge and skills and, as a result, their efficiency increases the following year. This can be seen by the significant positive link between the one-year lagged CE_{it-1} and cost efficiency at the 5% level. This result indicates that the Vietnam State bank has encouraged commercial banks to maintain efficiency over time (Adjei-Frimpong et al., 2014).

The Vietnam banking industry has maintained its stability over time. A new entry bank is selected and launched using a plan to maintain banking system stability and prevent a reduction in bank franchise value (Adjei-Frimpong et al., 2016). This can be seen by the significant positive link between the one-year lagged LER_{it-1} and market power at the 1% level. This outcome indicates that in Vietnam non-transparent information practices, network lending relationships and limited banking regulations encourage monopolistic banking practices. Therefore, the banks in the large super-class group can maintain their high level of market power and, ultimately, profitability (Berger et al., 2000; Delis, 2012). The SOM result provides evidence of a stable banking system because over 70% of local banks (19 of 27 banks) have maintained their position in group 1 or 2 banks from 2008 to 2017.

Large banks (those with greater assets) are stronger (have greater market power). These banks can set higher loan prices and therefore be more profitable. The reason is that *BANKSIZE* has a significant positive, robust connection with bank market power at the 1% level. This result also indicates that bigger banks have greater opportunities to access cheaper financial resources because of more depositor funds. These big banks are superior at handling moral hazard issues (borrowers) and override their competition to improve their market power (Delis, 2012). However, my results suggest that bank efficiency (in the bigger banks) suffers from diseconomies of scale. This is seen in the significant negative relationship between *BANKSIZE* and cost efficiency at the 5% level. This finding also indicates that bigger banks are worse at distributing their costs than smaller banks because of bureaucracy; this ultimately lowers their efficiency (Delis & Papanikolaou, 2009; Gardener et al., 2011).

Prior studies indicate that domestic Vietnam banks operate in response to the "quiet life hypothesis" (Nguyen & Nguyen, 2018; Nguyen et al., 2016b). This means that increased competition will improve bank cost-efficiency (Nguyen et al., 2016b). The statistically significant difference, in terms of market power, between the two bank super-class groups provides evidence that the two groups, with different levels of market power, exist side-by-side in the Vietnam banking industry. Group 2 banks enjoy greater market power and are less likely to be impacted by competition from other banks. As a result, bigger banks (group 2) are less efficient in response to the quiet life hypothesis. Bank size is an essential factor for banks to compete in financial markets. This study shows that *BANKSIZE* has a positive significant relationship with bank market power and has a negative significant relationship with cost efficiency during 2008 to 2017. This is because successful M&A activity may improve bank size quickly, but negatively impact on bank efficiency because of increased NPL value and operational costs (because of an increase in the number of bank staff).

Well-capitalised domestic Vietnam banks have greater market power. This is because well-capitalised banks have more available funds that allow them to seize opportunities considered riskier. This means they can lend to more borrowers, enjoy higher interest margins and gain higher market power (Abel & le Roux, 2017; Erler et al., 2017). This relationship was confirmed by the positive relationship between *CAP* and market power, which is statistically significant at the 1% level. In contrast, the source of funds (from equity owners or depositors) does not impact bank efficiency, as seen by the insignificant relationship between CAP_{it} and CE_{it}. Therefore, in the case of domestic Vietnam banks, increased capitalisation will increase market power but not cost efficiency.

Non-performing loans impact negatively on domestic Vietnam banks' market power and cost efficiency. This was confirmed by the negative significant relationships between NPL_{it} and LER_{it}/CE_{it}, which are significant at 5% and 10% levels, respectively. Higher NPL rates lead to a deterioration of a bank's loan portfolio (Abel & le Roux, 2017). In such a situation, banks can reduce their total number of loans and/or focus only on attracting good borrowers (Abel & le Roux, 2017). Banks are unable to set higher loan prices (or lower their market power) because they must compete to attract quality borrowers (Abel & le Roux, 2017). In addition, higher NPL rates increase the probability of bad debts. Banks must pay additional monitoring and enforcement costs, which ultimately decreases their efficiency costs. Thus, domestic Vietnam banks must work on decreasing their NPL to boost both their market power and their efficiency.

Domestic Vietnam banks can improve their efficiency during periods of rapid GDP growth. This is because economic growth contributes to higher incomes and lowers borrower default levels (Anwar, 2018; Erler et al., 2017). Hence, banks can reduce losses from bad debts and reserve lower provision for bad debts because of a decreased default rate (Anwar, 2018). Every Vietnam state bank is required by law to reserve 50% of the total loan amount for any overdue loan (up to 180 days). Over 180 days, the bank is required to reserve 100% of the total loan amount. Money reserved for these purposes increases costs and ultimately reduces the bank's profit. Therefore, a decreased default rate and fewer bad debts lead to decreased operating costs and, ultimately, to greater efficiency. This was confirmed by the significant positive relationship between GDP and CE at the 10% level. In contrast, economic growth does not affect the Vietnam banking environment; the relationship between GDP_t and LER_{it} is insignificant

High inflation decreases bank market power. This can be seen in the significant negative relationship between INF_t and LER_{it} at the 10% level. Banks cannot set higher prices for their financial products because they must compete to attract more borrowers. An inflationary environment increases bank efficiency. This is evident in the significant positive relationship between INF_t and CE_{it} at the 10% level. The reason is that domestic Vietnam banks increase their lending rates to compensate for losses associated with additional costs (salaries and/or employee benefits).

6.2.2 Vietnam Retail Banking

This section summarises the empirical results for the Vietnam retail banking sector. Retail banking is a profitable market for domestic Vietnam banks to embrace if they are looking for long-term expansion because of the increase in the number of internet users and emerging e-commerce. However, less than 50% of Vietnam banks are satisfied with their current internal credit rating system and 'poorly' designed CSMs (KPMG, 2013). This is a major factor that contributes to high NPL levels in Vietnam. NPL also negatively impacts bank market power and efficiency (see section 6.2.1). Therefore, I investigated the critical factors in lending and propose a more accurate combination method (a hybrid method) that can clearly interpret the relationship between input and output to build a more efficient CSM

The customer dataset was gathered from a bank in the group 1 banks. The bank is the Viet Nam Thuong Tin Commercial Bank (Vietbank). The dataset contained 15,865 credit profiles from 2006 to 2017. The weight of evidence and information value technique were used to transform data for modelling. Two of the 16 variables were removed from the original dataset because they were classified as having very weak predictive power. The remaining 14 variables were grouped into homogeneous clusters. No further variables were removed because absolute values of the correlating coefficients were less than 0.7. The transformed numerical dataset was screened and analysed using SOM. The 20x20 SOM grid was chosen to screen the dataset because this grid generated the most accurate prediction of 'bad' borrowers and depicted clearly ordinary clusters (see Appendix Figure D2 and D3).

Logit regression was used to investigate the critical factors associated with borrower creditworthiness to answer research objective four. This technique was selected because logit regression is broadly applied in financial practice and is easy to calculate and interpret (Chi & Hsu, 2012; Crook et al., 2007; Kruppa et al., 2013). Nine of 14 variables were classified as qualitative variables. To capture the effect of each attribute, dummy variables were generated (Wooldridge, 2012). The dummy dataset contained 81 variables, five of which were continuous and 76 were dummy variables (assigned a value of 0 and 1). The key empirical findings were (Table 6.2):

- First, borrowers' behaviours have changed. When a loan is issued after one year (over the study period from 2006 to 2017), the loan has a decreased probability of default by 0.92%. An individual who borrows money in 2006 is less likely (by 0.92%) to default on the loan than an individual who borrowed money in 2005. After the 2008 GFC, the borrower's income in Vietnam became more stable. In other words, borrowers did not want to lose their creditworthiness; the results show they were more willing to repay their financial obligations. Moreover, the findings indicate that interest rates have also recently reduced. The lending interest rate dropped 66 basis points when a loan was issued just one year later (during 2006 to 2017). This implies that an individual who borrowed money in 2006 will be charged a lower interest rate (6%) than an individual who borrowed money in 2005 (charged 6.66%). These results suggest that Vietbank wants to expand its market into the retail sector as indicated by reduced loan prices.
- Second, this study has identified some primary risk factors. Loan value is a risk factor; an increase in loan value increases the default rate by 0.25%. An increase of one year in loan duration increases the probability of default borrowers by 0.18%. This outcome confirms the commonly-held belief that a longer maturity date may indicate unaffordable short-term repayment capacity and higher future uncertainty, thus increasing the probability of default (Van Gool et al., 2012). The results

show that the lending interest rate is also a risk factor. An increase of 1% in the interest rate increases the default rate by 0.23%. Higher interest rates will, in theory, produce greater profitability. However, the bank may be confronted with a decline in credit demand and an increased level of bad debts. New customers are also risky borrowers. A new customer has an increased probability of default of 0.72%. This finding is not surprising because an old customer, who has a long-term relationship with the bank, is considered a good borrower (Dinh & Kleimeier, 2007; Kocenda & Vojtek, 2011).

Table 6.2 A Summary of the Empirical Findings of Critical Factors Impacting Creditworthiness of Borrowers and Interest Rate Charged

No	Variable		Lending Decision Empirical Equation		Interest Rate Empirical Equation		Note
			Coef.	Sign	Coef.	Sign	
1	Registration date		(-)	***	(-)	***	
2	Loan value		(+)	***	(-)	***	
3	Interest rate		(+)	**			
4	Opening date		(-)	***	(-)	***	
5	Loan duration		(+)	***	(+)	**	
6	Industry class	(industry_4.1…4.4)					Categories
7	New/old customer	(customer_8)					Categories
8	Credit term	(MIA_type_12.1…12.12)					Categories
9	Economic class	(economic_6)					Categories
10	Purpose group	(purpose_group_19.1…19.6)					Categories
11	Loan purpose	(Pupore_20.1…20.24)					Categories
12	Lending programs	(HTCV_21.1…21.11)					Categories
13	Individual credit	(TDCN_22. 1..22.12)					Categories
14	Sponsor	(TOFL_24.1…24.2)					Categories

Note: (+) and (-) denote a positive or negative relationship; *, ** and *** reveal significant levels at 10%, 5% and 1%;
Source: Author's computation

The major shortcoming of the logit regression technique is that the overall classification accuracy is lower than advanced computing and various hybrid approaches (Lessmann et al., 2015; Trinkle & Baldwin, 2016; Yu et al., 2015). Therefore, this study used a hybrid technique that integrates logit regression and ANNs to improve the classificatory power to answer research objective five. The hybrid technique used in this study combined three techniques: SOM, logit regression and MLP. SOM was used to screen the dataset. Logit regression was used to identify significant variables associated with borrowers' creditworthiness. Finally, all significant variables were put into MLP. The empirical findings for the hybrid technique are:

- Logit regression identified 43 statistically significant variables from 81 variables. Their significance levels differed from 1% – 5%. Hence, 43 variables were used as input data for MLP. In addition, the AUC, KS and Gini values generated by MLP were optimised when the number of neurons in the hidden layer was 20 neurons. Hence, an MLP network was built using 20 neurons in the hidden layer (see Figure 5.10).
- The component planes generated by SOM have a strong capacity to illustrate the data (see Figure 5.7). These component planes enable researchers to see the results visually and realise the motivation behind the classification. However, SOM was found to be less accurate in identifying significant risk factors than logit regression.
- The results reveal that the integrated model (a combination of logit regression and MLP) outperformed logit regression alone. The overall classification accuracy was improved by 1.2% by the hybrid technique. Type II errors decreased significantly, from 34.7% to 15.3% (a reduction of over 50 per cent) with the hybrid technique. This leads to the misclassification costs generated by the hybrid technique being reduced to 56% (at a cost ratio of 1:10). Other evaluating discriminatory power criteria (AUC, Gini and the KS coefficient value) provided robust evidence that the hybrid technique has greater classificatory power than logit regression. The goodness of fit test (Hosmer and Lemeshow test) results indicate that the hybrid technique fits better than logit regression.

Thus, the hybrid technique is a better option for bank executives to build a more efficient CSM. This is because this hybrid technique has a strong capacity for visualising data and produces more accurate results than logit regression alone.

Lastly, this study examined whether Vietbank uses the RBP strategy to determine borrowing interest rates. Our results showed that 27 of 42 cases (64%) followed the RBP strategy correctly. However, the bank failed to notice some critical risk factors and issued an interest rate that may cause future financial loss through bad debts.

6.3 Policy Implications of the Study

This study's findings have important implication for bank executives, policymakers and academics. This is the only study which uses the SOM technique to track dynamic changes in the financial situation of domestic Vietnam banks. Our results have shown that banks in Vietnam can be divided into two super-class bank groups (1 & 2). This classification provides academics with a new approach, which is based on an unsupervised algorithm. This is different from previous studies which divide domestic Vietnam banks into commercial state and joint stock banks (Nguyen & Nghiem, 2018; Vu & Turnell, 2010). As illustrated, the SOM has captured the differences in bank market power. Using this method, we have been able to divide Vietnam domestic banks into strong banks (group 2) and weak banks (group 1).

Moreover, the SOM 2D self-organisation map provides a clear picture of the financial trajectory patterns of domestic banks for the study period 2008 to 2017 (see Figure 4.3). This map provides bank executives with a visual image, which may enable them to better understand their financial situation in the market. Thus, they are able to select appropriate strategies or variables (*BANKSIZE*, *CAP* ratio, *NPL* ratio) to change their positions; from group 1 (the small group banks) to group 2 (bigger group banks) banks, or vice versa.

The existence of a bank group of strong market power makes it harder for smaller banks to survive. Group 2 banks (bigger banks) will survive better in a competitive environment and will seek continued expansion (Tabak et al., 2012; Wang, 2015). In such an environment, group 2 banks end up acquiring group 1 banks. This explains why the number of banks reduced from 43 to 32 over the study period 2008 to 2017. In short, group 1 banks will find it extremely difficult to survive. This is because banks in this group will struggle to cut costs to remain competitive (because of lower market power) and ultimately become a target for larger banks who are interested in mergers and acquisitions. Policymakers and regulators must take this phenomenon (group 2 banks acquiring group 1 banks) into consideration when creating any policies in order to maintain an optimal number of banks to ensure the stability of the banking system. As Le (2014) notes, regulators believe that 15 to 17 is the ideal number of banks to achieve stability in the Vietnamese banking system. However, there are currently 32 banks in operation. Thus, more than half of the total number of domestic Vietnam banks need to be merged.

The results show that domestic Vietnam banks' market power is persistent overtime. This is seen by the positive relationship between the one-year lagged Lerner and current bank market power. This finding indicates that Vietnam banking industry is characterised by non-transparent information, networking lending relationships and limited regulations around monopolistic practices, which leads to bank monopolies and reduces competition and innovation (Berger et al., 2000; Delis, 2012). Monopolies decrease competition and therefore consumers lose out. Banks which face lower levels of competition are less likely to be innovative, particularly in terms of adopting new technology. Policymakers and regulators must encourage greater transparency in relation to the information provided by the banking industry (for example, published bank information on loan interest rates), focus on decreasing monopolistic banking practices in order to ensure greater levels of competition and increase benefits for the customers. One possible strategy for reducing high market power is to encourage local banks to list shares on the Vietnamese stock exchange. In September 2018, 17 of 32 local owned banks were listed on the stock exchange (Ngo, Mac, & Tran, 2019). Therefore, policymakers and regulators must encourage banks to list on the stock exchange; this will increase information transparency in the banking industry (as the banks must make their financial data available to the public) and would also reduce monopolistic practices.

Well-capitalised and bigger banks have higher market power and profitability. This can be seen by the positive relationship between *BANKSIZE/CAP* ratio and bank market power. Size and capital ratio are important factors that bank executives in small banks should focus on to improve these ratios, in order to compete in the market. Besides, non-performing loans negatively affect bank market power and cost efficiency. As a result, bank executives and policymakers must work on diminishing NPL problems in order to maintain an efficient bank system. The Vietnamese government has already taken initial steps toward sorting this problem, by establishing the Vietnam Asset Management Company (VAMC), a firm which buys bad debts off banks to reduce their bad debt ratios in their loan portfolios. Until 2017, the VAMC collected and solved only 14% of the total bad debts (approximately 37,983 billion VND). Banks has recovered themselves the remaining 86% of total bad debt (approximately 224,071 billion VND) (Tran et al., 2017). Some difficulties associated with bad debts include: (i) Purchase and sale of bad debts were not implemented. The big contracts of the purchase and sale of bad debts were not generated because professional debt trading market was not developed; (ii) Credit institutions confront difficulty in identifying information on the position of collaterals (Ngo et al., 2019). Therefore, further steps could still be taken. For examples, the VAMC must have capital resources (supported from State budget) and have the rights to determine bad debts (Tran et al., 2017).

This study benefits the development of Vietnam retail banks by identifying several factors (loan value, loan duration, interest rate, and new customers) associated with default borrowers. More specifically, an increase in loan value will increase the default rate by 0.25%. An increase of one year in the loan duration will increase the probability of default borrowers by 0.18%. An increase of 1% in the interest rate will increase the default rate by 0.23%. A new customer will increase the probability of default by 0.72%. Some specific loans (loans for real estate trading and car loans) are considered unsafe loans. Loans for real estate trading will increase the probability of default by 9.93%. Car loans also increase the probability of default by 4.24%. Retail bank executives must consider these risk factors in their lending interest policies. Higher loan prices must be applied to such loans to compensate for potential losses.

In addition, by applying an automatic evaluation technique with standard procedures and criteria to assess a borrower's creditworthiness, banks can reduce their operation costs, increase repayment rates and boost their competitive advantage. Our results show that the hybrid technique can help retail bank executives build a more efficient CSM. Component planes generated using the SOM technique, can assist researchers in understanding results in a visual way and realise the motivation behind the classification process through seeing and evaluating the correlation coefficients between these planes. The integrated model (a combination of logit regression and MLP) outperforms logit regression alone. The hybrid technique can reduce misclassification costs from 100% to 56% compared with the logit regression alone.

6.4 Limitations of the Study and Future Research

One limitation of this study is the small sample size. The small sample size (because of the small number of Vietnam banks) means that the study cannot investigate additional determinants such as profitability (the return on assets/equity ratio), liquidity (the ratio of deposits to assets), the interest rate, market share, bank concentration (the HHI index) and control variables (the monetary policy, the growth of M2, the economic policy uncertainty, interest rate margin, etc). All of these impact both bank market power and efficiency. This is because increasing the determinants increases the number of instruments generated by the system GMM estimation. The problem of too many instruments (over-fitting endogenous variables) generated by system GMM estimations may weaken the Hansen test results and invalidate the system GMM results (Roodman, 2009a & b). In other words, the GMM results are valid only if the problem of endogeneity is partially solved.

A future study could include other countries in a cross-cultural analysis or examine individual countries such as Laos, Cambodia and Myanmar, to examine whether their bank industries suffer from similar issues. Including additional countries in the dataset would also enable scholars to investigate further the determinants that impact a bank's market power and efficiency.

Finally, the ANN was built using three layers: input, hidden, and output. Though the number of neurons in the input and output layer are known, the optimal structure of the hidden layer is still ambiguous and requires further research (Samarasinghe, 2016). This study used only one layer in the hidden layer to determine the optimal classification. The results show that the hybrid technique with one hidden layer outperformed logit alone. However, the classification ability of the hybrid technique could be improved if an optimal structure of the hidden layer were available. Therefore, future study may add more layer patterns (two or more) to optimise the hidden layer of neural networks.

References

Ab-Rahim, R. (2017). Efficiency and competition in the Malaysian banking market: Foreign versus domestic banks. Gadjah Mada International Journal of Business, 19(2), 193-221.

Abdou, H. A. (2009). Genetic programming for credit scoring: The case of Egyptian public sector banks. Expert Systems with Applications, 36 (9), 11402-11417.

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. Intelligent Systems in Accounting, Finance & Management, 18 (2-3), 59-88.

Abdou, H. A., Pointon, J., & El-Masry, A. (2008). Neural nets versus conventional techniques in credit scoring in Egyptian banking. Expert Systems with Applications, 35 (3), 1275-1292.

Abdou, H. A., Tsafack, M. D. D., Ntim, C. G., & Baker, R. D. (2016). Predicting creditworthiness in retail banking with limited scoring data. Knowledge-Based Systems, 103 89-103.

Abel, S. (2018). Cost efficiency and non-performing loans: An application of the Granger causality test. Journal of Economic and Financial Sciences (), 11(1), a170 doi:https://doi.org/10.4102/jef.v11i1.170

Abel, S., & le Roux, P. (2017). Evaluating Market Power in the Zimbabwean Banking Sector. Journal of Economic and Financial Sciences, 10(2), 274-291. doi: https://doi.org/10.4102/jef.v10i2.17

Abrahams, C., & Zhang, M. (2008). Fair Lending Compliance: Intelligence and Implications for credit risk management.

Adjei-Frimpong, K., Gan, C., & Hu, B. (2014). Cost efficiency of Ghana's banking industry: a panel data analysis. The International Journal of Business and Finance Research, 8(2), 69-86.

Adjei-Frimpong, K., Gan, C., & Hu, B. (2016). Competition in the banking industry: Empirical evidence from Ghana. Journal of Banking Regulation, 17(3), 159-175.

Al-Gasaymeh, A. (2016). Bank efficiency determinant: Evidence from the gulf cooperation council countries. Research in International Business and Finance, 38, 214-223.

Altman, E. I. (1968). Financial Ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(Sep 1968), 589-609.

Altunbas, Y., Carbo, S., Gardener, E. P., & Molyneux, P. (2007). Examining the relationships between capital, risk and efficiency in European banking. European Financial Management, 13(1), 49-70. doi:10.1111/j.1468-036X.2006.00285.x

Anderson, R. (2007). The Credit Scoring Toolkit. United States: Oxford University Press.

Angelini, P., & Cetorelli, N. (2003). The effects of regulatory reform on competition in the banking industry. Journal of Money, Credit, and Banking, 35(5), 663-684.

Antoniou, A., Guney, Y., & Paudyal, K. (2008). The determinants of capital structure: Capital market-oriented versus bank-oriented institutions. The Journal of Financial and Quantitative Analysis, 43(1), 59-92.

Anwar, M. (2018). Cost efficiency performance of Indonesian banks over the recovery period: A stochastic frontier analysis. The Social Science Journal. doi:https://doi.org/10.1016/j.soscij.2018.08.002

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. Review of Economic Studies, 58(2), 277-297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. Journal of Econometrics, 68(1), 29-51.

Ariff, M., & Can, L. (2008). Cost and profit efficiency of Chinese banks: A non-parametric analysis. China Economic Review, 19 (2), 260-273.

Ataullah, A., & Le, H. (2006). Economic reforms and bank efficiency in developing countries: the case of the Indian banking industry. Applied Financial Economics, 16(9), 653-663. doi:10.1080/09603100500407440

Baklouti, I., & Bouri, A. (2014). Can Credit-Scoring Models Effectively Predict Microloans Default? Statistical Evidence from the Tunisian Microfinance Bank. Global Credit Review, 3(1), 57-69.

Baltagi, B. H. (2008). Econometric Analysis of Panel Data (4th ed.). Chichester, UK: John Wiley & Sons Inc.

Barbour-Lacey, E. (2014). Vietnam's Retail Banks Seek to Entice New Customers. Retrieved from http://www.vietnam-briefing.com/news/vietnams-retail-banks-seek-entice-new-customers.html/

Barth, J. R., Lin, C., Ma, Y., Seade, J., & Song, F. M. (2013). Do bank regulation, supervision and monitoring enhance or impede bank efficiency. Journal of Banking & Finance, 37 (8), 2879-2892.

Bartlett, J. (2014). The Hosmer-Lemeshow goodness of fit test for logistic regression. Retrieved from https://thestatsgeek.com/2014/02/16/the-hosmer-lemeshow-goodness-of-fit-test-for-logistic-regression/

Baum, C. F. (2013). Dynamic Panel Data estimators. Retrieved from http://fmwww.bc.edu/EC-C/S2013/823/EC823.S2013.nn05.slides.pdf

Beck, R., Jakubik, P., & Piloiu, A. (2013). Non-performing loans - What matters in addition to the economic cycle? Paper presented at the Macroprudential Research Network, Working Paper Series No 1515

Bekhet, H. A., & Eletter, S. F. K. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. Review of Development Finance, 4(1), 20-28.

Berger, A. N., Bonime, S. D., Covitz, D. M., & Hancock, D. (2000). Why are bank profits so persistent? The roles of product market competition, informational opacity, and regional/macroeconomic shocks. Journal of Banking and Finance, 24(7), 1203-1235.

Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. Expert Systems with Applications, 40(1), 356-364.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87 (1), 115-143.

Bond, S., Hoeffler, A., & Temple, J. (2001). GMM estimation of empirical growth models. Economics Papers 2001-W21, Economics Group, Nuffield College, University of Oxford.

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. Portuguese Economic Journal, 1(2), 141-162.

Boone, J. (2008). A new way to measure competition. The Economic Journal, 118(531), 1245-1261. doi:10.1111/j.1468-0297.2008.02168.x

BIDV. (2017). Annual report 2017. Retrieved from https://www.bidv.com.vn

Chan, S.-G., & Karim, M. Z. A. (2010). Bank Efficiency and Macro-economic Factors: The Case of Developing Countries. Global Economic Review, 39(3), 269-289. doi:10.1080/1226508X.2010.513141

Chen, M.-Y. (2012). Visualisation and dynamic evaluation model of corporate financial structure with self-organising map and support vector regression. Applied Soft Computing, 12 (8), 2274-2288.

Chen, N., Ribeiro, B., Vieira, A., & Chen, A. (2013). Clustering and visualisation of bankruptcy trajectory using self-organising map. Expert Systems with Applications, 40 (1), 385-393.

Chi, B.-W., & Hsu, C.-C. (2012). A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model. Expert Systems with Applications, 39 (3), 2650-2661.

Choi, I. (2001). Unit root tests for panel data. Journal of International Money and Finance, 20(2), 249-272.

Cholongitas, E., Senzolo, M., Patch, D., Shaw, S., Hui, C., & Burroughs, A. K. (2006). Review article: scoring systems for assessing prognosis in critically ill adult cirrhotics. Alimentary Pharmacology and Therapeutics, 24(2006), 453-464.

Chuang, C., & Lin, R. (2009). Constructing a reassigning credit scoring model. Expert Systems with Applications, 36 (2), 1685-1694.

Coccorese, P. (2009). Market power in local banking monopolies. Journal of Banking & Finance, 33 (7), 1196-1210.

Coccorese, P. (2014). Estimating the Lerner index for the banking industry: a stochastic frontier approach. Applied Financial Economics, 24(2), 73-88. doi:10.1080/09603107.2013.866202

Cox, D. R. (1970). The Analysis of Binary Data. London: Chapman and Hall.

Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. European Journal of Operational Research, 183 (3), 1447-1465.

Cubiles-De-La-Vega, M.-D., Blanco-Oliver, A., Pino-Mejías, R., & Lara-Rubio, J. (2013). Improving the management of microfinance institutions by using credit scoring models based on Statistical Learning techniques. Expert Systems with Applications, 40 (17), 6910-6917.

Delis, M. D. (2012). Bank competition, financial reform, and institutions: The importance of being developed. Journal of Development Economics, 97(2), 450-465.

Delis , M. D., & Pagoulatos, G. (2009). Bank competition, institutional strength and financial reforms in Central and Eastern Europe and the EU. MPRA Paper No. 16494

Delis, M. D., & Papanikolaou, N. I. (2009). Determinants of bank efficiency: evidence from a semi-parametric methodology. Managerial Finance, 35 (3), 260-275. doi:10.1108/03074350910931771

Demirguc-Kunt, A., & Martínez-Pería, M. S. (2010). A Framework for Analysing Competition in the Banking Sector: An Application to the Case of Jordan. https://ssrn.com/abstract=1726248

Dinh, T. H. T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. International Review of Financial Analysis, 16(5), 471-495.

Drucker, H., Burges, C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. Advances in Neural Information Processing Systems, 9, 155-161.

Du Jardin, P., & Séverin, E. (2011). Predicting corporate bankruptcy using a self-organising map: An empirical study to improve the forecasting horizon of a financial failure model. Decision Support Systems, 51 (3), 701-711.

Du Jardin, P., & Séverin, E. (2012). Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. European Journal of Operational Research, 221 (2), 378-396.

Edelberg, W. (2003). Risk-based Pricing of Interest Rates in Household Loan Markets. FEDS Working Paper No. 2003-62. http://dx.doi.org/10.2139/ssrn.484522

Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. Journal of Monetary Economics, 53 (8), 2283–2298.

Equifax. (2014). Understanding Credit Scores. Retrieved from Equifax: https://www.equifax.com/ePort/Understanding_Credit_Scores_[FLASH_READY].ppt

Erler, A., Gischer, H., & Herz, B. (2017). Regional Competition in US Banking – Trends and Determinants. Paper presented at the FEMM Working Papers 170008, Otto-von-Guericke University Magdeburg, Faculty of Economics and Management.

Euromonitor. (2016). Consumer Lending in Vietnam. Retrieved from http://www.euromonitor.com/consumer-lending-in-vietnam/report

Fernández De Guevara, J., & Maudos, J. (2007). Explanatory factors of market power in the banking system. The Manchester School, 75 (3), 1463-6786.

Fernández de Guevara, J., Maudos, J., & Pérez, F. (2005). Market Power in European Banking Sectors. Journal of Financial Services Research, 27(2), 109-137.

FICO. (2015). FICO score - The score lenders use.

Fiordelisi, F., Marques-Ibanez, D., & Molyneux, P. (2011). Efficiency and risk in European banking. Journal of Banking & Finance, 35 (5), 1315-1326.

FirstData. (2016). FE Credit broadens product offerings to consumers in Vietnam. Retrieved from First Data Corporation: https://www.firstdata.com/downloads/pdf/FE-Credit-FirstVision-Client-Case-Study.pdf

Fred. (2019). Lerner Index in Banking Market for Viet Nam [DDOI04VNA066NWDB]. Retrieved from https://fred.stlouisfed.org/series/DDOI04VNA066NWDB

Gardener, E., Molyneux, P., & Nguyen-Linh, H. (2011). Determinants of efficiency in South East Asian banking. The Service Industries Journal, 31(16), 2693-2719. doi:10.1080/02642069.2010.512659

Getter, D. E. (2006). Consumer Credit Risk and Pricing. Journal of Consumer Affairs, 40(1), 41-63. doi:10.1111/j.1745-6606.2006.00045.x

Greene, W. H. (2012). Econometric Analysis (7th ed.). Boston; London: Pearson.

Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (5th ed.). Boston: McGraw-Hill.

Ha, D. (2016). Foreign banks take on different approach. Retrieved from http://www.vir.com.vn/foreign-banks-take-on-different-approach.html

Havrylchyk, O. (2006). Efficiency of the Polish banking industry: Foreign versus domestic banks. Journal of Banking & Finance, 30 (7), 1975-1996.

HDbank. (2016). Annual report 2016. Retrieved from https://hdbank.com.vn/api/download?fileName=/investor/editor/rjX8EL351iFyIKjIGFLO20190416180331%2F%40%40%40AR2016E_1555412740817.pdf

Hill, R. C., Griffiths, W. E., & Lim, G. C. (2011). Principles of Econometrics (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Ho, A., & Baxter, R. A. (2011). Banking Reform in Vietnam Retrieved from https://www.frbsf.org/banking/files/june-banking-reform-in-vietnam-June-2011.pdf

Hoang, T. T. H., Phan, D. V., & Bandaralage, J. (2016). Mergers, Acquisitions and Market Concentration in the Banking Sector: The Case of Vietnam. Asian Journal of Economics and Empirical Research, 3(1), 49-58.

Huang, E., & Scott, C. T. (2007). Credit Risk Scorecard Design , Validation and User Acceptance - A Lesson for Modellers and Risk Managers.

Huysmans, J., Baesens, B., Vanthienen, J., & Gestel, T. V. (2006). Failure prediction with self organising maps. Expert Systems with Applications, 30 (3), 479-487.

Jabareen, Y. (2009). Building a Conceptual Framework: Philosophy, Definitions, and Procedure. International Journal of Qualitative Methods, 8(4), 49-62. doi:10.1177/160940690900800406

Jagric, T., & Jagric, V. (2011). A Comparison of Growing Cell Structures Neural Networks and Linear Scoring Models in the Retail Credit Environment. Eastern European Economics, 49(6). doi:10.2753/EEE0012-8775490605

Jagric, V., Kracun, D., & Jagric, T. (2011). Does Non-linearity Matter in Retail Credit Risk Modeling? Czech Journal of Economics and Finance,, 4(2011).

Kablan, A. (2010). Banking Efficiency and Financial Development in Sub-Saharan Africa. No 10/136, IMF Working Papers, International Monetary Fund

Kader, H., Adams, M., & Hardwick, P. (2010). The Cost Efficiency of Takaful Insurance Companies. The Geneva Papers on Risk and Insurance - Issues and Practice, 35 (1), 161-181.

Kiruthika, & Dilsha, M. (2015). A Neural Network Approach for Microfinance Credit Scoring. Journal of Statistics & Management Systems, 18(1 & 2), 121-138. doi:10.1080/09720510.2014.961767

Kocenda, E., & Vojtek, M. (2011). Default Predictors in Retail Credit Scoring: Evidence from Czech Banking Data. William Davidson Institute Working Paper Number 1015.

Kočišová, K. (2014). Application of Data Envelopment Analysis to Measure Cost, Revenue and Profit Efficiency. 94, 47-57.

Kohonen, T. (1982). Self-organised formation of topologically correct feature maps. Biological Cybernetics, 43(1), 59-69.

KPMG. (2013). Vietnam banking survey. Retrieved from KPMG Limited

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. Expert Systems with Applications, 40 (13), 5125-5131.

Le, T. (2014). Market Concentration and Competition in Vietnamese Banking Sector. MPRA Paper No. 57406. http://mpra.ub.uni-muenchen.de/57406/

Le, T. T. V. (2009). Vietnam's policy responses to the Financial Crisis. Paper presented at the EAI Background Brief.

Lee, T.-S., & Chen, I.-F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications, 28(4), 743-752.

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247 (1), 124-136.

Limsombunchai, V., Gan, C., & Lee, M. (2005). An Analysis of Credit Scoring for Agricultural Loans in Thailand. American Journal of Applied Sciences, 2(8), 1198-1205.

Lin, F.-T., Wu, H.-Y., & Tran, T. N. N. (2015). Internet banking adoption in a developing country: an empirical study in Vietnam. Information Systems and e-Business Management, 13(2), 267-287.

Liu, H., Molyneux, P., & Nguyen, L. H. (2012). Competition and risk in South East Asian commercial banking,. Applied Economics, 44(28), 3627-3644. doi:10.1080/00036846.2011.579066

Lu, H., Liyan, H., & Hongwei, Z. (2013). Credit scoring model hybridizing artificial intelligence with logistic regression. Journal of Networks, 8(1), 253-261.

Lu, X., & White, H. (2014). Robustness checks and robustness tests in applied economics. Journal of Econometrics, 178 (1), 194-206.

Magri, S. (2018). Are Lenders Using Risk-Based Pricing in Consumer Loan Market? The effect of the 2008 crisis. Temi di discussione (Economic working papers) 1164, Bank of Italy, Economic Research and International Relations Area.

Magri, S., & Pico, R. (2011). The rise of risk-based pricing of mortgage interest rates in Italy. Journal of Banking & Finance, 35 (5), 1277-1290.

Marque, A., García, V., & Sánchez, J. (2013). A literature review on the application of evolutionary computing to credit scoring. Journal of the Operational Research Society, 64(9), 1384-1399.

Matroushi, S. (2011). Hybrid Computational Intelligence Systems based on Statistical and Neural Networks Methods for Time Series Forecasting: The Case of Gold Price. Master of Applied Science thesis, Lincoln University Library.

Nadeem, M., Gan, C., & Nguyen, C. (2017). Does intellectual capital efficiency improve firm performance in BRICS economies? A dynamic panel estimation. Measuring Business Excellence, 21 (1), 65-85. doi:10.1108/MBE-12-2015-0055

Ngo, P. A., Mac, D. T., & Tran, M. H. (2019). Banking Industry Report 2019. Retrieved from https://www.vcbs.com.vn/en/Communication/GetReport?reportId=6109

Nguyen, H. P. (2018). Profitability of Vietnamese Banks Under Competitive Pressure. Emerging Markets Finance and Trade, 55(9), 2004-2021. doi:10.1080/1540496X.2018.1511977

Nguyen, H. S., Tran, T. T. T., & Tran, T. H. Y. (2014). Bank Restructuring–International Perspectives and Vietnam Practices. Accounting and Finance Research, 3(2), 36-50.

Nguyen, H.-T. (2015). How is credit scoring used to predict default in China? EconomiX Working Papers 2015-1, University of Paris Nanterre, EconomiX.

Nguyen, M., Skully, M., & Perera, S. (2012). Bank market power and revenue diversification: Evidence from selected ASEAN countries. Journal of Asian Economics, 23 (6), 688-700.

Nguyen, T. P. T., & Nghiem, S. H. (2018). The effects of competition on efficiency: The Vietnamese banking industry experience. The Singapore Economic Review, 63(1), 1-30. doi:10.1142/S0217590817500114

Nguyen, T. P. T., Nghiem, S. H., Roca, E., & Sharma, P. (2016a). Bank reforms and efficiency in Vietnamese banks: evidence based on SFA and DEA. Applied Economics, 48(30), 2822-2835. doi:10.1080/00036846.2015.1130788

Nguyen, T. P. T., Nghiem, S. H., Roca, E., & Sharma, P. (2016b). Efficiency, innovation and competition: evidence from Vietnam, China and India. Empirical Economics, 51(3), 1235-1259. doi:10.1007/s00181-015-1045-5

Nguyen, T. P. T., Roca, E., & Sharma, P. (2014). How efficient is the banking system of Asia’s next economic dragon? Evidence from rolling DEA windows. Applied Economics, 46(22), 2665-2684. doi:10.1080/00036846.2014.909578

Nguyen, V. A. (2005), "Vietnam civil law code", in the National Assembly, (Ed. 33/2005/QH11).

Oliver, R. M., & Wells, E. (2001). Efficient Frontier Cutoff Policies in Credit Portfolios. The Journal of the Operational Research Society, 52(9), 1025-1033.

Othman, F. M., Mohd-Zamil, N. A., Rasid, S. Z. A., Vakilbashi, A., & Mokhber, M. (2016). Data Envelopment Analysis: A Tool of Measuring Efficiency in Banking Sector International Journal of Economics and Financial Issues, 6(3), 911-916.

Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. Expert Systems with Applications, 36(1), 2-17.

Pham, L., Cao, N. Y., Nguyen, T. D., & Tran, P. T. (2013). Structural Models for E-Banking Adoption in Vietnam. International Journal of Enterprise Information Systems, 91(1), 31 - 48.

Pham, T. T. T., & Lensink, R. (2008). Household Borrowing in Vietnam: A Comparative Study of Default Risks of Formal, Informal and Semi-formal Credit. Journal of Emerging Market Finance, 7(3), 237-261. doi:https://doi.org/10.1177/097265270800700302

Phan, H. T. M., Daly, K., & Akhter, S. (2016). Bank efficiency in emerging Asian countries. Research in International Business and Finance, 38 (2016), 517-530.

Potočnik, P., Berlec, T., Starbek, M., & Govekar, E. (2012). Self-organising neural network-based clustering and organisation of production cells. Neural Computing and Applications, 22(Suppl 1), 113-124.

Presbitero, A. (2005). The Debt-Growth Nexus: a Dynamic Panel Data Estimation. Working Papers 243, Universita' Politecnica delle Marche (I), Dipartimento di Scienze Economiche e Sociali.

Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. The Stata Journal, 9(1), 86-136.

Roodman, D. (2009b). A Note on the Theme of Too Many Instruments. Oxford Bulletin of Economics and Statistics, 71, 135-158. doi:10.1111/j.1468-0084.2008.00542.x

Samarasinghe, S. (2006). Neural Networks for Applied Sciences and Engineering. Auerbach Publications, Boston, MA.

Samarasinghe, S. (2016). Order in the Black Box: Consistency and Robustness of Hidden Neuron Activation of Feed Forward Neural Networks and Its Use in Efficient Optimization of Network Structure. Springer International Publishing Switzerland, Artificial Neural Network Modelling, Studies in Computational Intelligence, 628. doi:10.1007/978-3-319-28495-8_2

Saigonbank. (2017). Annual report 2017. Retrieved from https://www.saigonbank.com.vn/

SBV. (2018). The State Bank of Vietnam (SBV). Retrieved from https://www.sbv.gov.vn/webcenter/portal/en/home/sbv?_afrLoop=13989992297220095

SCB. (2016). Annual report 2016. Retrieved from

Schreiner, M. (2004). Scoring Arrears at a Microlender in Bolivia. Journal of Microfinance, 6(2), 65-88.

Scroggins, W. A., Fielding, W. T., & Clark, L. J. (2004). Borrower segmentation and credit scoring in bank consumer lending. Retrieved from http://www.westga.edu/~bquest/2006/segmentation.pdf

Sealey, C. W., & Lindley, J. T. (1977). Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions. Journal of Finance, 32(4), 1251-1266.

Sherman, H. D., & Zhu, J. (2006). Service Productivity Management: Improving Service Performance Using Data Envelopment Analysis (DEA). New York: Springer.

Siddiqi, N. (2006). Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. Hoboken, N.J: Wiley.

Simpasa, A. M. (2013). Competition and Market Structure in the Zambian Banking Sector. Working Paper Series N° 168 African Development Bank, Tunis, Tunisia

Soedarmono, W., Machrouh, F., & Tarazi, A. (2011). Bank market power, economic growth and financial stability: Evidence from Asian banks. Journal of Asian Economics, 22 (6), 460-470.

Staub, R. B., Souza, G. d. S. e., & Tabak, B. M. (2010). Evolution of bank efficiency in Brazil: A DEA approach. European Journal of Operational Research, 202 (1), 204-213.

Stewart, C., Matousek, R., & Nguyen, T. N. (2016). Efficiency in the Vietnamese banking system: A DEA double bootstrap approach. Research in International Business and Finance, 36 (2016), 96-111.

StoxPlus. (2016). Foreign Banks Increase Presence in Vietnam. Retrieved from https://biinform.com/News-Detail/223103/foreign-banks-increase-presence-in-vietnam-8355.html

Šušterši, M., Mramor, D., & Zupan, J. (2009). Consumer credit scoring models with limited data. Expert Systems with Applications, 36 (3), 4736-4744.

Tabak, B. M., Fazio, D. M., & Cajueiro, D. O. (2012). The relationship between banking market competition and risk-taking: Do size and capitalisation matter? Journal of Banking & Finance, 36 (12), 3366-3381.

Thomas, L. (2010). Consumer finance: challenges for operational research. Journal of the Operational Research Society, 61(1), 41--52.

Thorndike, R. L. (1953). Who Belongs in the Family? Psychometrika, 18 (4), 267-276. doi:10.1007/BF02289263

Tran, B. T., Ong, B., & Weldon, S. (2015). Vietnam Banking Industry Report. Retrieved from https://www.duxtonam.com/wp-content/uploads/2015/01/Vietnam-Banking-Report_2015.pdf

Tran, T. H., Tran, M. H., & Ly, H. A. T. (2017). Banking Report 2017. Retrieved from https://www.vcbs.com.vn/en/Communication/GetReport?reportId=4830

Tran, V. T. D., & Ho, Q. (2015). A proposed credit scoring model for Loan default Arobability: A Vietnamese bank case. Paper presented at the In International Conference on Qualitative and Quantitative Economics Research (QQE). Proceedings (p. 52). Global Science and Technology Forum. 10.5176/2251-2012_QQE15.32

Trinkle, B. S., & Baldwin, A. A. (2007). Interpretable Credit Model Development via Artificial Neural Networks. Intelligent Systems in Accounting, Finance & Management, 15(3-4), 123-147.

Trinkle, B. S., & Baldwin, A. A. (2016). Research opportunities for neural networks: the case for Credit. Intelligent Systems in Accounting, Finance And Management, 23(3), 240-254.

Tsai, C.-F. (2014). Combining cluster analysis with classifier ensembles to predict financial distress. Information Fusion, 16, 46-58.

Tsai, C.-F., & Chen, M.-L. (2010). Credit rating by hybrid machine learning techniques. Applied Soft Computing, 10 (2), 374-380.

Upadhyay, R. (2013). Information Value (IV) and Weight of Evidence (WOE) – A Case Study from Banking (Part 4). Retrieved from http://ucanalytics.com/blogs/information-value-and-weight-of-evidencebanking-case/

Van Gool, J., Wouter, V., Piet, S., & Bart, B. (2012). Credit scoring for Microfinance: is it worth? . International Journal of Finance & Economics, 17(2), 103-123.

Vasanthi, P., & Raja, P. (2011). Risk Management Model: an Empirical Assessment of the Risk of Default. Journal of Risk and Diversification - Issue 1.

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. New York: Springer.

VPbank. (2016). Annual report 2016. Retrieved from https://report.vpbank.com/en/2016.html

Vietbank. (2017). Annual report 2017. Retrieved from https://www.vietbank.com.vn/

Vu, H. T., & Turnell, S. (2010). Cost Efficiency of the Banking Sector in Vietnam: A Bayesian Stochastic Frontier Approach with Regularity Constraints. Asian Economic Journal, 24 (2), 115-139.

Waidyarathne, K. P., & Samarasinghe, S. (2014). Artificial neural networks to identify naturally existing disease severity status. Neural Computing and Applications, 25(5), 1031-1041. doi:10.1007/s00521-014-1572-6

Wang, T. (2015). Competition and Increasing Returns to Scale: A Model of Bank Size. The Economic Journal, 125(585), 989-1014.

Wang, W. (2010). The Probability of Chinese Mortgage Loan Default and Credit Scoring. (Master of Commerce and Management), Lincoln University.

West, D. (2000). Neural network credit scoring models. Computers & Operations Research, 27 (11-12), 1131-1152.

West, D., Dellana, S., & Qian, J. (2005). Neural network ensemble strategies for financial decision applications. Computers & Operations Research, 32 (10), 2543-2559.

Wintoki, M. B., Linck, J. S., & Netter, J. M. (2012). Endogeneity and the dynamics of internal corporate governance. Journal of Financial Economics, 105 (3), 581-606.

Wooldridge, J. M. (2010), Econometric Analysis of Cross Section and Panel Data, The MIT Press, Cambridge, Massachusetts.

Wooldridge, J. M. (2012). Introductory Econometrics : A Modern Approach. United States of America: Cengage Learning.

Worldbank. (2018a). Inflation, consumer prices (annual %). Retrieved from http://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=VN

Worldbank. (2018b). Lending interest rate (%). Retrieved from http://data.worldbank.org/indicator/FR.INR.LEND?end=2015&locations=VN&start=2000&view=chart

WorldTrade, S. (2008). Vietnam Trade Summary 2008 Data. Retrieved from http://wits.worldbank.org/CountryProfile/en/Country/VNM/Year/2008/Summary

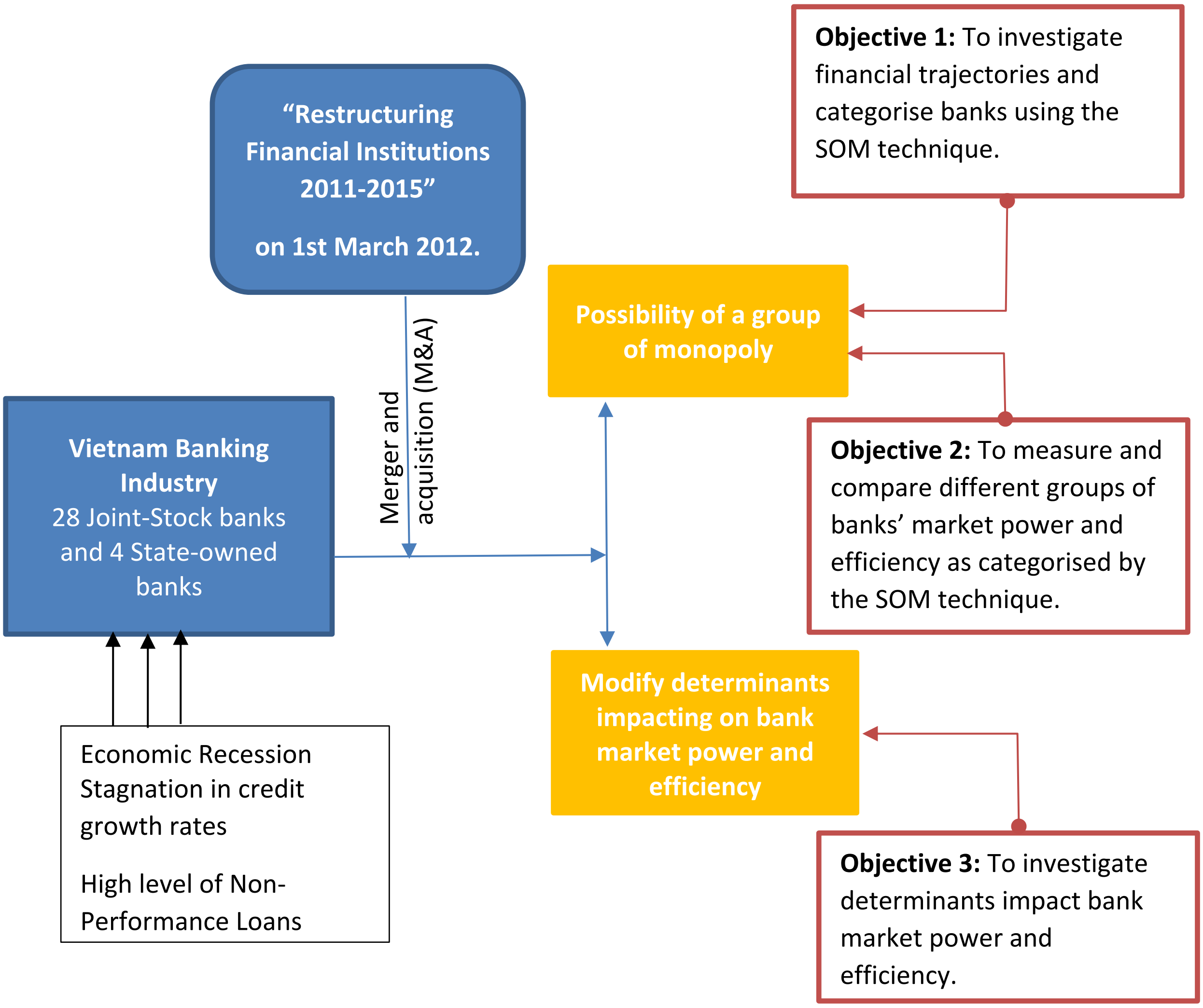
Yu, L., Li, X., Tang, L., Zhang, Z., & Kou, G. (2015). Social credit: a comprehensive literature review. Financial Innovation, 1(6). doi:https://doi.org/10.1186/s40854-015-0005-6

Zeitun, R., Temimi, A., & Mimouni, K. (2017). Do financial crises alter the dynamics of corporate capital structure? Evidence from GCC countries. The Quarterly Review of Economics and Finance, 63 21-33.

Appendix A

Domestic Vietnam Banks

Appendix Table A.0 Conceptual framework



Appendix Table A.1 List of Domestic Vietnam Banks from 2008 to 2017

No	Name	Trade name	Type
1	Bank for Investment and Development of Vietnam	BIDV, BID	State-owned
2	Vietnam Bank for Agriculture and Rural Development	Agribank	State-owned
3	Vietnam Bank for Industry and Trade	Vietinbank, CTG	State-owned
4	Joint Stock Commercial Bank for Foreign Trade of Vietnam	VCB, Vietcombank	State-owned
5	Sai Gon Joint Stock Commercial Bank	SCB	JSC
6	Saigon Thuong Tin Commercial Joint Stock Bank	Sacombank, STB	JSC
7	Military Commercial Joint Stock Bank	MBBank	JSC
8	Vietnam Technological and Commercial Joint- stock Bank	Techcombank	JSC
9	Saigon Hanoi Commercial Joint Stock Bank	SHB	JSC
10	Asia Commercial Bank	ACB	JSC
11	Vietnam Prosperity Joint-Stock Commercial Bank	VPBank	JSC
12	Ho Chi Minh Development Joint Stock Commercial Bank	HDbank	JSC
13	LienViet Post Joint Stock Commercial Bank	LienVietPostBank	JSC
14	Vietnam Commercial Joint Stock Export Import Bank	Eximbank	JSC
15	Vietnam Public Joint Stock Commercial Bank	PVcom bank	JSC
16	Tien Phong Commercial Joint Stock Bank	TPBank	JSC
17	Vietnam International Commercial Joint Stock Bank	VIB	JSC
18	Southeast Asia Commercial Joint Stock Bank	SeABank	JSC
19	Vietnam Maritime Commercial Stock Bank	Maritime Bank	JSC
20	Bac A Commercial Joint Stock Bank	BacABank	JSC
21	An Binh Commercial Joint Stock Bank	ABBank	JSC
22	National Citizen Commercial Joint Stock Bank	NCB Bank	JSC
23	Orient Commercial Joint Stock Bank	OCB Bank	JSC
24	Viet A Joint Stock Commercial Bank	VietABank	JSC
25	Nam A Commercial Joint Stock Bank	NamABank	JSC
26	BaoViet Commercial Joint Stock Bank	Bao Viet Bank	JSC
27	Vietnam Thuong Tin Commercial Joint Stock Bank	Viet Bank	JSC
28	Kien Long Commercial Joint Stock Bank	KienLongBank	JSC
29	Viet Capital Joint Stock Commercial Bank	Vietcapital Bank	JSC
30	The Co-operative Bank of Vietnam	Co-opbank	JSC
31	Petrolimex Group Commercial Joint Stock Bank	PG Bank	JSC
32	Saigon Bank for Industry and Trade	Saigonbank	JSC

Note: Several Vietnam banks which ceased operating in the Vietnam banking industry within the study period (2008 – 2017) are excluded from this study. Tin Nghia Bank and First Commercial Bank merged with the Saigon Commercial Bank (SCB) in 2011. In 2012, the Habubank merged with the Saigon Hanoi Bank (SHB). The Western Bank was acquired by the Petro Vietnam Financial Company. In 2014, the Contruction Bank merged with the Vietinbank. In 2015, Mekong Housing Bank merged with BIDV. Petrol Gas Bank, Global Petrol Bank and Ocean Bank were acquired by the Vietinbank in 2015.

Source: State Bank of Vietnam (2018); Hoang et al. (2016).

Appendix Table A.2 Summary Statistic of the Sample Banks' Financial Data (2008 to 2017)

Variables	Year	No. of banks	Obs.	Mean	Standard Deviation	Minimum	Maximum
Total Asset (Billion VND)	2008	25	258	64,557	97,829	2,419	396,993
	2009	25		86,254	115,208	7,478	470,000
	2010	25		117,862	136,439	12,628	524,000
	2011	25		136,651	156,460	15,942	560,000
	2012	26		145,174	169,344	14,842	617,859
	2013	26		163,160	189,509	14,684	697,037
	2014	26		188,750	217,589	15,823	763,590
	2015	26		221,872	263,696	17,749	874,807
	2016	27		253,590	306,428	19,047	1,006,404
	2017	27		302,639	365,945	21,319	1,202,283
Owner's Equity (Billion VND)	2008	25	258	4,890	4,685	1,021	14,040
	2009	25		5,806	5,187	1,093	17,639
	2010	25		8,238	7,063	2,022	24,749
	2011	25		9,656	8,486	2,591	28,639
	2012	26		11,560	11,010	3,002	41,547
	2013	26		13,088	13,475	3,003	54,075
	2014	26		13,498	14,026	3,211	55,013
	2015	26		15,103	14,975	3,217	56,110
	2016	27		16,297	16,200	3,066	60,399
	2017	27		17,976	17,675	3,216	63,765
Customer Deposits (Billion VND)	2008	25	258	47,102	83,874	1,172	375,033
	2009	25		58,259	95,327	4,230	434,331
	2010	25		72,666	107,605	5,782	474,941
	2011	25		94,194	129,251	6,242	506,316
	2012	26		110,615	144,840	9,269	557,028
	2013	26		127,034	154,903	10,823	568,591
	2014	26		151,113	182,928	12,064	656,271
	2015	26		181,721	222,793	14,086	763,361
	2016	27		207,308	263,413	15,202	886,004
	2017	27		229,930	302,206	14,849	1,011,314
Charter Capital (Billion VND)	2008	25	258	3,920	3,690	1,000	12,526
	2009	25		4,525	3,754	1,000	12,100
	2010	25		6,387	5,210	1,820	21,512
	2011	25		7,381	5,880	2,000	21,687
	2012	26		8,956	7,424	3,000	26,218
	2013	26		9,907	9,025	3,000	37,324
	2014	26		10,261	9,399	3,000	37,234
	2015	26		11,288	10,033	3,000	37,234
	2016	27		11,587	10,622	3,000	37,234
	2017	27		12,083	10,648	3,000	37,234
Customer Loans (Billion VND)	2008	25	258	37,448	66,961	275	294,697
	2009	25		51,955	83,877	3,172	365,952
	2010	25		69,091	108,129	5,156	474,941

	2011	25		76,288	112,718	3,604	443,877
	2012	26		89,066	129,080	5,990	480,453
	2013	26		102,333	146,584	10,670	548,774
	2014	26		118,957	163,666	11,232	580,493
	2015	26		146,769	200,959	11,612	676,688
	2016	27		168,155	223,277	12,533	744,814
	2017	27		197,041	261,047	13,989	863,575
Operating	2008	25	258	2,475	3,959	122	16,253
Income	2009	25		2,700	3,425	220	12,792
(Billion VND)	2010	25		3,937	5,079	400	19,679
	2011	25		5,664	7,683	603	31,533
	2012	26		5,543	7,437	513	30,433
	2013	26		5,347	7,011	449	25,609
	2014	26		5,839	7,294	480	25,736
	2015	26		6,909	8,456	700	30,441
	2016	27		8,007	10,287	717	35,476
	2017	27		10,032	12,451	755	42,680
Operating Expenses	2008	25	258	1,148	2,168	33	9,802
	2009	25		1,305	2,186	119	10,178
	2010	25		1,835	2,951	171	13,074
(including Staff	2011	25		2,503	3,654	245	16,464
Costs, Depreciation	2012	26		2,775	3,613	315	16,539
and Other	2013	26		2,760	3,429	343	15,152
Expenses)	2014	26		2,921	3,518	319	14,897
	2015	26		3,426	3,924	377	15,610
(Billion VND)	2016	27		3,992	4,914	419	18,217
	2017	27		4,575	5,148	402	19,100
Income before	2008	25	258	1,327	1,942	26	6,451
Provision for Credit	2009	25		1,396	1,605	91	5,793
Losses	2010	25		2,102	2,319	229	7,622
	2011	25		2,992	4,199	(1,278)	15,069
(Billion VND)	2012	26		2,770	3,967	92	13,894
	2013	26		2,592	3,777	48	11,874
	2014	26		2,918	3,939	59	13,283
	2015	26		3,483	4,625	111	14,831
	2016	27		4,015	5,508	144	17,259
	2017	27		5,456	7,423	212	23,581
Provision for Credit	2008	25	258	438	878	3	2,900
Losses	2009	25		413	940	17	4,470
(Billion VND)	2010	25		499	861	18	3,125
	2011	25		1,169	2,569	0	11,556
	2012	26		1,254	2,253	(564)	9,790
	2013	26		1,210	2,069	(492)	8,001
	2014	26		1,472	2,201	(50)	8,310
	2015	26		1,899	2,662	17	11,647
	2016	27		2,113	3,156	64	13,048
	2017	27		2,801	4,646	(51)	18,515
	2008	25	258	889	1,145	13	3,966

Profit before Tax (EBIT) (Billion VND)	2009	25		983	1,386	(1,856)	5,004
	2010	25		1,603	1,605	185	5,479
	2011	25		1,891	2,223	(1,371)	8,392
	2012	26		1,521	2,038	3	8,168
	2013	26		1,379	2,021	24	7,751
	2014	26		1,518	2,081	10	7,302
	2015	26		1,503	2,283	8	7,473
	2016	27		1,897	2,696	14	8,569
	2017	27		2,656	3,336	31	11,341
Profit after Tax (EAT) (Billion VND)	2008	25	258	716	946	10	3,319
	2009	25		737	1,120	(1,909)	3,945
	2010	25		1,228	1,234	139	4,236
	2011	25		1,396	1,672	(1,371)	6,259
	2012	26		1,145	1,518	2	6,169
	2013	26		1,051	1,520	18	5,808
	2014	26		1,191	1,628	8	5,727
	2015	26		1,194	1,833	6	6,377
	2016	27		1,515	2,164	11	6,858
	2017	27		2,092	2,688	22	9,091
Return on Asset (ROA) %	2008	25	258	1.35%	1.10%	0.16%	5.95%
	2009	25		1.35%	0.62%	-0.39%	3.11%
	2010	25		1.36%	0.80%	0.49%	4.73%
	2011	25		1.13%	1.50%	-5.51%	2.54%
	2012	26		0.95%	0.61%	0.02%	2.65%
	2013	26		0.98%	1.51%	0.03%	8.00%
	2014	26		0.64%	0.41%	0.03%	1.46%
	2015	26		0.51%	0.37%	0.02%	1.34%
	2016	27		0.63%	0.47%	0.02%	1.86%
	2017	27		0.72%	0.63%	0.02%	2.54%
Return on Equity (ROE) %	2008	25	258	11.17%	7.79%	0.78%	28.46%
	2009	25		12.03%	7.96%	-16.41%	23.61%
	2010	25		13.01%	5.50%	5.07%	22.55%
	2011	25		12.18%	14.16%	-45.75%	36.02%
	2012	26		8.91%	5.50%	0.10%	20.60%
	2013	26		6.84%	4.73%	0.35%	16.30%
	2014	26		7.29%	4.64%	0.30%	15.80%
	2015	26		6.75%	5.38%	0.23%	21.00%
	2016	27		7.96%	5.95%	0.35%	24.00%
	2017	27		9.06%	6.10%	0.68%	21.69%
Non-performance Loans Value (Billion VND)	2008	25	258	969	1,941	-	7,898
	2009	25		1,015	2,159	-	9,515
	2010	25		1,595	3,380	1	15,840
	2011	25		2,174	5,494	24	27,076
	2012	26		3,107	5,601	158	27,866
	2013	26		2,922	5,337	171	26,835
	2014	26		2,771	5,295	196	26,412
	2015	26		2,549	3,669	183	13,536

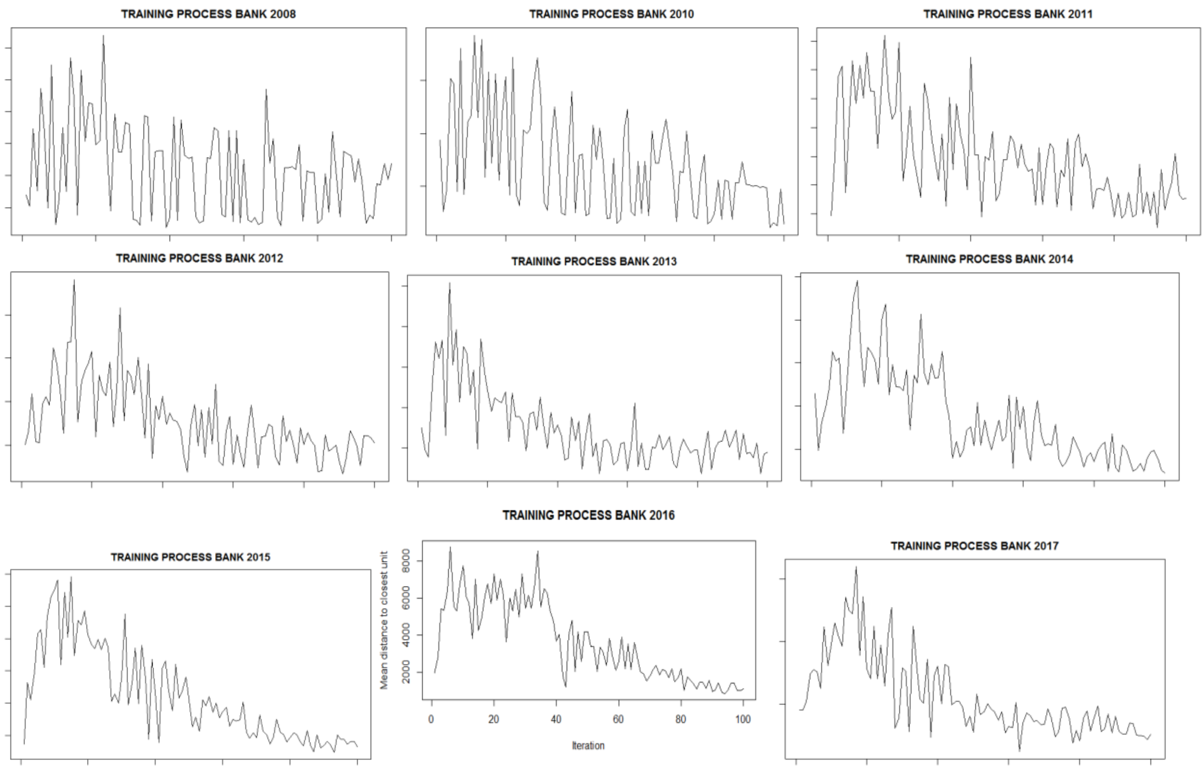
	2016	27		3,132	4,480	210	15,473
	2017	27		3,290	4,394	207	18,000
Non-performance Loans Ratio (%)	2008	25	258	2.06%	1.27%	0.00%	4.79%
	2009	25		1.55%	0.83%	0.00%	2.82%
	2010	25		1.97%	2.16%	0.02%	11.40%
	2011	25		2.30%	1.25%	0.56%	6.10%
	2012	26		3.69%	2.03%	1.35%	8.80%
	2013	26		3.00%	1.44%	0.82%	7.63%
	2014	26		2.29%	1.10%	0.52%	5.16%
	2015	26		1.82%	1.08%	0.34%	5.85%
	2016	27		1.93%	1.18%	0.68%	6.68%
	2017	27		1.85%	0.93%	0.45%	4.28%

Source: Author’s computation

Appendix B

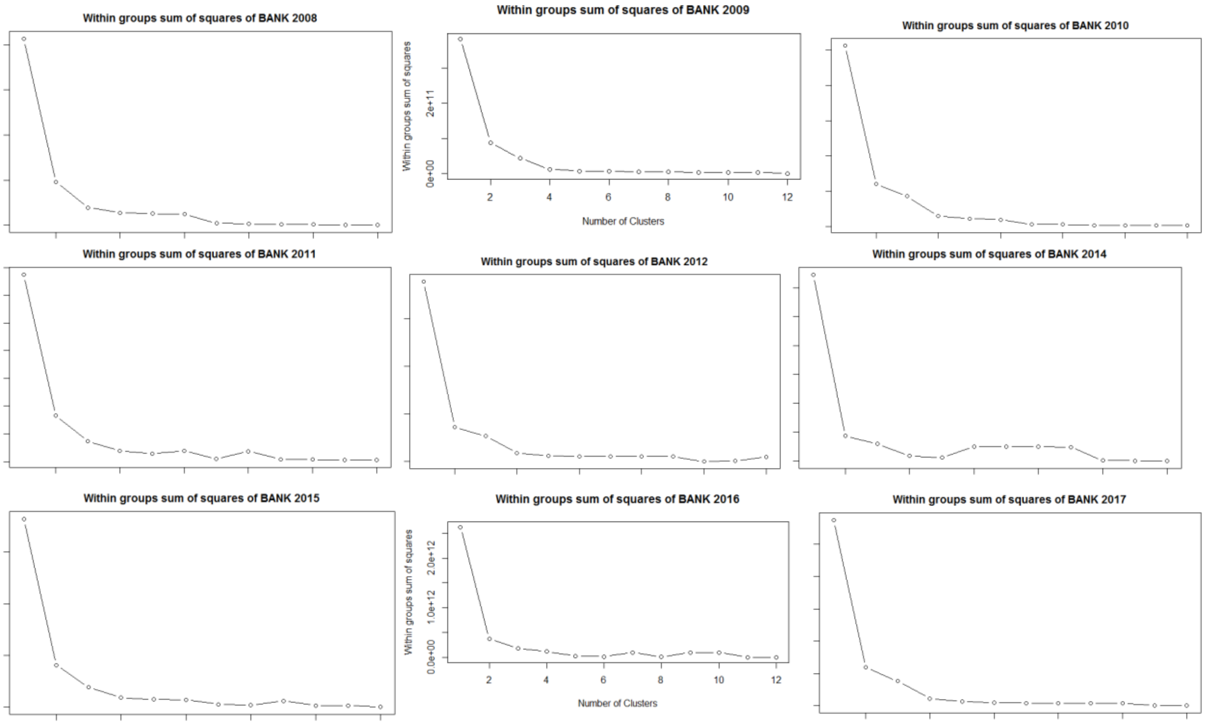
Self-Organisation Map Results

Appendix Figure B.1 The SOM Training Process



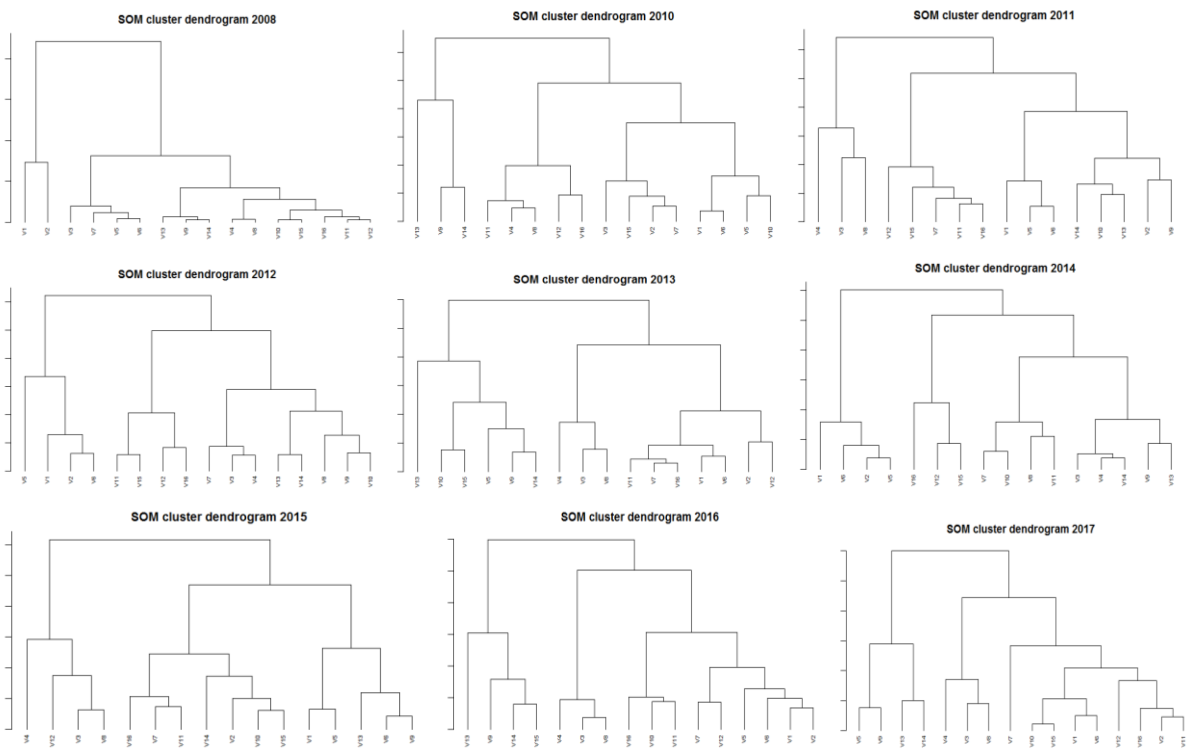
Note: These figures represent mean distance to closest unit value during the training process (from 2008 to 2017).
Source: Author's calculation

Appendix Figure B.2 Within Group Sum of Square Value



Note: These figures show within sum of square value in respond to number of clusters (from 2008 to 2017).
Source: Author's calculation

Appendix Figure B.3 SOM's Dendrograms



Note: These figures represent dendrogram (from 2008 to 2017).
Source: Author's calculations

Appendix Table B.4 Summary Asset Values for Supper-Class Group 1 and Group 2 Banks

ASSET VALUE		2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total average
All banks	Mean	64,557	86,254	117,862	136,965	145,174	160,118	190,771	221,939	253,727	302,639	**168,001**
	S.D.	97,829	115,208	136,439	156,460	169,344	181,040	217,852	263,721	306,542	365,945	**201,038**
Group 2	Mean	113,484	161,638	215,278	212,218	208,969	261,459	313,240	366,784	432,400	516,470	**280,194**
	S.D.	116,641	130,327	143,073	163,203	179,629	196,093	236,021	290,180	340,207	405,516	**220,089**
Group 1	Mean	11,552	16,668	27,940	23,301	24,673	41,886	47,890	52,953	61,309	72,360	**38,053**
	S.D.	12,039	6,913	14,745	7,316	9,955	26,939	25,833	26,402	29,280	34,816	**19,424**

Note: This table shows asset values of Vietnam domestic banks (group 2 and group 1 banks) from 2008 to 2017; S.D. = Standard deviation;

Source: Author’s calculations

Appendix C

Translog Cost Function

Appendix Table C.1 Estimates of Translog Cost Function

Variables	Parameter	Coefficient	t-statistic	P-value
$Ln(Q_{it})$	b_0*	1.36	4.32	0.00
$0.5Ln(Q_{it})^2$	b_1*	-0.04	-1.33	0.19
$Ln(W_{1it})$	a_1	0.64	1.28	0.20
$Ln(W_{2it})$	a_2	0.12	0.23	0.82
$Ln(W_{3it})$	a_3	-0.05	-0.2	0.84
$0.5Ln(Q_{it}) \times Ln(W_{1it})$	b_2*	-0.06	-1.17	0.24
$0.5Ln(Q_{it}) \times Ln(W_{2it})$	b_3*	0.00	-0.07	0.94
$0.5Ln(Q_{it}) \times Ln(W_{3it})$	b_4*	-0.03	-0.68	0.50
$Ln(W_{1it}) \times Ln(W_{2it})$	a_4	-0.05	-0.77	0.44
$Ln(W_{1it}) \times Ln(W_{3it})$	a_5	-0.03	-1.25	0.21
$Ln(W_{2it}) \times Ln(W_{3it})$	a_6	-0.04	-1.06	0.29
$0.5[Ln(W_{1it})]^2$	a_7	-0.02	-0.28	0.78
$0.5[Ln(W_{2it})]^2$	a_8	-0.06	-0.82	0.42
$0.5[Ln(W_{3it})]^2$	a_9	0.06	2.27	0.02
$Trend$	d_1	-0.10	-1.02	0.31
$Trend^2$	d_2	-0.01	-4.54	0.00
$Trend \times Ln(Q_{it})$	d_3 *	0.00	0.48	0.63
$Trend \times Ln(W_{1it})$	d_4	-0.02	-1.33	0.19
$Trend \times Ln(W_{2it})$	d_5	-0.02	-1.64	0.10
$Trend \times Ln(W_{3it})$	d_6	-0.01	-1.1	0.27
Intercept	a_0	-1.41	-0.61	0.54
R-squared: 0.9636		Number of observations: 258		
F-statistic: 278.9		Number of groups: 27		
Prob (F - statistic): 0.0000				

Note: This table shows regression results of the translog function equation (4.5)
* = Parameters will use for calculating marginal cost

Source: Author's computation

Appendix D

Retail Banking Empirical Results

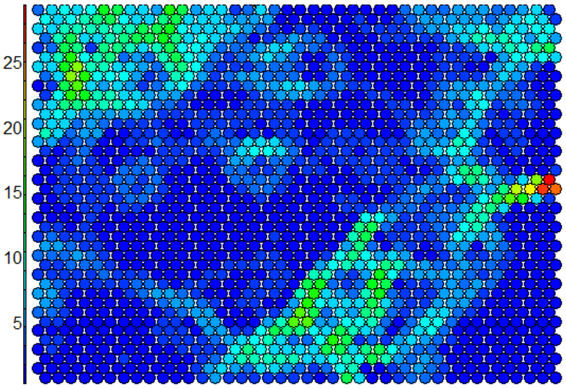
Appendix Table D.1 Summary Information of the Transformed Variables in the Lending Empirical Model (Equation 5.1)

Variable		Obs	Mean	Median	Standard Deviation	Min	Max
Dependent Variables							
Loan classified	Loan_group_2	15 865	0.05	0	0.21	0	1
Independent Variables							
Regiter date	Add_date_3	15 865	7.04	8	2.42	1	10
Industry class	industry_4	15 865	2.04	2	0.36	1	5
Economic class	economic_6	15 865	2.00	2	0.16	1	3
Customer	Customer_8	15 865	1.66	2	0.47	1	2
Loan value	Loan_value_9	15 865	4.99	5	1.86	1	11
Credit term	MIA_type_12	15 865	8.59	8	2.61	1	13
Interest rate	interest_13	15 865	2.67	3	1.61	1	12
Opening date	open_date_14	15 865	6.51	7	0.96	1	7
Loan duration	duration_16	15 865	5.99	5	4.16	1	12
Purpose group	purpose_group_19	15 865	6.76	7	0.88	1	7
Loan purpose	purpose_20	15 865	20.10	19	4.27	1	25
Lending program	HTCV_21	15 865	8.16	8	1.22	1	12
Individual credit	TDCN_22	15 865	9.23	9	3.83	1	15
Sponsor	TOFL_24	15 865	2.73	3	0.68	1	3

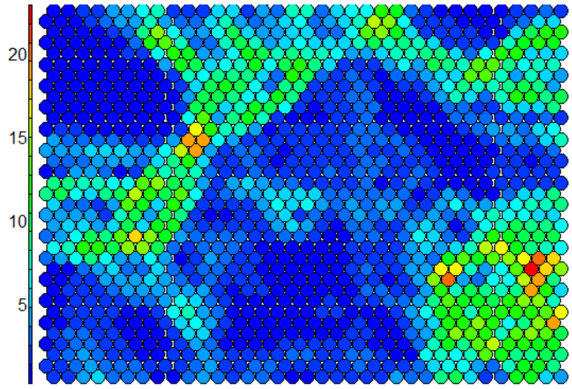
Note: Obs = number of observations

Source: Author's computation

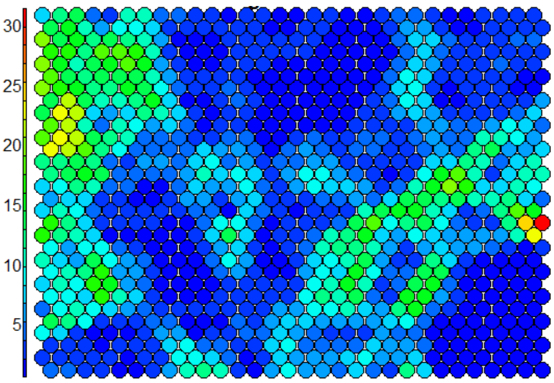
Appendix Figure D.2 SOM Neighbour Distance in Different Grids



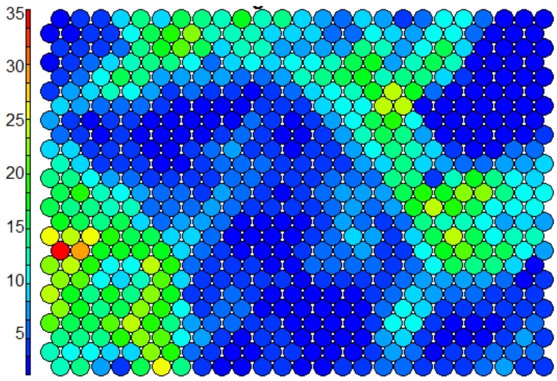
a). 40 x 40 SOM grid



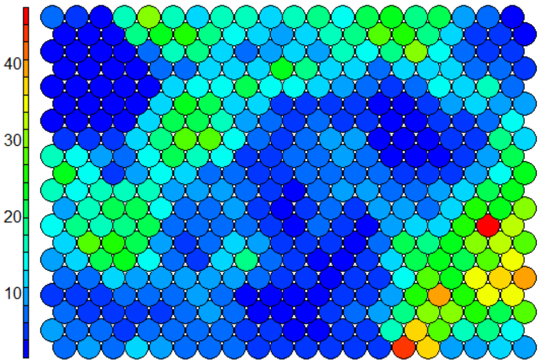
b) 35 x 35 SOM grid



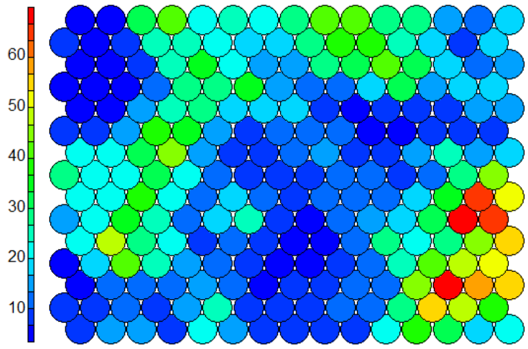
c). 30 x 30 SOM grid



d) 25 x 25 SOM grid



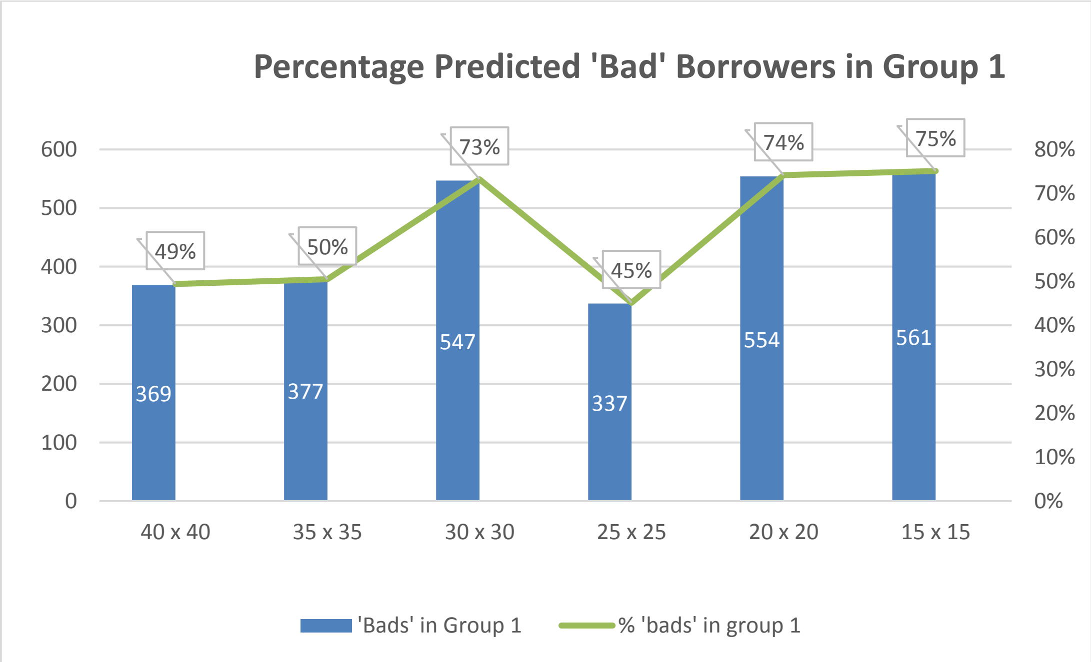
e). 20 x 20 SOM grid



f) 15 x 15 SOM grid

Source: Author's computation

Appendix Figure D.3 Prediction of "Bad" Borrowers in Different SOM grids



Source: Author's computation