The Relevance of Credit Risk in the determination of Commercial Banks’ Profitability: Evidence From Ghana

Godwin Kwabla Ekpe
godwin.rocksonekpe@gmail.com

Follow this and additional works at: https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations

Part of the Finance Commons, and the Statistics and Probability Commons

Recommended Citation

This Dissertation/Thesis is brought to you for free and open access by the Graduate Research & Artistry at Huskie Commons. It has been accepted for inclusion in Graduate Research Theses & Dissertations by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
ABSTRACT

THE RELEVANCE OF CREDIT RISK IN THE DETERMINATION OF COMMERCIAL BANKS’ PROFITABILITY: EVIDENCE FROM GHANA

Godwin Kwabla Ekpe, M.S.
Department of Statistics and Actuarial Science
Northern Illinois University, 2021
Dr. Duchwan Ryu, Director

Existing empirical literature on the relationship between credit risk and bank’s profitability is replete with mixed results. This research investigates the probable effect of credit risk on banks’ profitability by examining the nature of the relationship between two measures of credit risk (Loss provisioning rate and Actual provisioning charge rate) and two measures of profitability (Return on assets and Return on Equity). The investigation is conducted using data on the Ghanaian banking industry. Various modeling techniques are used to fit the data, including frequentist beta regression and Bayesian beta regression models. The results across all models suggest negative linear relationship between actual impairment charge rate and profitability, and a curvilinear relationship between impairment allowance rate and profitability. The curvilinear relationship is contrary to the positive linear relationship suggested by previous studies.

Key Words and Phrases: Within-between models, Beta regression, Bayesian beta regression, Credit risk, Banks profitability, Impairment allowance, ROA, ROE.
THE RELEVANCE OF CREDIT RISK IN THE
DETERMINATION OF COMMERCIAL BANKS’
PROFITABILITY: EVIDENCE FROM GHANA

BY
GODWIN KWABLA EKPE
© 2021 Godwin Kwabla Ekpe

A THESIS SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
MASTER OF SCIENCE

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Thesis Director:
Dr. Duchwan Ryu
ACKNOWLEDGEMENTS

My profound gratitude goes to God for His loving-kindness.

My sincere gratitude to my advisor Dr. Duchwan Ryu for his patience, guidance and contribution towards the success of this thesis in spite of his busy schedules.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 BACKGROUND OF THE STUDY</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Credit Risk and Commercial Banks’ Profitability</td>
<td>5</td>
</tr>
<tr>
<td>3 METHODOLOGY</td>
<td>8</td>
</tr>
<tr>
<td>3.1 Data and Variables Selection</td>
<td>8</td>
</tr>
<tr>
<td>3.1.1 Dependent Variables: Return on average total Asset (ROA) and Return on average total Equity (ROE)</td>
<td>9</td>
</tr>
<tr>
<td>3.1.2 Independent Variables - Internal and External Determinants of Banks’ Profit.</td>
<td>10</td>
</tr>
<tr>
<td>3.2 Data Sources and Choice of Models</td>
<td>12</td>
</tr>
<tr>
<td>3.2.1 Data Sources</td>
<td>12</td>
</tr>
<tr>
<td>3.2.2 Model Specifications</td>
<td>14</td>
</tr>
<tr>
<td>3.2.2.1 Within-Between Model</td>
<td>14</td>
</tr>
<tr>
<td>3.2.2.2 Generalized Linear Models (GLM)</td>
<td>16</td>
</tr>
<tr>
<td>4 DATA ANALYSIS AND EMPIRICAL RESULTS</td>
<td>20</td>
</tr>
<tr>
<td>4.1 Descriptive Statistics</td>
<td>20</td>
</tr>
<tr>
<td>4.2 Estimates for the Within-Between Regression Model</td>
<td>22</td>
</tr>
<tr>
<td>4.2.1 Residual Analysis of the Within-Between Model (WBM)</td>
<td>27</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Descriptive Statistics for full data</td>
<td>20</td>
</tr>
<tr>
<td>4.2</td>
<td>Descriptive Statistics for non-negative returns data</td>
<td>21</td>
</tr>
<tr>
<td>4.3</td>
<td>Within Effects - ROA</td>
<td>24</td>
</tr>
<tr>
<td>4.4</td>
<td>Between Effects from the hybrid model and the Differences from CRE Model - ROA</td>
<td>25</td>
</tr>
<tr>
<td>4.5</td>
<td>Within Effects for ROE</td>
<td>26</td>
</tr>
<tr>
<td>4.6</td>
<td>Between Effects from the hybrid model and the Differences from CRE Model - ROE</td>
<td>27</td>
</tr>
<tr>
<td>4.7</td>
<td>Results: Beta Regression with Logit vs. Arcsine Link Function - ROA</td>
<td>30</td>
</tr>
<tr>
<td>4.8</td>
<td>Results: Beta Regression with Logit vs. Arcsine Link Function - ROE</td>
<td>31</td>
</tr>
<tr>
<td>4.9</td>
<td>Constant Dispersion Bayesian Beta Regression Results - ROA</td>
<td>38</td>
</tr>
<tr>
<td>4.10</td>
<td>Non-constant Dispersion Bayesian Beta Regression Results - ROA</td>
<td>39</td>
</tr>
<tr>
<td>4.11</td>
<td>Constant Dispersion Bayesian Beta Regression Results - ROE</td>
<td>40</td>
</tr>
<tr>
<td>4.12</td>
<td>Non-constant Dispersion Bayesian Beta Regression Results - ROE</td>
<td>41</td>
</tr>
<tr>
<td>4.13</td>
<td>Akaike Information Criterion comparison of fit by logit and arcsine link functions</td>
<td>45</td>
</tr>
<tr>
<td>4.14</td>
<td>Comparison of constant and non-constant precision Bayesian beta models - ROA</td>
<td>45</td>
</tr>
<tr>
<td>4.15</td>
<td>Comparison of constant and non-constant precision Bayesian beta models - ROE</td>
<td>45</td>
</tr>
</tbody>
</table>
A1 Within-Effects from the Within-Between Models after deleting potential outliers.
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Residual Plots of the WBM - ROA.</td>
<td>28</td>
</tr>
<tr>
<td>4.2</td>
<td>Residual Plots of the WBM - ROE.</td>
<td>28</td>
</tr>
<tr>
<td>4.3</td>
<td>Residual Plots of the Beta regression (logit) - ROA</td>
<td>33</td>
</tr>
<tr>
<td>4.4</td>
<td>Residual Plots of the Beta regression (Arcsine) - ROA</td>
<td>34</td>
</tr>
<tr>
<td>4.5</td>
<td>Residual Plots of the Beta regression (logit) - ROE</td>
<td>35</td>
</tr>
<tr>
<td>4.6</td>
<td>Residual Plots of the Beta regression (Arcsine) - ROE</td>
<td>36</td>
</tr>
<tr>
<td>4.7</td>
<td>Comparison of fit by logit and arcsine link functions.</td>
<td>44</td>
</tr>
<tr>
<td>A1</td>
<td>Residual Plots of the WBM - ROA.</td>
<td>53</td>
</tr>
<tr>
<td>A2</td>
<td>Residual Plots of the WBM - ROE.</td>
<td>54</td>
</tr>
<tr>
<td>B1</td>
<td>Diagnostics Plots for the Bayesian Beta regression (Constant Precision) - ROA</td>
<td>55</td>
</tr>
<tr>
<td>B2</td>
<td>Diagnostics Plots for the Bayesian Beta regression (Non-Constant Precision) - ROA</td>
<td>55</td>
</tr>
<tr>
<td>B3</td>
<td>Diagnostics Plots for the Bayesian Beta regression (Non-Constant Precision) - ROA</td>
<td>56</td>
</tr>
<tr>
<td>B4</td>
<td>Diagnostics Plots for the Bayesian Beta regression (Constant Precision) - ROE</td>
<td>56</td>
</tr>
<tr>
<td>B5</td>
<td>Diagnostics Plots for the Bayesian Beta regression (Non-Constant Precision) - ROE</td>
<td>56</td>
</tr>
<tr>
<td>B6</td>
<td>Diagnostics Plots for the Bayesian Beta regression (Non-Constant Precision) - ROE</td>
<td>57</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

The advent of the Financial Services Modernization Act, 1999, gave banks the freedom to provide a wide range of financial intermediation services which they previously could not engage in. Some of these include information production, asset diversification, Investment services (indirect transfer), Brokerage services, Asset transformation (unbundling of investments), Reduction of transaction costs, Maturity intermediation, Denomination intermediation (investment size and foreign currencies), Provision of payment services and Risk management (insurance services) (Achou & Tenguh, 2008).

The provision of any one or combination of these services comes with an inherent associated risk. Depending on the type of services mix being offered, a bank may be faced with credit risk, liquidity risk, interest rate risk, market risk, off-balance sheet risk, foreign exchange risk, country or sovereign risk, technology and operational risk, insolvency risk, and quite recently, model risk. However, among these risks, credit risk is one of the greatest concern to most banks since for many commercial/universal banks around the globe, credit creation remains the major income generating activity.

Credit risk has proven to be the risk that can easily prompt bank failure. The greatest impact of financial crises experience over the years has been on the banking industry, where some banks which were previously performing well suddenly announced huge losses with some of them having to be bailed out by State or National governments ¹.

¹An example was the passage into U.S. law on October 3, 2008, of the $700 billion financial-sector rescue plan called the Emergency Economic Stabilization Act of 2008
Almost all of the major reported cases of havoc caused by the 2007/08 credit crunch came from the developed economies of the world. Indeed not much has been documented of the impact of the credit crunch on the banking industry in third world countries, particularly in Africa. In Ghana for example, one could argue that the 2007/08 credit crunch could not have had a significant impact on the Ghanaian banking industry. This is because the Ghana Stock Exchange (GSE) was arguably not really advanced and efficient and its integration with the Ghanaian economy was not at an advanced stage. Many Ghanaians prefer to hold their assets in landed property. Indeed, the Ghana Banking Survey (2009) by PricewaterhouseCoopers (PwC) reported that, despite the global financial crisis, the Ghana banking industry remained stable.

However, the Ghanaian banking industry did experience reduction in average profit level during the period of the credit crunch relative to previous years. The industry’s average impairment charge to gross loans almost doubled from 2.2% in 2008 to 4.2% in 2009 after increasing from 1.5% in 2007. Indeed, Barclays Bank of Ghana Ltd (BBGL) which prior to 2008 has been the most profitable bank in Ghana (PricewaterhouseCoopers’ 2008 Ghana Banking Survey Report) recorded unprecedented losses in 2008 and 2009 with profit before tax margins (PBTM) of −6.3% and −13.8% respectively having previously posted a PBTM of 36.8% in 2007. Not only did profit erode between the 2007-2009 period but BBGL’s asset quality deteriorated considerably over the period with Impairment allowance/gross loans and advances sky-rocketing from 2.6% in 2007 to 8% in 2008 and then to 17.7% in 2009. The bank’s Impairment charge/ gross loans and advances over the period had consequently seen unprecedented increment over the period, moving from 0.8% in 2007 to 6.0% and 9.8% in 2008 and 2009 respectively before subsiding to 3.8% in 2010. While BBGL was not listed on the GSE, it was a subsidiary bank of Barclays PLC, UK. As such, though these statistics give an indication of some problems with the soundness of its credit portfolio, they also present...
anecdotal evidence suggesting that the Global credit crunch might have affected the bank’s profitability.

Nonetheless, many previous studies on credit risk and profitability have reported positive relationship between credit or loan delinquency rates and profitability (see, Amidu and Hinson (2006); Ara, Bakaeva, and Sun (2009); Boahene, Dasah, and Agyei (2012); Kithinji (2010)).

This thesis therefore seeks to revisit the debate as to whether credit risk ultimately has a positive relationship with banks’ profitability. This is done by exploring a series of regression models based on data structure and deciding which modeling technique is best suited for an empirical investigation of the statistical relationship between credit risk and commercial banks’ profitability in the context of emerging markets using the Ghanaian banking industry as a test case.

The top ten commercial banks (by asset) were selected for this empirical investigation over twelve years (2005-2016). Profitability is measured by two proxies - Return on Asset (ROA) and Return on Equity (ROE) – and modeled as a function of some selected internal and external factors. The internal factors which are used as measures of credit risk include the ratio of Impairment Allowance to Total Loans & Advances (IMP_ALWR), and its squared term, and the ratio of Impairment Charge to Total Loans & Advances (IMP_CHGR). Bank size and Capital Adequacy Ratio(CAR) are treated as internal control variables while the external factors such as inflation rate and GDP growth rate are treated as external control variables.

The study employs a within-between regression model analysis as a preliminary exploratory model to investigate the relationship between credit risk exposure and bank’s profitability. This is done in order to make the results comparable to previous studies that use standard linear regression techniques although the response variable in many of those instances are ratio or proportion data. The results show that the effect of credit risk on banks’ profitability measured by the rates of Return on Assets of banks and Return on share-
holders Equity is significantly negative for actual impairment charge rate (IMP.CHGR). A one-percent increase in actual impairment charge rate is associated with 0.54% decrease in ROA and 4.1% decrease in ROE. However, a one-percent increase in impairment allowance (IMP.ALWR) corresponds to approximately one-percent increase in ROA and 4.9 increase in ROE. This agrees with previous studies. The point of departure, however, is the inclusion of the squared term of IMP.ALWR in the current study which then captures the cumulative effect provisioning rates.\textsuperscript{2} The thesis is that, increases in IMP.ALWR may initially lead to increases in profitability (ROA and ROE) but after a while subsequent increases in IMP.ALWR would eventually have negative cumulative effect on profitability (ROA and ROE). Results from predictive Generalized Linear Models (GLMs) such as frequentist Beta regression and Bayesian Beta regression models specified confirm this curvilinear relationship.

The rest of this thesis is structured as follows. Chapter 2 gives a background review of the literature relating to credit risk and banks’ profitability, Chapter 3 describes the data and methodology followed by data analysis and discussion of the empirical results in Chapter 4; Chapter 5 concludes.

\textsuperscript{2}Provisioning is the same terminology as impairment allowance.
CHAPTER 2
BACKGROUND OF THE STUDY

2.1 Credit Risk and Commercial Banks’ Profitability

The banking industry across the world has become much more competitive than it was a century ago. Though banking deals largely in financial products, it is also seen as a service industry. As such bank managers seek to optimize profit and service through protection of stakeholders’ investment and provision of consistent and enhanced customer service. A major threat to achieving the afore-stated goal is Credit Risk, which is thought as most serious risk faced by banks as it impacts directly on banks’ net worth.

Credit risk is the probability that a borrower or issuer of a debt instrument will not be able to honor their debt obligations - principal and interest repayments - to the lender at maturity (Van Greuning & Brajovic Bratanovic, 2009). In other words, credit risk arises when the debtor, a borrower, counterparty, or an obligor is unable to repay part or whole of the debt to the creditor at the due time and thereby exposing the lender to loss (Colquitt, 2007). With respect to the banking sector, delayed loan payments or total defaults can give rise to cash flow problems leading to a liquidity crisis.

Generally, credit and credit risk related items account for at least 70 percent of banks’ balance sheet and thus constitutes a major source of potential losses and bank failures (Mitku, 2015). Credit risk management in the banking industry therefore focuses on ensuring credit risk exposures are kept within tolerable limits so a bank’s risk-adjusted rate of return can be maximized. The proportion of a bank’s balance sheet accounted for by the credit risk items
is not by itself as problematic as when banks fail to diversify their credit or loans portfolio. Naturally, banks have the tendency to extend credit to entities or clients with whom they have an existing relationship. This often leads to banks concentrating their lending business to specific clients, or specific industries or geographic areas.

It is important to note that credit risk comprises both default risk - the risk that a borrower is unable to service their debt - and the risk of a decline in the credit worthiness of the borrower. When one or a few of a bank’s debtor-client suffers business mishaps that affect their ability to honor their debt obligations or issues that negatively impact their credit rating, such a bank is likely to face a huge credit risk crisis if they have a concentrated lending portfolio. The reason is that, the problems giving rise to the mishap may be industry-wide or may affect an entire geographical area and as such all clients in the concentrated industry or geographic area may be affected. Whilst default triggers a total or partial loss, a deteriorated credit rating of a debtor or counterparty is also likely to result in a loss because the required market yield would have to adjust upwards to compensate the higher risk. This rather often leads to a decline in value (Bessis, 2011). The worst case of credit risk is experienced when the borrower defaults either due to their unwillingness or inability to fulfill the obligations (Crouhy, Galai, & Mark, 2006). Such debts become ”bad debts”.

Not only do bad debts erode Banks profitability, but they also negatively affect their cash flows and liquidity. All of these compromises both their creditors’ (including customer savings) and shareholders’ funds and thus endangers the very existence of banks – the apparent reason credit risk is considered “the principal cause of bank failures” (Van Greuning & Brajovic Bratanovic, 2009).

Generally, an indicator of banks’ capacity to carry risk and/or increase their capital is their profitability. Profitability indicates banks’ competitiveness and measures the quality of management (Zou & Li, 2014). If profitability and asset quality erode due to bad debts, bank management’s credibility and capacity, particularly with respect to credit risk management,
is called in to question. These are but a few reasons why bank management often put in place a robust credit risk management framework to adequately deal with the threat posed by credit risk. To avoid or reduce the occurrence of both partial and total losses, banks normally give a serious consideration to any underlying collateral of potential borrowers relative to its financial adequacy as part of the credit risk assessment (Santomero, 1997). On the whole, since profitability is the key performance index for bank management, all efforts are made to ensure credit risk management function implements formally laid down policies and procedures that prescribe the scope of investment and financing assets and how they are originated, appraised, monitored, and collected with the least subjectivity or discretion as possible (Van Greuning & Brajovic Bratanovic, 2009).

From the foregoing, it is expected that when analyzing the performance of commercial banks, consideration is given to credit risk measures and how they affect their profitability. The determinants of commercial banks’ profitability can be categorized into two, namely those controlled by bank management (internal determinants) and those beyond the control of management (external determinants) (Guru, Staunton, & Balashanmugam, 2002; Kossidou, Tanna, & Pasiouras, 2005). Financial statements of commercial banks usually mirror the internal determinants as they are indices that often reflect policy decisions relating to sourcing and usage of debts, equity capital, liquidity management, overhead cost and general expenses management (Guru et al., 2002). The external factors mostly border on the micro- and macro-economic factors of the economy within which the banks operate.
CHAPTER 3
METHODOLOGY

3.1 Data and Variables Selection

The core objective of this thesis is to estimate the relationship between credit risk and profitability in the Ghanaian banking industry using regression techniques. Profitability, in this study, is measured or proxied by Return on Assets (ROA) and Return on Equity (ROE) since these are the two most important ratios for management and shareholders. Banks manage credit risk with the view of boosting interest income and minimizing loan losses from credit default. It is thus expected that banks with better credit risk management practice to have lower non-performing loans, impairment (provisioning) allowances and impairment charges. In other words, profitability (in terms of ROA and ROE) can be determined by the quality of credit risk management. In this light, this thesis considers credit risk as an independent variable. The study uses Impairment Allowance/Total Loans ratio (IMP_ALWR) and Impairment Charge/Total Loans ratio (IMP_CHGR) as proxies for credit risk management. The lower these ratios are the better or the higher the quality of credit risk management in the bank.

As stated earlier, banks’ profitability is likely to be influenced by both internal and external factors. As such the methodology of this thesis is to include some relevant external factors in the regression model. There are two alternative model specifications testing the nature of the relationship between profitability and credit risk management. The only difference is that the first model has ROA as the dependent variable while the second model
has ROE as its response variable. The full complements of the variables used in this study are discussed below.

### 3.1.1 Dependent Variables: Return on average total Asset (ROA) and Return on average total Equity (ROE)

There has been an extensive reliance on the Return on Asset (ROA) and the Return on Equity (ROE) as measures of profitability (Ara et al., 2009). Even though the ROA does not incorporate off-balance sheet items (such as operating leases, letters of credit, unused commitments) and thus may be overstated, it has become one of the most important measures of profitability in recent banking literature (Golin & Delhaise, 2013). A few of the previous studies that made use of ROA as a measure of profitability include Demirguc-Kunt, Huizinga, et al. (1998), Bashir (2001), Alrashdan (2002), Hasan-Rokem (2003), Naceur (2003), and Alkassim (2005). The ROA depicts the effectiveness of a bank’s management in using its assets to generate income and is computed as the ratio of net income (profit after tax) to average total assets.

An alternative measure of profitability adopted by this thesis is the Return on Equity (ROE). Computed as the ratio of net income to average equity of shareholders, the ROE measures the income earned on each unit of shareholders’ funds.

An established relationship between ROE and ROA can be stated as below (Achou & Tengu, 2008; Mishkin, 2013):

\[
ROE = ROA \times AE
\]

where ROA is defined explicitly as the ratio of net-interest income plus non-interest income less provisioning to total assets, and AE is Equity Multiplier which is defined as the ratio of
total assets to total equity. Given the above Equation (3.1), a bank is able to increase ROE by ensuring a higher ROA through increasing net-interest income and non-interest income and by decreasing loan provisioning through (Sensarma & Jayadev, 2009). Consequently, credit risk is expected to have a negative correlation with ROA and ROE.

### 3.1.2 Independent Variables - Internal and External Determinants of Banks’ Profit

**Impairment Allowance (Loan Loss Provisioning) and Impairment Charge Rates.**

Impairment Allowance and Impairment Charge Rates are the main explanatory variables of interest. These are indicators of the existence of Non-Performing Loans (NPLs) on a bank’s loan portfolio which are evidence of worsening credit risk in banking system (Miller & Noulas, 1997). Impairment Allowance Rate (IMP_ALWR) is the ratio of Impairment Allowance over gross loans and advances while Impairment Charge Rate (IMP_CHGR) is the ratio of the actual impairment charged over gross loans and advances. Lower values of these ratios point to a higher credit quality and vice versa. It is worth noting that loan loss provisioning is another term used for impairment allowance (see for example, Sufian (2009)).

**Capital Adequacy Ratio (CAR)**

There are various ways by which the CAR is measured. It can be measured as ratio of regulatory capital to Risk Weighted Assets, the ratio of tier 1 capital to Risk Weighted Assets or the ratio of Non-performing loans to Capital. Another variant that is often used as a proxy for CAR and adopted by this thesis is simply the ratio of total equity to total asset. This ratio reflects to the sufficiency of the amount of capital to absorb any shocks that the bank may experience. On the one hand, it is expected that with higher equity levels relative
to assets levels, there would be less need for external funding and therefore the higher the profitability of the bank (Kosmidou, 2008). On the other hand, based on Equation (3.1), we can express CAR as the inverse of AE. This implies that \( \text{ROE} = \frac{\text{ROA}}{\text{CAR}} \) and thus with a given ROA, the higher the capital, the lower the ROE.

**Bank Size**

The size of a bank is likely to affect its profitability (Haslem, 1968). The bigger the bank, the more diversified its asset portfolio is likely to be. A more diversified portfolio may impact favorably on risk and product portfolio as the cost of gathering and processing information can be reduced by Economies of scale (Boyd & Runkle, 1993; Kutsienyo, 2011). However, extremely large bank size may also breed huge bureaucracy and agency cost that may offset its gains (Berger, Hanweck, & Humphrey, 1987; Kutsienyo, 2011). Nonetheless, many studies in the past have found a positive and significant relationship between bank size, and profitability and capital ratios (Akhavein, Berger, & Humphrey, 1997; Bikker & Haaf, 2002; Bourke, 1989; Goddard, Molyneux, & Wilson, 2004; Haslem, 1968; Molyneux & Thornton, 1992; Short, 1979; Smirlock, 1985)

Since all other variables in this study are ratios or percentages, using total assets values may be incongruent with the covariates’ scale. As such, this thesis also follows previous studies and uses natural log of total assets (LN_TA) so as to avoid any such distortions.

**External Determinants**

External factors that can possibly affect the profitability of a bank may be political, economic, technological, and sociological in nature, depending on the environment in which it operates. It is important to note that even though banks do not have control over these external factors, the effectiveness of banks management can be seen in how these factors are anticipated by management and how the banks are strategically positioned to take advantage of them.
For the purpose of this study, a major external environment of interest is the economic environment as it is key to the corporate goal of maximizing shareholders’ wealth in any industry. Some key macroeconomic indices top managers, especially in the banking industry look out for include growth rate, inflation rate, and interest rates. I therefore include these as covariates in the regression models. Other variables considered are two binary indicator variables which are whether a bank has local origin (coded as 1) or foreign origin (coded 0), and whether the bank is listed on the Ghana stock exchange (coded 1) and zero (0) otherwise.

3.2 Data Sources and Choice of Models

3.2.1 Data Sources

The data used in this study come from a myriad of sources. Financial data on the Ghanaian Banking industry were obtained from the Bank of Ghana Fiscal Stability Reports, annual financial statement reports of the various banks and other reports issued by industry watchers including the annual Ghana Banking Survey Reports by PricewaterhouseCoopers (PwC) and the Ghana Association of Bankers (GAB). Data on the internal explanatory variables such as IMP_ALWR, IMP_CHGR, CAR and LN_TA as well as the two dependent variables (ie., ROA and ROE) were computed using figures from the financial statements (annual reports) of each of the ten commercial banks over a twelve year period spanning from 2005 to 2016. The banks included in this study are those that occupy the top ten position as categorized by their total operating assets in 2016. These banks include, Ecobank Ghana (EBG), GCB Bank (GCB), UT Bank Ghana Ltd (UGL), Barclays Bank Ghana Ltd (BBGL), Stanbic, Standard Chartered Bank (SCB), Fidelity, United Bank of Africa (UBA), Cal Bank
Ltd (CAL), and Zenith Bank Ltd (ZBL). These ten banks contribute to 73.1% of total operating assets of the Ghanaian Banking industry. Additionally, their industry share of both total deposits and total advances amount to 72.5%.¹

Data on the Ghanaian macroeconomic environment (external factors) such as Inflation (InfR), Average Policy (Lending) rate (Ave_IntRate) and real GDP growth rate (GDP_GR) were also obtained from the Bank of Ghana reports as well as FRED Data.

The models explored in this study focus more on determining whether Impairment Allowance/Total Loans ratio (IMP_ALWR) and/or Impairment Charge/Total Loans ratio (IMP_CHGR) are significant in explaining the rates of profitability of banks. The IMP_ALWR is used in place of Non-performing Loans/Total Loans (NPLR) which has been used in previous studies. Data for NPLR for all most all the banks used in this study are not publicly available. However, data for IMP_ALWR is available from the banks’ financial statements. Since the IMP_ALWR is a direct indicator or reflection of the level of non-performing loans on the books of banks, it can be used in place of NPLR as proxy for credit risk management. Additionally, for parametric models, I include the squared term of IMP_ALWR (IMP_ALWRsqr) to capture the long term or cumulative effect of this variable since impairment allowance, for example on a two-month delinquent loan which remains delinquent, is expected to increase the next period. Therefore, we can theoretically define the profitability function as:

\[
\text{Profit(ROA, ROE)} = f(\text{IMP_ALWR, MP_ALWRsqr, IMP_CHGR, CAR, LN_TA, InfR, GDP_GR, Ave_IntR, Origin, Listed})
\]

This study is distinct with respect to the previous studies on the subject of credit risk and profitability in Ghana in that the scope of data used and the number and nature of

¹A detailed analyses on the ranking of banks in Ghana can be found in the Ghana Banking Survey Report, 2017 by PricewaterhouseCoopers (PwC).
banks analyzed are different. None of the existing studies, to the best of my knowledge has covered up to a 12-year period and for 10 banks. The closest is by Opoku, Angmor, and Boadi (2016) who used data covering 2007 to 2014 and focusing only on the seven banks listed on the Ghana Stock Exchange (GSE). This current study includes four of the listed banks and six banks that are not listed on the GSE.

3.2.2 Model Specifications

3.2.2.1 Within-Between Model

To start with parametric specifications based on (3.1) for each of the two dependent variables – ROA and ROE, the longitudinal or panel structure of the data has been considered to specify a within-between (WB) model, initially proposed by Mundlak (1978) and extensively elaborated upon by Chamberlain (1984) and Wooldridge (2010). This model, which is sometimes called hybrid, between-within (see Allison (2009) and Bell and Jones (2015)) or a correlated random effects (CRE) model among econometricians (see Chamberlain (1984) and Wooldridge (2010)), allows for the estimation of both fixed and random effects in the same model. Given \( i = 1, 2, 3, \ldots, n \) entities (banks) observed and measured at times \( t = 1, 2, 3, \ldots, T \). Let \( y_{it} \) denote the outcome (profitability) for bank \( i \) in the period \( t \), \( x_{it} \) denote the vector of measured time-varying predictors that vary within and between individual entities(banks). Then, the hybrid linear WB model (Allison, 2009) can be described as:

\[
y_{it} = \alpha_0 + \beta'(x_{it} - \bar{x}_i) + \kappa'c_i + \lambda'\bar{x}_i + u_i + \epsilon_{it},
\]  

(3.3)
where $\alpha_0$, $\beta$, $\kappa$, $\lambda$ are unknown regression coefficients, $u_i$ is the random intercept and $\epsilon_i$ is the random error.

The within-effect (i.e., the fixed-effects) in Equation (3.3) is estimated by $\beta$, the between-effects by $\lambda$ and $\kappa$ estimates the effect of time-invariant variables (Mundlak, 1978; Neuhaus & Kalbfleisch, 1998; Schunck, 2013). The consistency of these parameter estimates relies on the assumptions that $E(u_i|x_{it},c_i) = 0$ and $u_i|x_{it},c_i \sim N(0, \sigma_u^2)$.

The assumption of zero correlation between $u_i$ and $(x_{it}, c_i)$ may be too strict to be met at times, especially where data being analyzed may not contain all relevant time-variant and time-invariant predictors. In such circumstances, a variant of the model in Equation (3.3) called correlated random effects (CRE) model is specified as:

$$y_{it} = \alpha_0 + \beta'x_{it} + \kappa'c_i + \tau'\bar{x}_i + \nu_i + \epsilon_{it}. \quad (3.4)$$

In this model, the random intercept $u_i$ takes the form:

$$u_i = \tau'\bar{x}_i + \nu_i,$$

where $\bar{x}_i$ accounts for any correlations between the intra-cluster error ($u_i$) and $x_{it}$ (Halaby, 2003; Schunck, 2013). The intra-cluster error structure as specified here implies that CRE does not strictly require $u_i$ to be uncorrelated with $x_{it}$. Thus the variable $\nu_i$ is the part of $u_i$ which is uncorrelated with $x_{it}$. Under these circumstances, the least square estimates of $\beta$ and $\kappa$ remain the unbiased and consistent estimates of the within fixed-effects and the individual effects of the time-invariant variables, respectively as in Equation (3.3). However, the coefficient of $\bar{x}_i$, $\tau$ now has a different interpretation. This coefficient now estimates the
difference of the within and the between effects (Mundlak, 1978; Schunck, 2013). In other words, the between effects in using Equation (3.4) is obtained by adding $\tau$ to $\beta$.

### 3.2.2.2 Generalized Linear Models (GLM)

The data analyzed has a few occasions where some of the banks recorded negative net returns (i.e., losses) at some point during the twelve-year period covered in this study. This implies that ROA and ROE of the affected banks are negative for the years the losses were made and such ROAs and ROEs are within the bound $(0, -1)$. It is important to note that most of the instances of the negative net returns were recorded by new banks in their first and/or second year of operations - a phenomenon that is commonplace for many banks as it takes time to recoup overhead cost and to break even. Since the occurrence of these negative returns do not necessarily reflect how well a bank will do in subsequent years, especially in relation to credit risk management, we can remove such observations from the data as they only serve as a source of undue bias in the estimation of the parameters. If we remove these few instances of negative returns from the data, then the dependent variables (ROA and ROE), which are all ratios now have all values in the open standard unit interval $(0, 1)$. Therefore, the second set of models I explored are Beta regression and Bayesian Beta regression.

The reason for exploring Beta regression is to be sure the fitted model to the data accounts appropriately for over dispersion that is often present in proportions or ratios which are beta distributed variables. Note also that, with the removal of the negative return observations, the data would not be a balanced panel. In this regard, the analysis here is not intended to draw on the panel structure.
Regarding the response $y$ as the ratio of returns to assets or equity, it can be modeled by the Beta distribution with the following pdf:

$$f(y; p, q) = \frac{\Gamma(p + q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1 - y)^{q-1}, \quad p > 0, q > 0, 0 < y < 1,$$

where $p$ and $q$ are two shape parameters and $\Gamma(\cdot)$ is the gamma function. Using the reparameterization as proposed by Ferrari and Cribari-Neto (2004) by setting the variate mean $\mu = p/(p + q)$ and the precision parameter $\phi = p + q$, we can rewrite the pdf as:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)(1 - \mu\phi)} y^{\mu\phi - 1}(1 - y)^{(1 - \mu)\phi - 1}, \quad 0 < y < 1,$$

with $0 < \mu < 1$ and $\phi > 0$. In this case, we have $y \sim Beta(\mu, \phi)$, where $E(y) = \mu$ and $var(y) = \mu(1 - \mu)/(\phi + 1)$.

Assuming a constant dispersion $\phi_i \equiv \phi$, and given random samples $y_1, ..., y_n$ of the dependent variable (ROA and ROE) such that $y_i \sim Beta(\mu_i, \phi), i = 1, ..., n$, we can model a Beta regression for $y$, on $x_{i1}, ..., x_{ik}$ using link function $g(\cdot)$ of mean values as:

$$g(\mu_i) = x_i'\beta, \quad i = 1, ..., n,$$

where $x_i = (x_{i1}, ..., x_{ik})'$ denotes the vector of $k$ predictors as defined earlier with $n = 110$ ones for the intercept and $\beta = (\beta_1, ..., \beta_k)'$ is $k \times 1$ parameters to be estimated. The log-likelihood function can be written as

$$l(\beta, \phi) = \sum_{i=1}^{n} \log \Gamma(\phi) - \log \Gamma(\mu_i\phi) - \log \Gamma((1 - \mu_i)\phi)$$

$$+ (\mu_i\phi - 1) \log y_i + \{(1 - \mu_i)\phi - 1\} \log(1 - y_i). \quad (3.5)$$
There are two main advantages for using a link function in the regression structure. First, applying a link function to $\mu_i$ ensures the two sides of the regression equation have values in the real line and secondly, it also affords us the flexibility to implement different link functions so as to choose the function that best fits the data (Zeileis, Cribari-Neto, Grün, & Kos-midis, 2010). Therefore in this study, I employ two link functions – logit and arcsine link functions as variance-stabilizing transformations – and decide which best fits the data.

For example, using the logit link function $g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$, the regression model can be written as:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1x_{i1} + \beta_2x_{i2} + \ldots + \beta_kx_{ik}, \quad i = 1, \ldots, n. \quad (3.6)$$

Or in matrix form, we can write

$$g(\mu) = X\beta \quad \Leftrightarrow \quad \mu = \frac{1}{1 - \exp(-X\beta)}.$$

Similarly, for the arcsine link function $g(\mu) = 9\{\sin^{-1}(\sqrt{\mu}) - \sin^{-1}(\sqrt{0.5})\}$, we have

$$g(\mu) = X\beta \quad \Leftrightarrow \quad \mu = \sin^2\left(\frac{1}{9}X\beta + \sin^{-1}(\sqrt{0.5})\right).$$

I use the `betareg()` function of `betareg` package in R to estimate parameters for the model with logit link function. With respect to the model with arcsine link function, I used an adjusted version of `betaglm.R` to estimate the parameters.\(^2\) It is worth noting that these

\(^2\)The relevant R code which is presented in the appendix was written with guidance provided by Thesis Director, Professor Duchwan Ryu.
estimations are Maximum Likelihood Estimations of the log-likelihood function in Equation (3.5).

The final model I explored is a Bayesian beta regression model. The aim is to examine how different the estimated coefficients would be compared to the classical beta regression model. Rather than treating the parameters as non-random, the Bayesian approach treats the mean and the dispersion parameters as random. This requires assigning prior distributions on these parameters. The empirical posterior distribution of these parameters are of interest in this study. Thus, letting $\Theta = (\beta, \gamma)$, where $\beta$ and $\gamma$ are mean and precision parameter vectors respectively, and denoting the likelihood function by $L(\Theta|y)$ and $\pi$ as the joint prior distribution, we can state the posterior distribution by applying the Bayes’ rule as:

$$p(\Theta|y) \propto p(y|\Theta)\pi(\Theta).$$

Often times, Bayesian practitioners assign a prior of normal distribution on the model parameters. While normal prior may be appropriate for other models, it would be inappropriate to assign a normal prior particularly to the dispersion parameter $\gamma$ in the present model. This is because we are dealing with a proportions data $y$ which is on the positive unit interval $(0, 1)$. Thus the dispersion cannot be expected to have negative values. As such a reasonable prior has to be a non-negative distribution.

I used the bugs function of the R2WinBUGS package in R as well as the bayes: betareg function in Stata to implement the Bayesian beta regression assuming beta distribution for the response variables, gamma prior for $\gamma$ and normal prior for $\beta$. The logit link function is specified for the mean and logarithm function for the dispersion.
CHAPTER 4
DATA ANALYSIS AND EMPIRICAL RESULTS

This chapter discusses the results from the data analysis, with focus on giving context to the estimated parameters, model diagnostics and selection.

4.1 Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROA</td>
<td>120</td>
<td>0.02</td>
<td>0.04</td>
<td>-3.15</td>
<td>13.68</td>
<td>-0.2</td>
<td>0.07</td>
</tr>
<tr>
<td>ROE</td>
<td>120</td>
<td>0.21</td>
<td>0.22</td>
<td>-1.7</td>
<td>4.39</td>
<td>-0.82</td>
<td>0.54</td>
</tr>
<tr>
<td>IMP_ALWR</td>
<td>120</td>
<td>0.06</td>
<td>0.05</td>
<td>1.47</td>
<td>2.03</td>
<td>0.004</td>
<td>0.22</td>
</tr>
<tr>
<td>I(IMP_ALWR^2)</td>
<td>120</td>
<td>0.01</td>
<td>0.01</td>
<td>2.83</td>
<td>8.42</td>
<td>0.000016</td>
<td>0.05</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>120</td>
<td>0.03</td>
<td>0.03</td>
<td>2</td>
<td>4.87</td>
<td>-0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>CAR</td>
<td>120</td>
<td>0.12</td>
<td>0.05</td>
<td>-0.27</td>
<td>-0.31</td>
<td>0.02</td>
<td>0.24</td>
</tr>
<tr>
<td>LN_TA</td>
<td>120</td>
<td>2.95</td>
<td>0.58</td>
<td>-0.77</td>
<td>-0.06</td>
<td>1.27</td>
<td>3.86</td>
</tr>
<tr>
<td>InfR</td>
<td>120</td>
<td>0.14</td>
<td>0.03</td>
<td>0.1</td>
<td>-1.51</td>
<td>0.09</td>
<td>0.19</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>120</td>
<td>0.07</td>
<td>0.03</td>
<td>1.07</td>
<td>0.54</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>120</td>
<td>0.17</td>
<td>0.04</td>
<td>1.09</td>
<td>0.03</td>
<td>0.13</td>
<td>0.26</td>
</tr>
<tr>
<td>Origin</td>
<td>120</td>
<td>0.4</td>
<td>0.49</td>
<td>0.4</td>
<td>-1.85</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Listed</td>
<td>120</td>
<td>0.48</td>
<td>0.5</td>
<td>0.07</td>
<td>-2.01</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 4.2: Descriptive Statistics for non-negative returns data

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROA</td>
<td>110</td>
<td>0.08</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.86</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>ROE</td>
<td>110</td>
<td>0.26</td>
<td>0.13</td>
<td>0.1</td>
<td>-0.73</td>
<td>0.01</td>
<td>0.54</td>
</tr>
<tr>
<td>IMP_ALWR</td>
<td>110</td>
<td>0.06</td>
<td>0.05</td>
<td>1.47</td>
<td>2.09</td>
<td>0.004</td>
<td>0.22</td>
</tr>
<tr>
<td>I(IMP_ALWR^2)</td>
<td>110</td>
<td>0.01</td>
<td>0.01</td>
<td>2.87</td>
<td>8.74</td>
<td>0.000016</td>
<td>0.05</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>110</td>
<td>0.02</td>
<td>0.02</td>
<td>2.21</td>
<td>6.14</td>
<td>-0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>CAR</td>
<td>110</td>
<td>0.13</td>
<td>0.04</td>
<td>-0.23</td>
<td>-0.05</td>
<td>0.03</td>
<td>0.24</td>
</tr>
<tr>
<td>LN_TA</td>
<td>110</td>
<td>3.03</td>
<td>0.52</td>
<td>-0.84</td>
<td>0.57</td>
<td>1.27</td>
<td>3.86</td>
</tr>
<tr>
<td>InfR</td>
<td>110</td>
<td>0.14</td>
<td>0.04</td>
<td>0.12</td>
<td>-1.54</td>
<td>0.09</td>
<td>0.19</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>110</td>
<td>0.07</td>
<td>0.03</td>
<td>1.03</td>
<td>0.34</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>110</td>
<td>0.17</td>
<td>0.04</td>
<td>1</td>
<td>-0.24</td>
<td>0.13</td>
<td>0.26</td>
</tr>
<tr>
<td>Origin</td>
<td>110</td>
<td>0.42</td>
<td>0.5</td>
<td>0.33</td>
<td>-1.91</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Listed</td>
<td>110</td>
<td>0.55</td>
<td>0.5</td>
<td>-0.18</td>
<td>-1.99</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

With regards to the dependent variables (ROA and ROE), we note differences in their distributions when the negative returns are removed from the data. Table 4.1 shows some descriptive statistics of the full sample data where the values of skewness and excess kurtosis that indicate both ROA and ROE are negatively skewed with tails heavier than a normal distribution. On the other hand, with removal of the negative returns from the data, the values of skewness and excess kurtosis from Table 4.2 show slightly positively skewed and thinner tailed distribution relative to a normal distribution.

Based on the full sample, the minimum impairment allowance made was as low as 0.4% while the maximum was about 22% of gross loans and advances. Clearly, there were signs of deterioration of some banks loan portfolios prompting a provisioning rate as high as 22% and this perhaps culminated in a maximum impairment charge of 14% which was way above the
industry average. As reported by Boahene et al. (2012), the ratio of non-performing loans to gross loans and advances tend to be high— an indication of high default among borrowers in Ghana. This also could possibly be a reason for high lending rates that is prevalent in the industry. Bank size (log of total asset) averaged 2.95 with a standard deviation of 0.58 signifying that among the top ten banks competition in terms asset is quite keen.

### 4.2 Estimates for the Within-Between Regression Model

Results from fitting the two variants of the within-between model to the data with negative values show the within-bank effects are similar as expected. For both responses, ROA and ROE, the results show that all the internal factors are significant at least at the 90% confidence level, except CAR which is not significant in the ROE response model. In particular, the key variables of interest (i.e., IMP\_CHGR and IMP\_ALWR) which are proxies of credit risk are highly significant (at 99% confidence level) in explaining the variability in ROA and ROE.

Turning to the signs, as expected, impairment charge to gross loans and advances ratio (IMP\_CHGR) is significantly negatively related to profitability (ROA and ROE). However, across both models, loan loss provisioning rate (IMP\_ALWR) indicates a significant positive relationship with profitability. Previous studies (e.g., Boahene et al (2012), Amidu and Hinson (2006), Kithinji (2010), Ara, Bakaeva and Sun (2009)) also found this rather unexpected relationship. Since IMP\_ALWR is an indicator of risk of default, a significant positive relation between IMP\_ALWR and profitability would suggest that as a bank’s risk of customer loan default increases, the bank is able to increase its profitability. During any giving financial year, as non-performing loans graduate from one bucket to a higher bucket, provisioning rises alongside. The positive relationship may be possible if interest accruing on
such loans get paid early enough and this goes to beef up interest income and hence higher profits.

However, even if this phenomenon is plausible, it may not reflect the cumulative effect of movement of non-performing loans within various delinquency levels (buckets). The inclusion of the squared term of the impairment allowance ratio (IMP_ALWRsqr) in the ROA and ROE equations in this study helps to highlight this cumulative effect of increasing loan loss provisioning on profitability.

We see from the results that IMP_ALWRsqr is not only significant but also has a huge negative relationship with profitability (ROA and ROE) relative to the coefficient estimate for IMP_ALWR. The real importance of this term is that in the medium to long term, the cumulative effect of rising impairment allowance ratio would eventually have a telling effect on the ROA and ROE. This is due to the fact that as more and more impairment (provisioning) allowance is made for rising non-performing loans, not only do banks experience a lull in their credit creation activity as their lending ability shrinks or is constrained because loanable funds are tied up, but their risk appetite level might potentially go down as well. High NPLs can give cause to liquidity constraints and thus, as side their inability to lend and earn interest, banks may be unable to also invest in other areas to generate revenue due to liquidity constraints. It thus stands to reason why we may see a positive relationship between impairment allowance and profitability in the interim but observe a very high and negative significant effect eventually. Given the specified model, the statistically significant negative coefficient of IMP_ALWRsqr reveals curvilinearity of IMP_ALWR with profitability.

Indeed, from the model specification, we see that a partial derivative with respect to IMP_ALWR would show that the expected overall effect of a change in IMP_ALWR on profitability (ROA and ROE) is \( \beta_1 - 2 \times \beta_2 \text{IMP}_{-}\text{ALWR} \). Thus, the total expected effect is at the least less than the positive coefficient of IMP_ALWR and would at higher values of IMP_ALWR be negative.
Table 4.3: Within Effects - ROA

<table>
<thead>
<tr>
<th>Dependent variable: ROA</th>
<th>Hybrid Model</th>
<th>CRE Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMP_ALWR</td>
<td>0.996***</td>
<td>0.996***</td>
</tr>
<tr>
<td></td>
<td>(0.204)</td>
<td>(0.204)</td>
</tr>
<tr>
<td>‘I(IMP_ALWR^2)‘</td>
<td>-4.052***</td>
<td>-4.052***</td>
</tr>
<tr>
<td></td>
<td>(1.003)</td>
<td>(1.003)</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>-0.535***</td>
<td>-0.535***</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.123)</td>
</tr>
<tr>
<td>CAR</td>
<td>0.123*</td>
<td>0.123*</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.049***</td>
<td>0.049***</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>InfR</td>
<td>0.196*</td>
<td>0.196*</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>-0.205*</td>
<td>-0.205*</td>
</tr>
<tr>
<td></td>
<td>(0.110)</td>
<td>(0.110)</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>-0.516***</td>
<td>-0.516***</td>
</tr>
<tr>
<td></td>
<td>(0.151)</td>
<td>(0.151)</td>
</tr>
<tr>
<td>Listed</td>
<td>-0.043</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Origin</td>
<td>-0.014**</td>
<td>-0.014**</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.107**</td>
<td>-0.032</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.055)</td>
</tr>
</tbody>
</table>

Observations          | 120          | 120        |
Pseudo $R^2$          | 0.6          | 0.6        |
Log Likelihood        | 252.367      | 252.367    |
Akaike Inf. Crit.     | -466.735     | -466.735   |
Bayesian Inf. Crit.   | -413.772     | -413.772   |

Note: *p<0.1; **p<0.05; ***p<0.01
### Table 4.4: Between Effects from the hybrid model and the Differences from CRE Model - ROA

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>ROA</th>
<th>Between Effects</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘immean(IM_ALWR)’</td>
<td>0.822</td>
<td>0.174</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.753)</td>
<td>(0.780)</td>
<td></td>
</tr>
<tr>
<td>‘immean(I(IM_ALWR^2))’</td>
<td>-3.855</td>
<td>0.197</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.754)</td>
<td>(3.886)</td>
<td></td>
</tr>
<tr>
<td>‘immean(IM_CHGR)’</td>
<td>0.413</td>
<td>0.947</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.574)</td>
<td>(0.587)</td>
<td></td>
</tr>
<tr>
<td>‘immean(CAR)’</td>
<td>0.223</td>
<td>0.100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.210)</td>
<td>(0.221)</td>
<td></td>
</tr>
<tr>
<td>‘immean(LN_TA)’</td>
<td>0.021</td>
<td>-0.028*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>‘immean(Listed)’</td>
<td>0.022**</td>
<td>0.065**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.028)</td>
<td></td>
</tr>
</tbody>
</table>

Observations: 120
Log Likelihood: 252.367 252.367
Akaike Inf. Crit.: -466.735 -466.735
Bayesian Inf. Crit.: -413.772 -413.772

Note: *p<0.1; **p<0.05; ***p<0.01

The between effects for the hybrid model and the difference of the between and within effects from the CRE models are presented in Table 4.4 and 4.6. A significant between-effect suggests that the effect of a change in the covariate within a bank is different from the effect across banks. For the ROA response model, the only variable with a significant between-effect is Listed (which equals 1 if the bank is listed on the Ghana Stock Exchange and zero otherwise). Thus, being listed on the Ghana Stock Exchange has a pay-off in terms of reaping higher ROA. This is not the case with the ROE response model where the between-effect is not significant for being listed.
Table 4.5: Within Effects for ROE

<table>
<thead>
<tr>
<th></th>
<th>Hybrid Model</th>
<th>CRE Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMP_ALWR</td>
<td>4.903***</td>
<td>4.903***</td>
</tr>
<tr>
<td></td>
<td>(1.290)</td>
<td>(1.290)</td>
</tr>
<tr>
<td>'I(IMP_ALWR^2)'</td>
<td>-20.214***</td>
<td>-20.214***</td>
</tr>
<tr>
<td></td>
<td>(6.335)</td>
<td>(6.335)</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>-4.085***</td>
<td>-4.085***</td>
</tr>
<tr>
<td></td>
<td>(0.775)</td>
<td>(0.775)</td>
</tr>
<tr>
<td>CAR</td>
<td>0.171</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>(0.423)</td>
<td>(0.423)</td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.254***</td>
<td>0.254***</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>InfR</td>
<td>0.198</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>(0.748)</td>
<td>(0.748)</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>-1.830***</td>
<td>-1.830***</td>
</tr>
<tr>
<td></td>
<td>(0.695)</td>
<td>(0.695)</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>-2.243**</td>
<td>-2.243**</td>
</tr>
<tr>
<td></td>
<td>(0.954)</td>
<td>(0.954)</td>
</tr>
<tr>
<td>Listed</td>
<td>-0.204</td>
<td>-0.204</td>
</tr>
<tr>
<td></td>
<td>(0.167)</td>
<td>(0.167)</td>
</tr>
<tr>
<td>Origin</td>
<td>-0.055</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.966***</td>
<td>-0.491</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.333)</td>
</tr>
</tbody>
</table>

Observations: 120 120
Pseudo $R^2$: 0.56 0.56
Log Likelihood: 62.622 62.622
Akaike Inf. Crit.: -87.243 -87.243
Bayesian Inf. Crit.: -34.281 -34.281

Note: *p<0.1; **p<0.05; ***p<0.01
### Table 4.6: Between Effects from the hybrid model and the Differences from CRE Model - ROE

<table>
<thead>
<tr>
<th></th>
<th>Between Effects</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{imean(IMP_ALWR)} )</td>
<td>-3.659 (4.560)</td>
<td>-8.562* (4.739)</td>
</tr>
<tr>
<td>( \text{imean(I(IMP_ALWR^2))} )</td>
<td>11.012 (22.731)</td>
<td>31.227 (23.597)</td>
</tr>
<tr>
<td>( \text{imean(IMP_CHGR)} )</td>
<td>3.581 (3.476)</td>
<td>7.666** (3.561)</td>
</tr>
<tr>
<td>( \text{imean(CAR)} )</td>
<td>2.129* (1.273)</td>
<td>1.959 (1.341)</td>
</tr>
<tr>
<td>( \text{imean(LN_TA)} )</td>
<td>0.332*** (0.080)</td>
<td>0.079 (0.092)</td>
</tr>
<tr>
<td>( \text{imean(Listed)} )</td>
<td>0.046 (0.052)</td>
<td>0.250 (0.175)</td>
</tr>
</tbody>
</table>

Observations: 120 120  
Log Likelihood: 62.622 62.622  
Akaike Inf. Crit.: -87.243 -87.243  
Bayesian Inf. Crit.: -34.281 -34.281  

*Note:*  
*p<0.1; **p<0.05; ***p<0.01

#### 4.2.1 Residual Analysis of the Within-Between Model (WBM)

The residual diagnostics plots for the between effects models are presented in Figure 4.1 for the ROA response model and Figure 4.2 for the ROE response model. From the residual diagnostics plots, the within-between models appear to fit the data that include the negative observations quite fairly well in terms of residual variance homogeneity as there are no observed patterns in the residuals for both ROA and ROE response models. Some observations for both ROA and ROE response variables appear to be influential as they
Figure 4.1: Residual Plots of the WBM - ROA.

Figure 4.2: Residual Plots of the WBM - ROE.

have relatively large cook’s distance values. The rule of thumb has been to consider the as influential any observation for which the cook’s distance is greater than one. But given
the small sample size under consideration, cook’s distances that stick out relative to all observations may require further investigation as they distort the predictions by the model.

From the onset, since we are dealing with proportion or ratio data, residual normality may not be possible. However, the Q-Q normal plot helps to in indicating which observations may be outliers. An investigation of the deviance residuals that fall outside the simulated envelope proposed by Atkinson (1985) shows the associated observations to be mainly of the instances where some banks made losses during their initial stages of operation. Table in the appendix presents results for these within-between models after removing the potential outliers from the data. There are slight changes to the size of the coefficient estimates of the internal variables, including IMP_ALWR but they retain their respective signs as in the full data model. The residual diagnostics plots show that, for the ROA response model, all deviance residuals lie within the simulated envelop.

4.3 Estimates for the Beta Regression Model

The generalized linear models (GLMs) explored in this study are for the purpose of addressing possible overdispersion as indicated by the Q-Q plots of the within-between models. Additionally, the GLMs afford us the chance to specify what may be a more appropriate variance-stabilizing transformations link function.

Estimates of the coefficients on the internal determinants of profitability from both logit and arcsine link beta regressions are similar to what is observed under the within-between models. Impairment allowance rate and its squared term have similar significance as well as their respective signs or direction of the relationship with profitability (ROA, ROE). The same situation holds for impairment charge rate. Additionally, just as capital adequacy ratio
Table 4.7: Results: Beta Regression with Logit vs. Arcsine Link Function - ROA

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>ROA</th>
<th>ROA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Logit Link</td>
<td>Arcsine Link</td>
</tr>
<tr>
<td>IMP_ALWR</td>
<td>4.470***</td>
<td>5.526***</td>
</tr>
<tr>
<td></td>
<td>(1.068)</td>
<td>(1.340)</td>
</tr>
<tr>
<td>I(IMP_ALWR^2)</td>
<td>−18.478***</td>
<td>−22.941***</td>
</tr>
<tr>
<td></td>
<td>(5.047)</td>
<td>(6.280)</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>−4.627***</td>
<td>−5.5805***</td>
</tr>
<tr>
<td></td>
<td>(0.741)</td>
<td>(0.890)</td>
</tr>
<tr>
<td>CAR</td>
<td>1.098***</td>
<td>1.300***</td>
</tr>
<tr>
<td></td>
<td>(0.383)</td>
<td>(0.474)</td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.186***</td>
<td>0.217***</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>InfR</td>
<td>−0.552</td>
<td>−0.763</td>
</tr>
<tr>
<td></td>
<td>(0.762)</td>
<td>(0.936)</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>−1.104*</td>
<td>−1.238*</td>
</tr>
<tr>
<td></td>
<td>(0.653)</td>
<td>(0.818)</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>−0.500</td>
<td>−0.420</td>
</tr>
<tr>
<td></td>
<td>(0.859)</td>
<td>(1.089)</td>
</tr>
<tr>
<td>Origin</td>
<td>−0.175***</td>
<td>−0.215***</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>Listed</td>
<td>0.088***</td>
<td>0.106***</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Constant</td>
<td>−2.933***</td>
<td>−5.071***</td>
</tr>
<tr>
<td></td>
<td>(0.142)</td>
<td>(0.176)</td>
</tr>
</tbody>
</table>

Observations  | 110  | 110  |
R²              | 0.631 | 0.562 |
Log Likelihood  | 344.747 | 112.683 |

Note: *p<0.1; **p<0.05; ***p<0.01
### Table 4.8: Results: Beta Regression with Logit vs. Arcsine Link Function - ROE

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROE</td>
<td>ROE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Logit Link</td>
<td>Arcsine Link</td>
<td></td>
</tr>
<tr>
<td>IMP_ALWR</td>
<td>13.103***</td>
<td>23.931***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.864)</td>
<td>(6.941)</td>
<td></td>
</tr>
<tr>
<td>I(IMP_ALWR^2)</td>
<td>−59.602***</td>
<td>−106.756***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.744)</td>
<td>(32.364)</td>
<td></td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>−18.865***</td>
<td>−32.445***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.959)</td>
<td>(4.397)</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>−0.542</td>
<td>−0.452</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.377)</td>
<td>(2.469)</td>
<td></td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.643***</td>
<td>1.160***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.262)</td>
<td></td>
</tr>
<tr>
<td>InfR</td>
<td>−0.779</td>
<td>−2.838</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.714)</td>
<td>(4.866)</td>
<td></td>
</tr>
<tr>
<td>GDP_GR</td>
<td>−4.108*</td>
<td>−7.363*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.276)</td>
<td>(4.284)</td>
<td></td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>−4.517</td>
<td>−0.500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.079)</td>
<td>(5.66)</td>
<td></td>
</tr>
<tr>
<td>Origin</td>
<td>−0.567***</td>
<td>−1.089***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td>(0.210)</td>
<td></td>
</tr>
<tr>
<td>Listed</td>
<td>0.336***</td>
<td>0.685***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.113)</td>
<td>(0.213)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>−2.018***</td>
<td>−4.087***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.502)</td>
<td>(0.914)</td>
<td></td>
</tr>
</tbody>
</table>

|                | Observations        | 110            | 110            |
|                | R^2                 | 0.562          | 0.631          |
|                | Log Likelihood      | 112.683        | 344.747        |

*Note:* *p<0.1; **p<0.05; ***p<0.01
is not significant under the ROE within-between models, it is also not significant under any of the beta regression models for the ROE response variable.

4.3.1 Residual Analysis of the Beta Regression Models

The residual diagnostic plots for the ROA model using both logit and arcsine link functions are presented in Figure 4.3 and Figure 4.4, respectively. For the logit link model, while there is not observed residual variance heterogeneity, the cook’s distance for some of the observations are relatively high. But these are below 0.2. A review of the half-normal plot shows both link functions quite fairly model the data equally well with all but one deviance residual lying within the simulated envelop. Similar observation as described above for the ROA model applies to the ROE model as well.
Figure 4.3: Residual Plots of the Beta regression (logit) - ROA.
Figure 4.4: Residual Plots of the Beta regression (Arcsine) - ROA.
Figure 4.5: Residual Plots of the Beta regression (logit) - ROE.
Figure 4.6: Residual Plots of the Beta regression (Arcsine) - ROE.
4.4 Estimates for the Bayesian Beta Regression Model

Regarding the Bayesian beta regression, I first fit the model with a constant dispersion parameter assumption. I also fit the model assuming a non-constant dispersion parameter modeled as the function of all independent variables. As stated before, the Bayesian beta models explored have a logit link function for the mean and a logarithmic function for the precision. The results for the constant dispersion and non-constant dispersion models for ROA are presented in Tables 4.9 and 4.10 while the ROE counterparts are presented in Tables 4.11 and 4.12. Results for all four models are quite similar to the frequentist beta regression in terms of the sign on the mean parameters, especially the credit risk measures. From the Bayesian perspective, we are 95% certain that the posterior parameters for the credit risk measures lie within the credible intervals. This bolsters the results from the frequentist beta regression.
### Table 4.9: Constant Dispersion Bayesian Beta Regression Results - ROA

**Dependent Variable:**

<table>
<thead>
<tr>
<th>ROA</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Equal-tailed [95% Cred. Interv]</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMP_ALWR</td>
<td>1.394</td>
<td>0.518</td>
<td>0.499 2.435</td>
</tr>
<tr>
<td>IMP_ALWR_Sq</td>
<td>-3.570</td>
<td>1.726</td>
<td>-6.906 -0.897</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>-3.985</td>
<td>0.838</td>
<td>-5.597 -2.364</td>
</tr>
<tr>
<td>CAR</td>
<td>1.573</td>
<td>0.398</td>
<td>0.770 2.350</td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.222</td>
<td>0.037</td>
<td>0.150 0.293</td>
</tr>
<tr>
<td>InfR</td>
<td>0.719</td>
<td>0.357</td>
<td>0.099 1.439</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>-1.597</td>
<td>0.770</td>
<td>-2.987 -0.037</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>-2.047</td>
<td>0.388</td>
<td>-2.797 -1.304</td>
</tr>
<tr>
<td>Origin</td>
<td>-0.172</td>
<td>0.035</td>
<td>-0.239 -0.106</td>
</tr>
<tr>
<td>Listed</td>
<td>0.053</td>
<td>0.032</td>
<td>-0.012 0.112</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.874</td>
<td>0.162</td>
<td>-3.187 -2.569</td>
</tr>
</tbody>
</table>

**Gamma**

| Intercept | 6.281 | 0.144 | 5.993 6.556 |

Note. N = 110.
Table 4.10: Non-constant Dispersion Bayesian Beta Regression Results - ROA

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>ROA</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Equal-tailed</td>
<td>[95% Cred. Interv]</td>
</tr>
<tr>
<td>IMP_ALWR</td>
<td>4.901</td>
<td>0.927</td>
<td>2.954</td>
<td>6.613</td>
</tr>
<tr>
<td>IMP_ALWR_Sq</td>
<td>-21.536</td>
<td>4.027</td>
<td>-29.095</td>
<td>-13.137</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>-3.538</td>
<td>0.642</td>
<td>-4.765</td>
<td>-2.298</td>
</tr>
<tr>
<td>CAR</td>
<td>1.291</td>
<td>0.390</td>
<td>0.569</td>
<td>2.067</td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.198</td>
<td>0.041</td>
<td>0.119</td>
<td>0.280</td>
</tr>
<tr>
<td>InfR</td>
<td>-0.243</td>
<td>0.392</td>
<td>-0.956</td>
<td>0.593</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>-0.146</td>
<td>0.612</td>
<td>-1.359</td>
<td>1.020</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>-0.607</td>
<td>0.545</td>
<td>-1.670</td>
<td>0.457</td>
</tr>
<tr>
<td>Origin</td>
<td>-0.167</td>
<td>0.035</td>
<td>-0.237</td>
<td>-0.102</td>
</tr>
<tr>
<td>Listed</td>
<td>0.080</td>
<td>0.037</td>
<td>0.004</td>
<td>0.150</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.107</td>
<td>0.127</td>
<td>-3.364</td>
<td>-2.869</td>
</tr>
</tbody>
</table>

Gamma

<table>
<thead>
<tr>
<th></th>
<th>IMP_ALWR</th>
<th>IMP_ALWR_Sq</th>
<th>IMP_CHGR</th>
<th>CAR</th>
<th>LN_TA</th>
<th>InfR</th>
<th>GDP_GR</th>
<th>Ave_IntR</th>
<th>Origin</th>
<th>Listed</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0168</td>
<td>0.0129</td>
<td>2.33E-10</td>
<td>0.0233</td>
<td>0.0023</td>
<td>1.36E-07</td>
<td>0.00057</td>
<td>0.00796</td>
<td>0.00038</td>
<td>0.00305</td>
<td>6.31957</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.0000113</td>
<td>1.83E-06</td>
<td>4.13E-10</td>
<td>0.0000122</td>
<td>5.33E-06</td>
<td>2.14E-07</td>
<td>0.000021</td>
<td>8.12E-06</td>
<td>0.000015</td>
<td>2.06E-06</td>
<td>0.142527</td>
</tr>
<tr>
<td>Equal-tailed</td>
<td>0.01680</td>
<td>0.01292</td>
<td>5.66E-14</td>
<td>0.02327</td>
<td>0.00233</td>
<td>7.72E-11</td>
<td>0.00054</td>
<td>0.00794</td>
<td>0.00036</td>
<td>0.00304</td>
<td>6.02836</td>
</tr>
<tr>
<td>[95% Cred. Interv]</td>
<td>0.01684</td>
<td>0.01293</td>
<td>1.63E-09</td>
<td>0.02331</td>
<td>0.00235</td>
<td>7.52E-07</td>
<td>0.00006</td>
<td>0.00797</td>
<td>0.00041</td>
<td>0.00305</td>
<td>6.58737</td>
</tr>
</tbody>
</table>

Note. N = 110.
## Table 4.11: Constant Dispersion Bayesian Beta Regression Results - ROE

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>ROE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>IMP_ALWR</td>
<td>4.031</td>
</tr>
<tr>
<td>IMP_ALWR_sq</td>
<td>-14.937</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>-8.061</td>
</tr>
<tr>
<td>CAR</td>
<td>1.177</td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.715</td>
</tr>
<tr>
<td>InfR</td>
<td>4.279</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>0.428</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>-6.710</td>
</tr>
<tr>
<td>Origin</td>
<td>-0.597</td>
</tr>
<tr>
<td>Listed</td>
<td>0.308</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.756</td>
</tr>
</tbody>
</table>

### Gamma

| Intercept | 2.659 | 0.140 | 2.380 | 2.925 |

Note. N = 110.
### Table 4.12: Non-constant Dispersion Bayesian Beta Regression Results - ROE

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>ROE</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Equal-tailed [95% Cred. Interv]</td>
<td></td>
</tr>
<tr>
<td>IMP_ALWR</td>
<td>4.579</td>
<td>1.255</td>
<td>2.026</td>
<td>7.229</td>
</tr>
<tr>
<td>IMP_ALWR_Sq</td>
<td>-15.612</td>
<td>2.577</td>
<td>-20.585</td>
<td>-10.575</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>-16.502</td>
<td>2.267</td>
<td>-21.095</td>
<td>-12.244</td>
</tr>
<tr>
<td>CAR</td>
<td>-0.054</td>
<td>1.225</td>
<td>-2.395</td>
<td>2.440</td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.381</td>
<td>0.114</td>
<td>0.156</td>
<td>0.610</td>
</tr>
<tr>
<td>InfR</td>
<td>-3.124</td>
<td>1.993</td>
<td>-7.117</td>
<td>0.903</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>-4.104</td>
<td>2.296</td>
<td>-8.521</td>
<td>0.706</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>-0.109</td>
<td>0.734</td>
<td>-1.428</td>
<td>1.378</td>
</tr>
<tr>
<td>Origin</td>
<td>-0.625</td>
<td>0.114</td>
<td>-0.842</td>
<td>-0.383</td>
</tr>
<tr>
<td>Listed</td>
<td>0.270</td>
<td>0.114</td>
<td>0.051</td>
<td>0.500</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.190</td>
<td>0.512</td>
<td>-2.209</td>
<td>-0.243</td>
</tr>
</tbody>
</table>

**Gamma**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMP_ALWR</td>
<td>0.376</td>
<td>0.227</td>
<td>0.050</td>
<td>0.905</td>
</tr>
<tr>
<td>IMP_ALWR_Sq</td>
<td>0.271</td>
<td>0.166</td>
<td>0.042</td>
<td>0.648</td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>0.315</td>
<td>0.181</td>
<td>0.048</td>
<td>0.724</td>
</tr>
<tr>
<td>CAR</td>
<td>0.651</td>
<td>0.347</td>
<td>0.112</td>
<td>1.434</td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.403</td>
<td>0.155</td>
<td>0.126</td>
<td>0.716</td>
</tr>
<tr>
<td>InfR</td>
<td>0.174</td>
<td>0.080</td>
<td>0.038</td>
<td>0.349</td>
</tr>
<tr>
<td>GDP_GR</td>
<td>0.225</td>
<td>0.081</td>
<td>0.087</td>
<td>0.400</td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>0.226</td>
<td>0.130</td>
<td>0.037</td>
<td>0.503</td>
</tr>
<tr>
<td>Origin</td>
<td>0.093</td>
<td>0.049</td>
<td>0.011</td>
<td>0.193</td>
</tr>
<tr>
<td>Listed</td>
<td>0.304</td>
<td>0.185</td>
<td>0.051</td>
<td>0.745</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.266</td>
<td>0.472</td>
<td>0.334</td>
<td>2.135</td>
</tr>
</tbody>
</table>

Note. N = 110.
4.4.1 Diagnostic Analysis of the Bayesian Beta Regression Models

The estimates from the Bayesian Regression Models corroborates the negative relationship between actual impairment allowance charge rate and profitability and the curvilinear (eventual negative) relationship between impairment allowance rate and profitability. However, there appears to be convergence issues with the MCMC samples for the posterior distribution of the parameters. The situation is particular worse in the non-constant dispersion models with burn-in iteration values below 20,000. Though convergence improves significantly with burn-in iteration values above 200,000, one needs to exercise caution in interpreting and utilizing results from the Bayesian models as the improvements may just also be pseudo-convergence. Included in the Appendix B are samples of the diagnostic plots using burn-in iteration value of 300,000. I note that the precision estimates tend to quite mimic gamma distribution as modeled.

4.5 Model Comparison and Discussions

The within-between models (WBs) have been found to outperform traditional random effects and fixed effects estimation particularly in small samples for longitudinal data (Dieleman & Templin, 2014). The two variants of the WBs considered in this study give us the same within-fixed-effects estimates with the added benefit of identifying which between-effects are significant. A causal influence interpretation is plausible for these models except that the outcome variables analyzed are ratio or proportion data.

Shifting focus from causality to prediction, the GLM models explored are warranted given the proportion data. Both logit and arcsine link function beta regressions produce quite
similar fits to data. Not much difference is noted by visual inspection of their respective residual diagnostic plots. An additional plot of estimates against the observed ROA and ROE for both link functions as shown in Figure 4.7(a) and 4.7(b) also provide only very little on deciding which link function best models these response variables. However, using the AIC values presented in Table 4.13, we able to make a guided but cautious choice of one model over the other.

Regarding the ROA response, the logit link function model can be said to marginally outperform the arcsine link function in fitting the data, based on their AIC values. However, with respect to the ROE response variable, the arcsine link marginally outperforms the logit link function model with a slightly lower AIC value.

Finally I compare the results from the constant precision and non-constant precision specifications of the Bayesian beta regression models. The model tests and information criterion results are presented in Tables 4.14 and 4.15 for the ROA and the ROE responses, respectively. With respect to the ROA, we note from Table 4.14 that the constant variance model outperforms the non-constant variance counterpart based on the various indicators.
Figure 4.7: Comparison of fit by logit and arcsine link functions
Table 4.13: Akaike Information Criterion comparison of fit by logit and arcsine link functions

<table>
<thead>
<tr>
<th>Link Functions</th>
<th>ROA</th>
<th>ROE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logit</td>
<td>-665.494</td>
<td>-201.366</td>
</tr>
<tr>
<td>Arcsine</td>
<td>-665.0885</td>
<td>-201.576</td>
</tr>
</tbody>
</table>

Table 4.14: Comparison of constant and non-constant precision Bayesian beta models - ROA

| Model                              | DIC         | log(ML)     | log(BF)    | P(M|y) |
|------------------------------------|-------------|-------------|------------|-------|
| Constant precision (ROA)           | -670.4244   | 285.8259    | .          | 1     |
| Non-constant precision (ROA)       | -665.5486   | 249.0735    | -36.7524   | 0     |

Table 4.15: Comparison of constant and non-constant precision Bayesian beta models - ROE

| Model                              | DIC         | log(ML)     | log(BF)    | P(M|y) |
|------------------------------------|-------------|-------------|------------|-------|
| Constant precision (ROE)           | -227.9693   | 61.50585    | .          | 1     |
| Non-constant precision (ROE)       | -201.8223   | 29.49392    | -32.0119   | 0     |

The smaller DIC and larger log(ML) show the constant precision model provides a better fit. The log Bayes-factor for the non-constant variance model relative to the constant variance model is $-36.75$, indicating the former has a lower predictive score (Kass & Raftery,
1995). We observe similar situation in the ROE model where the constant variance model outperforms the constant variance marginally (see Table 4.15). It is however noted that both models show signs of weak convergence.
CHAPTER 5

CONCLUSIONS

This thesis investigates the probable effect of two measures (proxies) of credit risk on commercial banks profitability. Profitability is measured by return on asset (ROA) and return on equity (ROE).

An exploratory preliminary causal within-between model is fitted to the data with negative returns to compare results with similar estimates in the literature. Results suggest negative linear relationship between actual impairment charge rate and profitability, and a curvilinear relationship between impairment allowance rate and profitability. The curvilinear relationship is such that the slope of the effect of impairment allowance on profitability changes sign from positive to negative as the former increases.

To appropriately account for possible over dispersion in the profitability ratios (ROA, ROE) data, beta regression models are explored. The results reflect similar relationship between credit risk (impairment allowance rate, actual impairment charge rate) and profitability as seen in the standard linear within-between models. The logit link function performs better marginally relative to the arcsine link in modeling of the ROA response. The logit link function on the other hand does slightly better modeling the ROE response.

A Bayesian beta regression modeling is also explored and the constant precision model provides a better fit than the random precision model, all of which affirm the relationship estimated by the frequentist beta regression models.
REFERENCES


## APPENDIX A

### WBM RESULTS WITHOUT POTENTIAL OUTLIERS

**Table A1:** Within-Effects from the Within-Between Models after deleting potential outliers

<table>
<thead>
<tr>
<th></th>
<th>ROA</th>
<th>ROE</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMP_ALWR</td>
<td>Hybrid Model</td>
<td>CRE Model</td>
<td>Hybrid Model</td>
<td>CRE Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.435***</td>
<td>0.435***</td>
<td>3.319***</td>
<td>3.319***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.103)</td>
<td>(1.032)</td>
<td>(1.032)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{I}(\text{IMP_ALWR}^2) )</td>
<td>-1.897***</td>
<td>-1.897***</td>
<td>-14.585***</td>
<td>-14.585***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.493)</td>
<td>(0.493)</td>
<td>(4.914)</td>
<td>(4.914)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMP_CHGR</td>
<td>Hybrid Model</td>
<td>CRE Model</td>
<td>Hybrid Model</td>
<td>CRE Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.492***</td>
<td>-0.492***</td>
<td>-4.238***</td>
<td>-4.238***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.057)</td>
<td>(0.569)</td>
<td>(0.569)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>0.115***</td>
<td>0.115***</td>
<td>0.136</td>
<td>0.136</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.315)</td>
<td>(0.315)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN_TA</td>
<td>0.015***</td>
<td>0.015***</td>
<td>0.140***</td>
<td>0.140***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.039)</td>
<td>(0.039)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>InfR</td>
<td>-0.031</td>
<td>-0.031</td>
<td>0.046</td>
<td>0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.059)</td>
<td>(0.585)</td>
<td>(0.585)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP_GR</td>
<td>-0.059</td>
<td>-0.059</td>
<td>-0.942*</td>
<td>-0.942*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.052)</td>
<td>(0.515)</td>
<td>(0.515)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ave_IntR</td>
<td>-0.014</td>
<td>-0.014</td>
<td>-0.839</td>
<td>-0.839</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.077)</td>
<td>(0.767)</td>
<td>(0.767)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listed</td>
<td>-0.023*</td>
<td>-0.023*</td>
<td>-0.193</td>
<td>-0.193</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.123)</td>
<td>(0.123)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin</td>
<td>-0.006</td>
<td>-0.006</td>
<td>-0.069</td>
<td>-0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.132)</td>
<td>(0.132)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.525</td>
<td>-0.514</td>
<td>-6.807</td>
<td>-6.149</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.554)</td>
<td>(5.553)</td>
<td>(41.055)</td>
<td>(26.518)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>115</td>
<td>324.024</td>
<td>-606.048</td>
<td>-548.404</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>324.024</td>
<td>103.268</td>
<td>-106.892</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>324.024</td>
<td>-164.536</td>
<td>-164.536</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>324.024</td>
<td>-106.892</td>
<td>-106.892</td>
</tr>
</tbody>
</table>

*Note: *p<0.1; **p<0.05; ***p<0.01
Figure A1: Residual Plots of the WBM after deleting potential outliers - ROA.
Figure A2: Residual Plots of the WBM after deleting potential outliers - ROE.
APPENDIX B

SAMPLE DIAGNOSTICS PLOTS FOR BAYESIAN BETA REGRESSION

**Figure B1:** Diagnostics Plots for the Bayesian Beta regression (Constant Precision) - ROA.

**Figure B2:** Diagnostics Plots for the Bayesian Beta regression (Non-Constant Precision) - ROA
Figure B3: Diagnostics Plots for the Bayesian Beta regression (Non-Constant Precision) - ROA

Figure B4: Diagnostics Plots for the Bayesian Beta regression (Constant Precision) - ROE

Figure B5: Diagnostics Plots for the Bayesian Beta regression (Non-Constant Precision) - ROE
Figure B6: Diagnostics Plots for the Bayesian Beta regression (Non-Constant Precision) - ROE
# Codes for the Within-between (WBM) Models

```r
#install.packages('lmerTest')
#install.packages('jtools')
library(lmerTest)
library(jtools)
library(sandwich)

#install.packages('D:/Downloads/panelr_0.7.4.tar.gz', source = TRUE, repos = NULL)
#install.packages("stargazer") #outputing tables
library(stargazer)
library(panelr)

#install.packages('skimr') #Helps with summary stats
library(skimr)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("car")
library(car)

# Read data

# Set working directory
setwd("C:/Users/godwi/Dropbox/Research/Stats_paper")
```
#full data
panel1_O <- read.csv("NEWPANEL1.csv")
attach(panel1_O)
#Declare data as panel
datta1<- panel_data(panel1_O, id = ID, wave = Year)
# creating the factor variables
Origin.f<-factor(datta1$Origin)
Listed.f<-factor(datta1$Listed)
datta1<-cbind(datta1,Origin.f,Listed.f)
colnames(datta1)[19] <- "Origin.f"
colnames(datta1)[20] <- "Listed.f"
datta1<- panel_data(panel1_O, id = ID, wave = Year)

#1. Model 1: The Hybrid Within-Between Effects Model for ROA
model1 <- wbm(ROA_O ~ IMP_ALWR + I(IMP_ALWR^2) + IMP_CHGR + CAR + LN_TA + InfR + GDP_GR + Ave_IntR + Listed | Origin, data = datta1)
summary(model1)
#The summ output shows the wbm report robust-cluster standard errors
summ(model1, robust = "HC3", cluster = "ID")

#2. Model 2: Correlated Random Effects Model (Contextual Model) for ROA
model2 <- wbm(ROA_O ~ IMP_ALWR + I(IMP_ALWR^2) + IMP_CHGR + CAR + LN_TA + InfR + GDP_GR + Ave_IntR + Listed | Origin, data = datta1, model="contextual")
summary(model2)

# Output the two models into latex table for ROA

class(model1) <- "lmerMod"
class(model2) <- "lmerMod"
stargazer(model1, model2, title="Results", align=TRUE)

# 3. Model 3: The Hybrid Within-Between Effects Model for ROE

model3 <- wbm(ROE_O ~ IMP_ALWR + I(IMP_ALWR^2) + IMP_CHGR + CAR + LN_TA + InfR + GDP_GR + Ave_IntR + Listed | Origin, data = datta1)

summary(model3)

# 4. Model 4: Correlated Random Effects Model (Contextual Model) for ROE

model4 <- wbm(ROE_O ~ IMP_ALWR + I(IMP_ALWR^2) + IMP_CHGR + CAR + LN_TA + InfR + GDP_GR + Ave_IntR + Listed | Origin, data = datta1, model="contextual")

summary(model4)

# Output the two models into latex table for ROE

class(model3) <- "lmerMod"
class(model4) <- "lmerMod"
stargazer(model3, model4, title="Results", align=TRUE)

# Combine Output of the two models into latex table for ROA and ROE

stargazer(model1, model2, model3, model4, title="Results", align=TRUE)
# Model 1 Residual Diagnostics

par(mfrow = c(3, 2))

## Pearson Residuals vs index:
plot(residuals(model1), ylab = "Pearson residual", xlab = "Obs. number", main = "Pearson Residuals vs indices of obs.")

## Cook's Distance
plot(cooks.distance(model1), type = "n", ylab = "Cook's distance", xlab = "Obs. number", main = "Cook's distance plot.")
cdist <- cooks.distance(model1)
segments(1:length(cdist), 0, 1:length(cdist), cdist)

## leverage vs fitted
plot(fitted(model1), hatvalues(model1), ylab = "Leverage", xlab = "Linear predictor", main = "Leverage vs predicted values.")

## residuals vs linear predictor
plot(fitted(model1), residuals(model1), ylab = "Pearson residual", xlab = "Linear predictor", main = "Residuals vs linear predictor.")

## normal quantiles
car::qqPlot(residuals(model1), ylab = "Deviance residuals", xlab = "Normal quantiles", main = "Q-Q normal plot of residuals.", pch = 1, grid = FALSE)

## Deviance Residuals vs index:
plot(residuals(model1, type = "deviance"), ylab = "Deviance residual", xlab = "Obs. number", main = "Deviance Residuals vs indices of obs.")

# Model 2 Diagnostics

par(mfrow = c(3, 2))
## Pearson Residuals vs index:
plot(residuals(model2), ylab = "Pearson residual", xlab="Obs. number",main = "Pearson Residuals vs indices of obs.")

## Cook's Distance
plot(cooks.distance(model2), type="n", ylab = "Cook’s distance", xlab="Obs. number",main = "Cook’s distance plot.")
cdist <- cooks.distance(model2)
segments(1:length(cdist), 0, 1:length(cdist), cdist)

## leverage vs fitted
plot(fitted(model2), hatvalues(model2), ylab = "Leverage", xlab="Linear predictor",main = "Leverage vs predicted values.")

## residuals vs linear predictor
plot(fitted(model2), residuals(model2),ylab = "Pearson residual", xlab="Linear predictor",main = "Residuals vs linear predictor.")

## normal quantiles
car::qqPlot(residuals(model2),ylab = "Deviance residuals", xlab="Normal quantiles",main = "Q-Q normal plot of residuals.",pch = 1)

## Deviance Residuals vs index:
plot(residuals(model2, type ="deviance"),ylab = "Deviance residual", xlab="Obs. number",main = "Deviance Residuals vs indices of obs.")

#Model3 Residual Diagnostics
par(mfrow=c(3,2))

## Pearson Residuals vs index:
plot(residuals(model3), ylab = "Pearson residual", xlab="Obs. number",main = "Pearson Residuals vs indices of obs.")
## Cook's Distance

plot(cooks.distance(model3), type="n", ylab = "Cook's distance", xlab="Obs. number",main = "Cook's distance plot.")
cdist <- cooks.distance(model3)
segments(1:length(cdist), 0, 1:length(cdist), cdist)

## leverage vs fitted

plot(fitted(model3), hatvalues(model3), ylab = "Leverage", xlab="Linear predictor",main = "Leverage vs predicted values.")

## residuals vs linear predictor

plot(fitted(model3), residuals(model3),ylab = "Pearson residual", xlab="Linear predictor",main = "Residuals vs linear predictor.")

## normal quantiles

car::qqPlot(residuals(model3),ylab = "Deviance residuals", xlab="Normal quantiles",main = "Q-Q normal plot of residuals.",pch = 1,grid=FALSE)

## Deviance Residuals vs index:

plot(residuals(model3, type ="deviance"),ylab = "Deviance residual", xlab="Obs. number",main = "Deviance Residuals vs indices of obs.")

##### END OF ANALYSIS OF DATA WITH NEGATIVE RETURNS #####

##### BEGINNING OF ANALYSIS OF NON-NEGATIVE RETURNS DATA #####

#Read Non-negative returns data

Banksdata_Non<- read.csv("NEWPANEL1_nonneg.csv")

attach(Banksdata_Non)
#install.packages("betareg")
library(betareg)

#### Beta Regressions #######

#Logit Link Function Beta Regression

### THE ROA Model - Model 5 ###

model5=betareg(ROA ~IMP_ALWR + I(IMP_ALWR^2) + IMP_CHGR + CAR + LN_TA + InfR + GDP_GR + Ave_IntR + Origin + Listed, link = "logit", data=Banksdata_Non)
summary(model5)

#Results table
stargazer(model5, title="Results", align=TRUE)

### Diagnostics #plot5

set.seed(123)
par(mfrow=c(3,2))
plot(model5, which = 1:4, type = "pearson")
plot(model5, which = 5, type = "deviance", sub.caption = "")
plot(model5, which = 1, type = "deviance", sub.caption = "")

### THE ROE Model - Model 6 ###

model6=betareg(ROE ~IMP_ALWR + I(IMP_ALWR^2) + IMP_CHGR + CAR + LN_TA + InfR + GDP_GR
+ Ave_IntR + Origin + Listed, link = "logit", data=Banksdata_Non)

summary(model6)

#Results table

stargazer(model6, title="Results", align=TRUE)

###Diagnostics #plot6

set.seed(124)

par(mfrow=c(3,2))

plot(model6, which = 1:4, type = "pearson")

plot(model6, which = 5, type = "deviance", sub.caption = "")

plot(model6, which = 1, type = "deviance", sub.caption = "")

########################################################

### Arcsine link Beta Regression ###

########################################################

## Arcsine link Model for ROA - Model 7

y <- Banksdata_Non$ROA

n <- length(y)

X <- cbind(rep(1,n),Banksdata_Non$IMP_ALWR,Banksdata_Non$IMP_ALWR_Sq,
           Banksdata_Non$IMP_CHGR,Banksdata_Non$CAR,Banksdata_Non$LN_TA,
           Banksdata_Non$InfR,Banksdata_Non$GDP_GR, Banksdata_Non$Ave_IntR,Banksdata_Non$Origin,Banksdata_Non$Listed)

bet <- c(0,0,0,0,0,0,0,0,0,0,0)

betold <- c(0,0,0,0,0,0,0,0,0,0,0)

aphi <- 0.5 # initial a(phi)

phi <- 1/aphi-1
Xb <- X %*% betold
#mu <- sin((1/9)*Xb+asin(sqrt(0.5)))^2
mu <- (sin((1/9)*Xb+asin(sqrt(0.5))))^2
num <- 1000
for (k in (1:num)){
  W <- diag(c((2/9)*sqrt(mu*(1-mu))))
  digams <- digamma(phi*(1-mu))-digamma(phi*mu)
  logys <- log(y/(1-y))
  ys <- digams + logys
  trigams <- trigamma(phi*(1-mu))+trigamma(phi*mu)
  Ws <- diag(c(phi*((4/81)*mu*(1-mu))*trigams + (2/81)*(2*mu-1)*(
            digams+logys)))

  bet <- betold + solve(t(X) %*% Ws %*% X) %*% t(X) %*% ys # phi is canceled
  Xb <- X %*% bet
  #mu <- sin(1/9*Xb+asin(sqrt(0.5)))^2
  mu <- (sin((1/9)*Xb+asin(sqrt(0.5))))^2

  aphi <- sum(((y-mu)^2/(mu*(1-mu)))/(n-2))
  phi <- 1/aphi-1

  if (max(abs(bet-betold)) < 1e-8){
    break
  }
}
else{
    betold <- bet
}

if (k>=num){
    print("Newton-Raphson does not converge!")
}

# Estimated beta
bet

# Estimated phi
phi

# Approximated Fisher Information
digams <- digamma(phi*(1-mu))-digamma(phi*mu)
trigams <- trigamma(phi*(1-mu))+trigamma(phi*mu)
logmus <- log(phi*mu/(phi*(1-mu)-1))

# Wss <- diag(c(phi*(mu*(1-mu))^2*trigams + mu*(1-mu)*(2*mu-1)*(digams+logmus)))
Wss <- diag(c(phi*((4/81)*mu*(1-mu))*trigams + (2/81)*(2*mu-1)*(digams+logmus)))

FI <- phi*t(X)%*%Wss%*%X

# var(betahat)
Varbet <- solve(FI)
Varbet
# z-test statistic

#Calculate z values for significance

z <- bet / sqrt(diag(Varbet))

table

#P-values

p_values<-2*pnorm(-abs(z))

#Standard error

se<-sqrt(diag(Varbet))

# Assign names and report the results.

result<-cbind(bet,se,z,p_values);

# Get fitted values

logitest <- model5$fitted.values

#Calculate Pearson Residuals

pearson=(y-mu)/sqrt(mu)

# Calculate Deviance Residuals
di <- (2*y*log(y/mu) + 2*(n-y)*log((n-y)/(n-mu)))
res<- y-mu
Devresid <-sign(res)*sqrt(di)
#Plot the diagnostics plots
par(mfrow=c(2,2))
## Pearson Residuals vs index:
plot(pearson,ylab = "Pearson residual", xlab="Obs. number",main = "Pearson Residuals vs indices of obs.")
## residuals vs linear predictor
plot(asinest, pearson,ylab = "Residual", xlab="Linear predictor",main = "Residuals vs linear predictor.")
#normal quantiles
car::qqPlot(Devresid,ylab = "Deviance residuals", xlab="Normal quantiles",main = "Q-Q normal plot of residuals.",pch = 1,grid=FALSE)
#Deviance Residuals vs indices of observation
plot(Devresid,ylab = "Deviance residual", xlab="Obs. number",main = "Deviance Residuals vs indices of obs.")
#### Model Comparison Logit vs Arcsine for ROA ####
#Table of results : Logit vs Arcsine for ROA - Table#4.7
stargazer(model5,model7, title="Results: Beta Regression with Logit vs. Arcsine Link Function - ROA", align=TRUE)
#Get fitted values
logitest <- model5$fitted.values
asinest <- mu
#Comparing the two link functions
```r
plot(y, logitest, pch=1, xlab="Observations - ROA", ylab="Estimated ROA")
points(y, asinest, pch=4)
abline(0, 1, col="red")
legend(0.05, 0.108, legend=c("logit", "arcsine"), pch=c(1, 4), cex=0.8)

######## Using AIC values ##########
# AIC for model 5
AIC(model5)

# Beta shape values
estBetaParams <- function(mu, var) {
  alpha <- ((1 - mu) / var - 1 / mu) * mu ^ 2
  beta <- alpha * (1 / mu - 1)
  return(params = list(alpha = alpha, beta = beta))
}
parEst<-estBetaParams(mu, var)
alpha<-parEst$alpha
beta<-parEst$beta

# Log-likelihood
LL <- sum(dbeta(y, alpha, beta, log=TRUE))
# AIC, 1 intercept, 11 slopes and 1 dispersion
AIC.7<- -2*LL + 2*12
AIC.7

############################ ROE - Arcsine link #################################

########## Arcsine link Model for ROE - Model 8 ###########################
y <- Banksdata_Non$ROE
n <- length(y)
```
X <- cbind(rep(1,n), Banksdata_Non$IMP_ALWR, Banksdata_Non$IMP_ALWR_Sq,
        Banksdata_Non$IMP_CHGR, Banksdata_Non$CAR, Banksdata_Non$LN_TA,
        Banksdata_Non$InfR, Banksdata_Non$GDP_GR, Banksdata_Non$Ave_IntR, Banksdata_Non$Origin,
        Banksdata_Non$Listed)

bet <- c(0,0,0,0,0,0,0,0,0,0,0)
betold <- c(0,0,0,0,0,0,0,0,0,0,0)
aphi <- 0.5 # initial a(\phi)
phi <- 1/aphi-1

Xb <- X %*% betold

#mu <- sin((1/9)*Xb+asin(sqrt(0.5)))^2
mu <- (sin((1/9)*Xb+asin(sqrt(0.5))))^2

num <- 1000
for (k in (1:num)){
    W <- diag(c((2/9)*sqrt(mu*(1-mu))))
    digams <- digamma(phi*(1-mu))-digamma(phi*mu)
    logys <- log(y/(1-y))
    ys <- digams + logys
    trigams <- trigamma(phi*(1-mu))+trigamma(phi*mu)
    Ws <- diag(c(phi*((4/81)*mu*(1-mu))*trigams + (2/81)*(2*mu-1)*
               digams+logys)))

    bet <- betold + solve(t(X)%*%Ws%*%X) %*% t(X)%*%Ws%*%ys # phi is canceled
    Xb <- X %*% bet
    #mu <- sin(1/9*Xb+asin(sqrt(0.5)))^2
mu <- (sin((1/9)*Xb+asin(sqrt(0.5))))^2
aphi <- sum((y-mu)^2/(mu*(1-mu))/(n-2))
phi <- 1/aphi-1

if (max(abs(bet-betold)) < 1e-8){
  break
}
else{
  betold <- bet
}
if (k>=num){
  print("Newton-Raphson does not converge!")
}

# Estimated beta
bet

# Estimated phi
phi

# Approximated Fisher Information
digams <- digamma(phi*(1-mu))-digamma(phi*mu)
trigams <- trigamma(phi*(1-mu))+trigamma(phi*mu)
logmus <- log(phi*mu/(phi*(1-mu)-1))
# Wss <- diag(c(phi*(mu*(1-mu))^2*trigams + mu*(1-mu)*(2*mu-1)*(digams+logmus )))
Wss <- diag(c(phi*((4/81)*mu*(1-mu))*trigams + (2/81)*(2*mu-1)*(digams + logmus)))
FI <- phi*t(X)%*%Wss%*%X
FI
# var(betahat)
Varbet <- solve(FI)
Varbet
# z-test statistic
#Calculate z values for significance
z <- bet / sqrt(diag(Varbet))
z
#P-values
p_values<-2*pnorm(-abs(z))
#Standard error
se<-sqrt(diag(Varbet))
# Assign names and report the results.
result<-cbind(bet,se,z,p_values);
colnames(result)[1]<-"estimate"
colnames(result)[2]<-"SE"
colnames(result)[3]<-"z value"
colnames(result)[4]<-"Pr(>|t|)"
rownames(result)<-c("Intercept","IMP_ALWR","IMP_ALWR_Sq","IMP_CHGR","CAR","LN_TA","InfR","GDP_GR","Ave_IntR","Origin","Listed")
print(result)
## Model Diagnostics Plotting for model 8 -- plot#8

# Get fitted values

\[ \text{logitest6} \leftarrow \text{model6}\$\text{fitted.values} \]

\[ \text{asinest8} \leftarrow \text{mu} \]

# Calculate Pearson Residuals

\[ \text{pearson} = (y - \text{mu}) / \sqrt{\text{mu}} \]

# Calculate Deviance Residuals

\[ \text{di} \leftarrow (2 \times y \times \log(y/\text{mu}) + 2 \times (n-y) \times \log((n-y)/(n-\text{mu}))) \]

\[ \text{res} \leftarrow y - \text{mu} \]

\[ \text{Devresid} \leftarrow \text{sign(res)} \times \sqrt{\text{di}} \]

### Plot the diagnostics plots for model 8

\[
\text{par(mfrow=c(2,2))}
\]

## Pearson Residuals vs index:

\[
\text{plot(pearson,ylab = "Pearson residual", xlab="Obs. number",main = "Pearson Residuals vs indices of obs." )}
\]

## residuals vs linear predictor

\[
\text{plot(asinest8, pearson,ylab = "Residual", xlab="Linear predictor",main = "Residuals vs linear predictor." )}
\]

# normal quantiles

\[
\text{car::qqPlot(Devresid,ylab = "Deviance residuals", xlab="Normal quantiles",main = "Q-Q normal plot of residuals.",pch = 1, grid=FALSE)}
\]

# Deviance Residuals vs indices of observatio

\[
\text{plot(Devresid,ylab = "Deviance residual", xlab="Obs. number",main = "Deviance Residuals vs indices of obs." )}
\]
### Model Comparison Logit vs Arcsine for ROE ###

#Table of results : Logit vs Arcsine for ROA - Table#4.7

stargazer(model6,model8, title="Results: Beta Regression with Logit vs. Arcsine Link Function - ROE", align=TRUE)

#Comparing the two link function

#Get fitted values
logitest6 <- model6$fitted.values
asinest8 <- mu

plot(y,logitest6,pch=1,xlab="Observation - ROE",ylab="Estimated ROE")
points(y,asinest8,pch=4)
abline(0,1,col="blue")
legend(0.05, 0.48, legend=c("logit", "arcsine"), pch=c(1,4), cex=0.8)

######## Using AIC values ##########

#AIC for model 6
AIC(model6)

#Variance of Y
#var(Y) = V(mu)/(1+phi), where V(mu) = mu(1-mu).
var<-(mu*(1-mu))/(1+phi)

#Beta shape values
estBetaParams <- function(mu, var) {
    alpha <- ((1 - mu) / var - 1 / mu) * mu ^ 2
}
beta <- alpha * (1 / mu - 1)
return(params = list(alpha = alpha, beta = beta))

parEst<-estBetaParams(mu,var)
alpha<-parEst$alpha
beta<-parEst$beta

#Log-likelihood
LL <- sum(dbeta(y, alpha,beta,log=TRUE))

#AIC, 1 intercept, 11 slopes and 1 dispersion
AIC.8<- -2*LL + 2*12
AIC.8

###Bayesian Beta regression using R2winBUGS
install.packages("R2WinBUGS", dependencies = TRUE, repos = "https://cloud.r-project.org")
library("R2WinBUGS")
library(coda)
library(car)

#Read Non-negative returns data
Banksdata_Non<- read.csv("NEWPANEL1_nonneg.csv")
attach(Banksdata_Non)

#Prepare data for winbugs
ROA<-Banksdata_Non$ROA
IMP_ALWR<-Banksdata_Non$IMP_ALWR
IMP_ALWR_Sq<-Banksdata_Non$IMP_ALWR_Sq
IMP_CHGR<-Banksdata_Non$IMP_CHGR
CAR<-Banksdata_Non$CAR
LN_TA<-Banksdata_Non$LN_TA
InfR<-Banksdata_Non$InfR
GDP_GR<-Banksdata_Non$GDP_GR
Ave_IntR<-Banksdata_Non$Ave_IntR
Origin<-Banksdata_Non$Origin
Listed<-Banksdata_Non$Listed
N<-nrow(Banksdata_Non)
data<-list("ROA", "IMP_ALWR", "IMP_ALWR_Sq","IMP_CHGR","CAR",
"LN_TA","InfR", "GDP_GR","Ave_IntR", "Origin", "Listed", "N")

#initial values

set.seed(145)
inits <- function(){
  list("gamma" =rgamma(5, 10),"beta0"=rnorm(1), "beta1"=rnorm(1), "
  beta2"=rnorm(1),"beta3"=rnorm(1),
  "beta4"=rnorm(1),"beta5"=rnorm(1),"beta6"=rnorm(1),"beta7"=rnorm(1)
  ,"beta8"=rnorm(1),
  "beta9"=rnorm(1),"beta10"=rnorm(1))
}

#Define the parameters to monitor:
parameters<- c("beta0", "beta1", "beta2", "beta3", "beta4", "beta5", "beta6", "beta7", "beta8", "beta9", "beta10","gamma")

#Set up the model
bayesbeta.model<- function() {
  for(i in 1:N){
    ROA[i] ~ dbeta(a[i],b[i])
    a[i]<- mu[i]*gamma
    b[i]<-(1-mu[i])*gamma
  }
  #priors
  beta0~dnorm(0.1,100)
  beta1~dnorm(0.1,100)
  beta2~dnorm(0.1,100)
  beta3~dnorm(0.1,100)
  beta4~dnorm(0.1,100)
  beta5~dnorm(0.1,100)
  beta6~dnorm(0.1,100)
  beta7~dnorm(0.1,100)
  beta8~dnorm(0.1,100)
beta9~dnorm(0.1,100)
beta10~dnorm(0.1,100)
gamma~dgamma(0.1,100)
}

#Run the model
myoutput<-bugs(model.file = "bayesbeta.model",
data = data,
parameters.to.save = parameters,
inits = inits,
n.chains = 2,
n.iter = 15000,
n.burnin = 5000,
n.thin = 3,
bugs.directory="C:/Program Files (x86)/winbugs14_full_patched/WinBUGS14/",
depbug=T)

#### Bayesian Beta Regression using Stata

log using C:\Users\godwi\Dropbox\Research\Stats_paper\bayesbeta, replace
import delimited C:\Users\godwi\Dropbox\Research\Stats_paper\NEWPANEL1_nonneg.csv

**ROA Model: constant precision**
bayes, thinning(5) initial({roa:} {scale:} .03) saving(ROA\_noncons\_simdata00011) ///
prior({roa:imp\_alwr imp\_alwr\_sq imp\_chgr car ln\_ta infr gdp\_gr ave\_intr i. origin i.listed _cons},normal(0,100)) ///
burnin(300000) rseed(123) dots: betareg roa imp\_alwr imp\_alwr\_sq imp\_chgr car ln\_ta infr gdp\_gr ave\_intr i.origin i.listed

estimates store ROA\_const\_model1

**ROA Model: non-constant precision

bayes, thinning(5) prior({scale:imp\_alwr imp\_alwr\_sq imp\_chgr car ln\_ta infr gdp\_gr ave\_intr i. origin i.listed},gamma(2,2)) initial({roa:} {scale:} .03) ///
saving(ROA\_noncons\_simdata1) prior({roa:imp\_alwr imp\_alwr\_sq imp\_chgr car ln\_ta infr gdp\_gr ave\_intr i. origin i.listed _cons},normal(0,100)) burnin(300000) ///
rseed(124) dots: betareg roa imp\_alwr imp\_alwr\_sq imp\_chgr car ln\_ta infr gdp\_gr ave\_intr i.origin i.listed, ///
scale(imp\_alwr imp\_alwr\_sq imp\_chgr car ln\_ta infr gdp\_gr ave\_intr i.origin i.listed)
estimates store ROA\_nonconst\_model1

outreg2 [ROA\_const\_model1 ROA\_nonconst\_model1] using bayesbeta, excel dec(4) replace
***Compare models by comparing the actual probability associated with each of the models.

bayestest model ROA_const_model ROA_nonconst_model

***Compare models by comparing the information criterion of the models.

*DIC

bayesstats ic ROA_const_model ROA_nonconst_model

**ROE Model: constant precision

bayes, thinning(5) initial({roe:} {scale:} .01) saving(ROEcons_simdata2) ///
prior({roe:imp_alwr imp_alwr_sq imp_chgr car ln_ta infr gdp_gr ave_intr i.
    origin i.listed _cons},normal(0,100)) ///
burnin(300000) rseed(123) dots: betareg roe imp_alwr imp_alwr_sq imp_chgr
car ln_ta infr gdp_gr ave_intr i.origin i.listed

estimates store ROE_const_model2

**ROE Model: non-constant precision

bayes, thinning(5) prior({scale:imp_alwr imp_alwr_sq imp_chgr car ln_ta infr
gdp_gr ave_intr i.origin i.listed},gamma(2,2)) initial({roe:} {scale:}
  .03) ///
saving(ROEnoncons_simdata2) prior(\{roe:imp_alwr imp_alwr_sq imp_chgr car ln_ta infr gdp_gr ave_intr i.origin i.listed _cons\}, normal(0,100)) burnin (300000) ///
  rseed(125) dots: betareg roe imp_alwr imp_alwr_sq imp_chgr car ln_ta infr gdp_gr ave_intr i.origin i.listed, ///
  scale(imp_alwr imp_alwr_sq imp_chgr car ln_ta infr gdp_gr ave_intr i.origin i.listed)
estimates store ROE_nonconst_model2

***Compare models by comparing the actual probability associated with each of the models.
bayestest model ROE_const_model2 ROE_nonconst_model2

***Compare models by comparing the information criterion of the models.
*DIC
bayesstats ic ROE_const_model2 ROE_nonconst_model2

*** Sample Model Diagnostics for the imp_alwr variable

bayesgraph diagnostics \{imp_alwr\}, traceopts(lwidth(0.2) lcolor(teal))
  acopts(lag(100)) histopts(bin(100)) kdensopts(show(all))