2022

Exploring the Use of Rasch Models to Construct Measures of Firms' Profitability with Multiple Discretization Ratio-type Data

Chen Qiu

University of Kentucky, chenqiu_qpm.uky2022@outlook.com

Author ORCID Identifier:

https://orcid.org/0000-0002-2542-7794

Digital Object Identifier: https://doi.org/10.13023/etd.2022.294

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Chen Qiu, Student

Dr. Kelly D. Bradley, Major Professor

Dr. Kenneth M. Tyler, Director of Graduate Studies

Exploring the Use of Rasch Models to Construct Measures of Firms' Profitability with Multiple Discretization Ratio-type Data

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Education
at the University of Kentucky

By
Chen Qiu
Lexington, Kentucky
Co-Directors: Dr. Kelly D. Bradley, Professor of Educational Policy Studies and Evaluation
and Dr. Michael Peabody, Adjunct Assistant Professor of Educational, School, & Counseling Psychology
Lexington, Kentucky
2022

https://orcid.org/0000-0002-2542-7794

ABSTRACT OF DISSERTATION

Exploring the Use of Rasch Models to Construct Measures of Firms' Profitability with Multiple Discretization Ratio-type Data

Ratio-type data plays an important role in real-world data analysis. Mass ratios have been created for different purposes, depending on time and people's needs. Then, it is necessary to create a comprehensive score to extract information from those mass ratios when they measure the same concept from different perspectives. Therefore, this study adopts the same logic of psychometrics to systematically conduct scale development on ratio-type data under the Rasch model. However, it is first necessary to discretize the ratio-type data for use in the Rasch model. Therefore, this study also explores the effect of different data discretization methods on scale development by using financial profitability ratios as a demonstration. Results show that retaining more ratio categories can benefit Rasch modeling because it can better inform the model. The dynamic clustering algorithm, k-median is a better method for extracting characteristic patterns of the ratio-type data and preparing the data for the Rasch model. This study illustrates that there is no one-way good discretization method for ratio-type data under the Rasch model. It is more reasonable to use the traditional algorithm if each ratio has target benchmark/benchmarks, whereas the k-median clustering algorithm achieves good modeling results when benchmark information is lacking.

KEYWORDS: Rasch, Data Discretization, K-median, Financial Profitability Ratio

07/29/2022

Chen Qiu

07/29/2022

Date

Exploring the Use of Rasch Models to Construct Measures of Firms' Profitability with Multiple Discretization Ratio-type Data

By
Chen Qiu

Kelly D. Bradley
Co-Director of Dissertation

Michael Peabody
Co-Director of Dissertation

Kenneth M. Tyler
Director of Graduate Studies

07/29/2022
Date

DEDICATION

To all my family members.

ACKNOWLEDGMENTS

For the dissertation, I want to thank my Dissertation Co-Chairs, Dr. Kelly D. Bradley, and Dr. Michael Peabody. Next, I want to thank the complete Dissertation Committee, and outside reader, respectively: Dr. Joseph Waddington, Dr. Margaret Bausch, and Dr. Shannon Sampson. Each individual provided insights that guided and challenged my thinking, substantially improving the finished product. I also want to thank Dr. Carol Hanley who provided me with detailed assistance with my writing.

For my doctoral academic life, I want to thank the consistent support and companionship from all my family members, especially my mother Liping Pan, my father Ruigang Qiu, and my girlfriend Zhimi Kuang.

In the QPM program, I would like to say thanks to Dr. Xin Ma who brought me into the QPM program in 2017; to Dr. Gregory Heileman who directly and decisively provided me with the funding to help me start my program in 2017; to Dr. Michael Toland who provided me the great mentorship and friendship; to all the APS lab members, especially Carol, David, Hao, and John Eric; to Dr. Michael Peabody who provided me the mentorship in 2019 at ABFM.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 Introduction

To facilitate comparisons, people have created various ratios in the practice of business and daily life based on their experiences and needs. These ratios are usually calculated using two numbers, one number is treated as the numerator, and the other number as the denominator. These ratio-type data are a convenient way of expressing ideas. For example, we cannot simply use weight as an indicator of people's fitness because their heights are different. Therefore, the Body Mass Index (BMI) was created, using weight divided by the square of height, and usually multiplied by 10,000 (CDC, 2022) for comparison.

During the analysis of ratio-type data, people usually compare the single ratio with the specific benchmark/benchmarks to determine whether the subject meets the targeted criterion/criteria. When comparing the single ratio to its benchmark/benchmarks, people transform the ratio-type data from a continuous variable into a categorical variable based on the benchmark/benchmarks, which indicates whether the subject's performance meets the criteria/criterion. This process of transferring data from a continuous to a categorical format is called data discretization and is a logical, common way of comparative thinking. For example, people can be classified into four weight categories based on their BMI and the benchmarks: underweight, healthy weight, overweight, and obesity (CDC, 2022). That is, in some real-world situations, we are more concerned with the categories in which values fall, rather than the differences between each value.

We can introduce a common term from the field of measurement to describe this logical process, namely grading. In this case, the grading process occurs at the item level, and a grading system can either be dichotomous or polytomous. The dichotomous grading system allows participants to receive full or zero credit for the item, whereas, the polytomous grading system allows participants to earn partial credit.

Sometimes, there is little information on benchmark/benchmarks or particular targeted criterion/criteria for the ratio-type data. In such cases, we shift attention to the data itself, meaning that we let the data introduce itself to us, such as its characteristic pattern under clustering algorithms. In fact, these techniques, such as simplifying algorithms or preprocessing data for classification models which only handle categorical data, are widely used in data science fields.

The description thus far primarily applies to a single ratio analysis. However, more ratios are constantly being created, and many describe the same latent construct or similar characteristics from different perspectives. For example, profitability ratios include gross profit margin, operating profit margin, pretax margin, net profit margin, operating ROA, ROA, return on total capital, ROE, and return on common equity (CFA, 2020, p. 274). In many cases, a researcher must unsystematically compare multiple ratio data from the mass available ratios because he or she doesn't want to lose perspective or ignore option by simply using a single ratio instead of considering all others. Comparing ratios individually can be tedious, complex, and confusing, especially when some ratios are above the benchmarks and others are below. Therefore, it is necessary to explore how these ratios can be converted into a useful comprehensive score. One of the benefits of this process is a comprehensive result from multiple perspectives in less time with less complexity.

In summary, ratio-type data analysis reflects two characteristics: (1) the logic of data discretization is consistent with how people process information through ratio-type data comparisons during decision-making, and (2) many of the ratios conceptually or practically measure the same latent variable. The first characteristic ensures that it has practical meaning to use the discretization methods to discretize data into categories, and the second characteristic ensures that the assumption of unidimensional is tenable for most models when they are designed to analyze unidimensional data.

It is not hard to see that the scenario discussed above corresponds to an examination. In the case of an exam, the examiner designs questions or items to test the ability of the examinees, such as algebraic calculations. In the exam, each item has its own answer key. In this case of extracted information from the mass of ratio-type data under the examination framework, the ratios can be treated as items, and benchmarks can be treated as answer keys. In order to obtain accurate results on subjects' abilities, psychometricians usually use advanced psychometric models to analyze the measurement data.

Rasch models have been widely incorporated into the field of education and medicine, including examinations and survey-based studies. This study adopts the same analytic philosophy of educational testing with Rasch models by using financial profitability ratio data to create a measurement of financial profitability as a practical example for demonstration purposes. The financial profitability performance will be set as the content of the examination; firms will be treated as examinees; financial profitability ratios will be treated as exam items; and finally, benchmarks will be treated as the answer keys or ranking categories for each financial ratio item. The benchmarks can be determined a priori or post hoc. For example, prior benchmarks can be determined based on commonly used benchmarks, such as the industry standards, whereas post hoc benchmarks can be determined based on clustering patterns, such as the clustering algorithms.

Therefore, this study explores how Rasch models can be used as tools to extract comprehensive information from the mass of ratio-type data under different kinds of data discretization methods. In addition, data discretization methods include the data transformation based on the target benchmarks and also include pattern extraction from data by clustering algorithm in the absence of the specific benchmarks. The advantage of this approach is that using the Rasch model embeds the decision-making strategies by the data discretization algorithm according to the empirical benchmarks or data-driven supported patterns. Namely, this combination of data discretization and psychometric modeling method is in line with the process of human thinking and decision-making.

1.2 The Goal of the Study

The goal of this study is to investigate how the combination of data discretization algorithms and Rasch models can be used as a method to extract information from ratio-type data by using the financial profitability ratios as a demonstration. The focus of this study will be twofold: (1) to systematically explore how Rasch models can be adopted as tools to effectively perform information extraction on countless financial profitability ratios according to different data discretization methods, respectively, and (2) to systematically explore how different data discretization methods will affect the performance of Rasch models. Therefore, this study will use the prescriptive psychometric approach, regarded by Rasch models, to analyze the financial profitability of ratio-type data under different data discretization methods as a demonstration of scale development. This study uses two Rasch models: dichotomous Rasch model and Partial Credit model.

Therefore, two research questions will be explored:

1. How are the items/financial profitability ratios constructed as a measure under the Rasch framework based on different discrete ratio-type data, respectively?

2. How do the different data discretization methods affect scale development?

1.3 The Design of the Study

The process of scale development starts with identifying the construct, creating items, and implementing the measurement models. The example of creating a scale to measure firms' financial profitability based on five financial profitability ratios is displaced in Figure 1.1. The first step of identifying the construct starts with financial statements based on theoretical or practical needs. Based on this construct, we start to create items. Firstly, we select several financial profitability variables from financial statements. Secondly, we convert these profitability variables into profitability ratios to eliminate the effect of different firm sizes when making comparisons. Thirdly, we use data discretization methods to create response strings for each profitability ratio, respectively, and the profitability ratios will be transformed into profitability items. In the next step, the items are inserted into a preliminary measurement model, such as the dichotomous Rasch model or Partial Credit model, to develop a scale and conduct a final model based on several criteria. Finally, we can obtain the comprehensive score (theta) from the final model to measure firms' "ability of profitability," which is comparable to any ability measured by the exam.

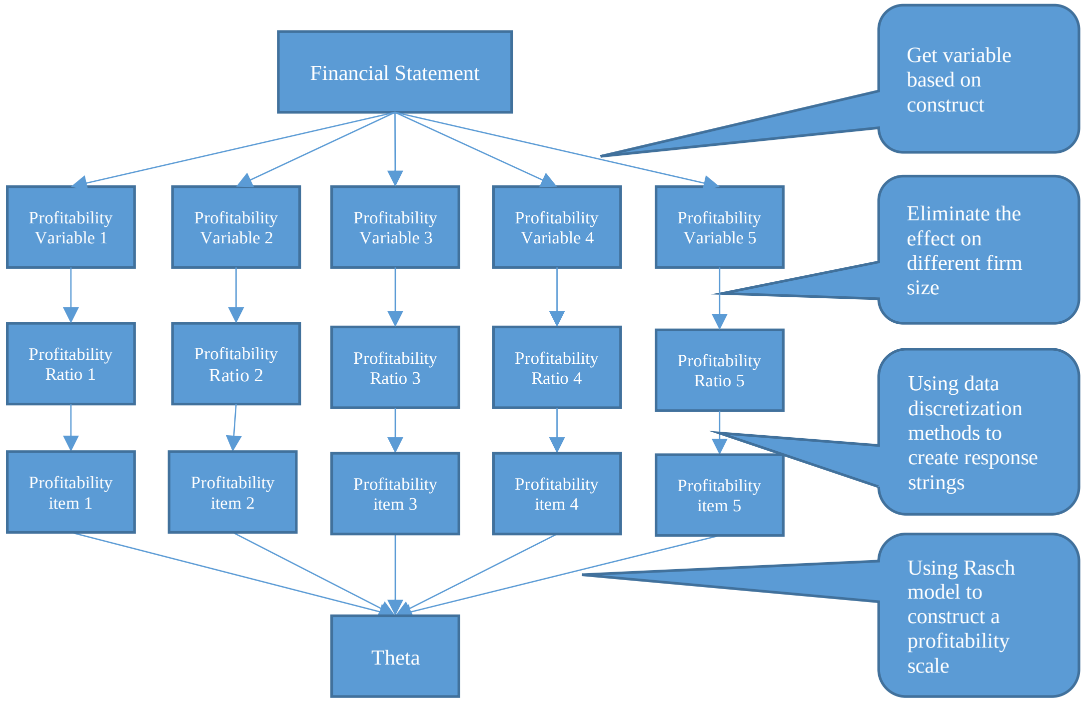

Figure 1.1 The Flow Chart of Information Extraction

1.4 Significance of the Study

The primary purpose of this study is to show how different discretization methods and measurement models can be used in the scale development of ratio-type data. Measurement models can be used as a feasible and standardized mathematical integrated solution to analyze discrete ratio-type data. This study extracts firms' financial performance of profitability from financial ratios to provide immediate information to stakeholders to retain all-around and multi-angle ratio comparison information. Meanwhile, the psychometric structure allows content experts to create new items and select data discretization methods, then incorporate Rasch models based on future needs.

This study makes two major contributions to the literature: (1) this study uses Rasch modeling to systematically create a scale for measuring firms' profitability based on the commonly used financial profitability ratio data in Wharton Research Data Services (WRDS) databases, and (2) this study systematically explores how to prepare discretized data and how different data discretization methods affect scale development. In addition to using traditional median, quartiles, and deciles to discrete the ratio-type data, this study also employed the k-median clustering algorithm to discrete the ratio-type data. Therefore, this study will compare and analyze the performance of different Rasch models under different discretization methods by using financial profitability ratios as an example.

This study contributes two critical pieces of information to the applied field. The first piece of information is the theta score which is the financial profitability indices in this study. The theta score is defined as the score of a firm's financial profitability ability, and each firm receives a theta score measuring their financial profitability performance in their respective industry. The second piece of information is the beta score, which is defined as the score of financial profitability ratio difficulty, which indicates how hard or easy for firms to achieve each financial profitability ratio.

Finally, this study can be treated as a guide for practitioners who are interested in extracting information from ratio-type data within the Rasch framework. The potential audience of this study is very broad. Researchers or practitioners can select an appropriate Rasch model to extract a latent score according to each ratio-type data with particular benchmarks or data discretization methods. Moreover, researchers can involve non-ratio-type quantitative or qualitative data with particular benchmarks in the Rasch model if all items are treated as unidimensionality. Related to this study, firms' board members can use the profitability indices produced in the study as a guide to understand their firm’s industry-standard position, then adjust the firm's operating strategy.

CHAPTER 2. LITERATURE REVIEW

2.1 Significance of Financial Ratios

In the field of business, financial statements provide useful information related to the financial and operational status of a single firm. A wealth of information about public firms can be found in these financial statements (e.g., the annual reports on Form 10-K; the quarterly reports on Form 10-Q) (SEC, 2021). The balance sheet as a static blueprint describes the assets, liabilities, and equity of the firm at a fixed time point (e.g., fiscal year, or calendar year), and includes three main components: assets (e.g., Current assets, and Fixed assets), liabilities (e.g., Current liabilities, and long-term liabilities), and stockholders' equity (e.g., Common stock, Preferred stock, and accumulated Retained earnings), which are presented in formula 2.1(Ross et al., 2018, pp. 119-120). The income statement records the firm's financial performance within a fiscal or calendar year between two adjacent years' balance sheets, and includes revenue, cost of goods sold, depreciation, interest, and taxes (Ross et al., 2018, pp. 127-128). The statement of cash flow shows the cash flow changes (i.e., operating activities, financing activities, and investing activities) during a firm's financial year (Ross et al., 2018, p.171). In practice, people can obtain different kinds of information from those financial statements to make objective and reasonable decisions based on their needs.

$$\text{Assets} \equiv \text{Liabilities} + \text{Stockholders' Equity} \tag{2.1}$$

However, in practice, it is time-consuming and cumbersome to compare firms by directly obtaining information from financial statements, especially for firms of different sizes. Fortunately, financial ratios provide us with a feasible and efficient way to compare firms of different sizes (Ross et al., 2018, p. 179). Namely, financial ratios are an effective method to extract information from financial statements. In fact, financial ratios have been used in business for over a century and can be traced back to the late 19th century (Horrigan, 1968). Over time, several analysis methods using financial ratios have been employed in industrial applications, including common-size analysis (i.e., common-size analysis of the balance sheet, common-size analysis of the income statement, crosse-sectional analysis, and trend analysis) and regression analysis (CFA Institute, 2020, pp. 248–256).

Financial ratios play an important role as the epitome of financial statements to provide a series of important information about a firm's performance in the daily work of finance related. Unfortunately, the comparison work among multiple firms can become extremely cumbersome and extraordinarily complex with countless financial ratios, especially when a firm is above others in some ratios, but below in other ratios. Meanwhile, many financial ratios measure similar content or the same dimension. Therefore, it is meaningful to utilize a method that effectively extracts and combines similar financial ratios into a single latent trait indicator. To achieve this goal, this study focuses on one of the significant financial ratio categories, financial profitability, for demonstration purposes.

2.2 Previous Related Studies

The Rasch model is designed for scale development, scale validation, and test score reporting which is popular in educational, psychological, and clinical research. However, the application of the Rasch model is still in its early stage in the financial world (Gori & Gori, 2018). For example, Ridzak (2011) first introduced to use of the logic of the examination to rank the banks based on their strictness under the Rasch model by treating banks as examiners and the firms as examinees.

Few researchers have conducted financial ratio analyses based on the Rasch model. Schellhorn and Sharma (2013) first adopted the dichotomous Rasch measurement model to rank firms by managerial ability. A similar methodological approach was performed in Jambulingam et al.'s (2016) paper. Gori and Gori (2018) used the Rasch model based on stacked data. The common feature of these studies is that they put selected financial ratios into a unidimensional structure within the Rasch framework. However, many details still need to be addressed, such as exploring the optimal threshold distinguishing method (Schellhorn & Sharma 2013), which is one of the major research goals and contributions of this study.

In detail, Schellhorn & Sharma (2013) used the dichotomous Rasch model and industry averages as benchmarks to rank the firms in the aerospace/defense industry and foods industry by managerial ability based on 13 financial ratios, resulting in 8 and 9 financial ratio models, respectively. Simultaneously, this approach was performed 10 times each year from 2002 to 2011 within each industry. The sample size (number of firms) of the aerospace/defense industry was between 17 to 23 from 2002 to 2011, and the sample size of the food industry was between 31 to 45 from 2002 to 2011. Jambulingam et al. (2016) used the dichotomous Rasch model and industry averages as benchmarks to rank the firm in the pharmaceutical industry by financial performance based on 24 market and accounting ratios, resulting in an 18 financial ratios model. Meanwhile, the same approach was performed 12 times from 2002 to 2013 with a sample size of 15 firms. Gori & Gori, (2018) used the Rating Scale Rasch model and percentiles as benchmarks to conduct a new method for credit ratings in the sector of Consumer Discretionary with a sample of 44 firms from 2004 to 2014. In addition, Gori & Gori, (2018) included 13 financial ratios at the beginning, resulting in a 10 financial ratios model.

In order to meet the requirements of the Rasch model data format, the loss of information is inevitable when transforming from continuous to categorical data (Jambulingam et al., 2016; Schellhorn & Sharma, 2013). Therefore, it is necessary to ensure that the logic of data discretization is consistent with how people process information through ratio-type data comparisons during decision-making, thereby ensuring that data discretization has practical meaning. Assuredly, keeping more categories can reduce the loss of information. Meanwhile, the firm (person) reliability, item reliability, and Cronbach alpha increase as the number of categories on each item increases (Gori and Gori, 2018). However, the thresholds can become disordered or some thresholds are too close to each other as the number of categories increases (Gori and Gori, 2018). Furthermore, some ratio-type data may not have clear benchmarks as a reference in practice (Schellhorn & Sharma, 2013). Therefore, in this case, the unsupervised machine learning clustering algorithms will better extract clustering information to ensure minimizing within-group differences and maximizing between-groups differences.

2.3 Financial Profitability Ratios

A profitability ratio measures the firm's ability to make profits. In other words, it measures the efficiency of the firm's use of assets and the firm's management and operations (Ross et. al., 2018, p. 50). Meanwhile, "Profitability reflects the firm's competitive position in the market, and by extension, the quality of its management" (CFA Institute, 2020, p. 273). In the CFA Institute's classification scheme (2020), profitability ratios include gross profit margin, operating profit margin, pretax margin, net profit margin, operating ROA, ROA, return on total capital, ROE, and return on common equity (p. 274). Under Koh and Killough's (1990) financial ratios framework, profitability ratios include net profit margin, return on investment, return on equity, basic earning power, and retained earnings to total assets.

Different researchers or stakeholders have different preferences and purposes for financial ratio selection. Meanwhile, it is worth noting, that there is no uniform industry standard for the naming and calculation formula of financial ratios, and this lack of standard is reflected by different analysts and by different databases (CFA, 2020, p. 244). Therefore, this study chose the common financial profitability ratios from the WRDS database, a well-known trusted commonly used database.

Several criteria guided the selection of financial profitability ratios in this study:

(1) All the profitability ratios were selected from the WRDS database to ensure the standardized calculation method of financial ratios.

(2) The profitability ratios are highly related to the financial profitability category by definition to achieve a single latent dimension structure.

(3) The profitability ratios are monotonic or nearly monotonic to meet the assumption of the Rasch model.

(4) All the profitability ratios in the database were selected to enhance the measurement coverage more comprehensively.

Using the criteria outlined above, the 15 commonly used financial profitability ratios selected from the WRDS database to create the financial profitability scale are: Effective Tax Rate, Gross Profit/Total Assets, After-tax Return on Average Common Equity, After-tax Return on Total Stockholders' Equity, After-tax Return on Invested

Capital, Gross Profit Margin, Net Profit Margin, Operating Profit Margin After Depreciation, Operating Profit Margin Before Depreciation, Pre-tax Return on Total Earning Assets, Pre-tax return on Net Operating Assets, Pre-tax Profit Margin, Return on Assets, Return on Capital Employed, and Return on Equity (WRDS Research Team, 2016).

Additional information on profitability ratios, the effective tax rate measures a firm's average income tax rate. The lower effective tax rate enables a firm to retain more earnings after tax. Based on the WRDS database, the formula for calculating effective tax rate is presented below:

$$\text{Effective Tax Rate} = \frac{\text{Income Tax}}{\text{Pretax Income}} \tag{2.2}$$

The gross profit to total assets ratio compares the gross profit to total assets. A higher gross profit to total assets ratio indicates the higher profit per unit of total assets. Based on the WRDS database, the formula for calculating gross profit to total assets ratio is presented below:

$$\text{Gross Profit/Total Assets} = \frac{\text{Gross Profit}}{\text{Total Assets}} \tag{2.3}$$

The after-tax return on average common equity ratio measures the net income earned per unit of the average of common equity. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating after-tax return on average common equity ratio is presented below:

$$\text{After} - \text{tax Return on Average Common Equity} = \frac{\text{Net Income}}{\text{Average of Common Equity}} \tag{2.4}$$

The after-tax return on total stockholders' equity ratio measures the net income earned per unit of total shareholders' equity. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating after-tax return on total stockholders' equity ratio is presented below:

$$\text{After} - \text{tax Return on Total Stockholders' Equity} = \frac{\text{Net Income}}{\text{Total Shareholders}'\text{Equity}} \tag{2.5}$$

The after-tax return on invested capital ratio measures a combination of net income earned and interest expenses per unit of invested capital. The higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating after-tax return on invested capital ratio is presented below:

$$\text{After} - \text{tax Return on Invested Capital} = \frac{\text{Net Income} + \text{Interest Expenses}}{\text{Invested Capital}} \tag{2.6}$$

The higher gross profit margin shows the comprehensive product situation of higher pricing and lower cost (CFA Institute, 2020, p. 274). Based on the WRDS database, the formula for calculating gross profit margin is presented below:

$$\text{Gross Profit Margin} = \frac{\text{Gross Profit}}{\text{Sales}} \tag{2.7}$$

The net profit margin measures the net income earned per unit of sale. A higher value indicates the higher profit earned per unit of sales. Based on the WRDS database, the formula for calculating net profit margin is presented below:

$$\text{Net Profit Margin} = \frac{\text{Net Income}}{\text{Sales}} \tag{2.8}$$

The operating profit margin after depreciation ratio measures the operating income after depreciation earned per unit of sales. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating operating profit margin after depreciation ratio is presented below:

$$\text{Operating Profit Margin After Depreciation} = \frac{\text{Operating Income after Depreciation}}{\text{Sales}} \tag{2.9}$$

The operating profit margin before depreciation ratio measures the operating profit margin before depreciation earned per unit of sales. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating operating profit margin before depreciation ratio is presented below:

$$\text{Operating Profit Margin Before Depreciation} = \frac{\text{Operating Income before Depreciation}}{\text{Sales}} \tag{2.10}$$

The pre-tax return on total earning assets ratio measures the operating income after depreciation per unit of total earning assets. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating pre-tax return on total earning assets ratio is presented below:

$$\text{Pre} - \text{tax Return on Total Earning Assets} = \frac{\text{Operating Income after Depreciation}}{\text{Total Earning Assets}} \tag{2.11}$$

The pre-tax return on net operating assets ratio measures the operating income after depreciation per unit of net operating assets. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating pre-tax return on net operating assets ratio is presented below:

$$\text{Pre} - \text{tax return on Net Operating Assets} = \frac{\text{Operating Income after Depreciation}}{\text{Net Operating Assets}} \qquad (2.12)$$

The pre-tax profit margin measures the pretax income per unit of sales. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating pre-tax profit margin is presented below:

$$\text{Pre} - \text{tax Profit Margin} = \frac{\text{Pretax Income}}{\text{Sales}} \qquad (2.13)$$

The return on assets ratio measures the operating income before depreciation per unit of average total assets. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating return on asset ratio is presented below:

$$\text{Return on Assets} = \frac{\text{Operating Income Before Depreciation}}{\text{Average Total Assets}} \qquad (2.14)$$

The return on capital employed ratio measures the earnings before interest and taxes per unit of interest expense. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating return on capital employed ratio is presented below:

$$\text{Return on Capital Employed} = \frac{\text{EBIT}}{\text{Interest Expense}} \qquad (2.15)$$

The return on equity ratio measures the net income earned per unit of average book equity. A higher value indicates the firm generalized higher profit during the period. Based on the WRDS database, the formula for calculating return on equity ratio is presented below:

$$\text{Return on Equity } = \frac{\text{Net Income}}{\text{Average Book Equity}} \tag{2.16}$$

2.4 Category of Industry

Because firms operate differently among industries, it is necessary to compare firms in the same industry category. Firms in different industries usually have different benchmarks for each ratio. For example, the financial industry doesn't need to have higher inventory than the manufactory. The Global Industry Classification Standard (GICS) is an industry classification structure that was created in 1999 and is constantly updated by MSCI and the S&P Dow Jones Indices since 1999 (MSCI, 2020). The GICS includes 11 sectors, 24 industry groups, 69 industries, and 158 sub-industries (MSCI, 2020). The 11 sectors include Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Communication Services, Utilities, and Real Estate (MSCI, 2020). This study uses the industrial sector data with a GICS code of 20.

2.5 Discretization of Financial Ratios

Data discretization is a significant topic in the data analysis process, especially in feature engineering under machine learning. If there are existing targeted benchmarks related to each item or variable, we can use these benchmarks to discretize the data. An example of a discretization method is the use of industry medians but other commonly used data discretization methods include quartiles and deciles. Last, some clustering algorithms can help show patterns in the data, which is a form of unsupervised machine learning. The target of unsupervised machine learning is to recognize patterns based on input vectors and without corresponding values (Bishop, 2006).

This study involves the use of four types of technical data discretization methods: median-based data discretization method; quartiles-based data discretization method; deciles-based data discretization method; and k-median clustering-based data discretization method.

2.5.1 Median-based data discretization method

Financial ratios are commonly compared to their industry medians within the same industry to determine the performance measured by the financial ratios. Therefore, it is meaningful to use industrial medians as the common benchmarks in this study. In this study, the industry median at the sector level has been used as the “answer key.” The GICS-20 industry medians of each financial ratio are downloaded from the WRDS database, and the data is dichotomized into a pass or fail format based on the “answer key.” For instance, if a firm’s financial ratio is equal to or better than the industry median, it will receive a “pass” score on this question (financial ratio) or 1, otherwise, the firm will receive a “fail” score or 0.

2.5.2 Quartiles-based data discretization method

Conceptually, the quartiles-based data discretization method can be seen as an extension of the median-based data transformation method which involves two extra thresholds of the 25th percentile and 75th percentile. However, it is worth noting that this sample median and industry median would be different if the sample did not include all firms in the same sector. Data is transformed into polytomous data based on the "answer keys." Therefore, if a firm's financial ratio is equal to or better than the 75^{th} percentile of the sample, it will receive a score of 4 on this question (financial ratio); if a firm's financial ratio is equal to or better than the 50^{th} percentile of the sample and lower than 75^{th} percentile of the sample, they will receive a score of 3 on this question (financial ratio); if a firm's financial ratio is equal or better than the 25^{th} percentile of the sample and lower than 50^{th} percentile of the sample, they will receive a score of 2 on this question (financial ratio); and if a firm's financial ratio is lower than 25^{th} percentile of the sample, they will receive a score of 1 on this question (financial ratio).

2.5.3 Deciles-based data discretization method

The logic of deciles-based data discretization methods is the same as the logic of quartiles-based data discretization method, which separates the data into 10 categories. The data is transferred into polytomous data based on the "answer keys." Therefore, if a firm's financial ratio is equal to or better than the 90^{th} percentile of the sample, it will receive a score of 10 on this question (financial ratio); if a firm's financial ratio is equal to or better than the 80^{th} percentile of the sample and lower than 90^{th} percentile of the sample, they will receive a score of 9 on this question (financial ratio); if a firm's financial ratio is equal or better than the 70^{th} percentile of the sample and lower than 80^{th} percentile of the sample, they will receive a score of 8 on this question (financial ratio) and so on until a firm's financial ratio is lower than the 10^{th} percentile of the sample, and they will receive a score of 1 on this question (financial ratio).

2.5.4 K-median clustering-based data discretization method

The k-median clustering-based data discretization method is different than the traditional approach of those previously mentioned. The k-median clustering method puts more emphasis on the data distribution in the samples of the sample clustering. The reason for choosing the k-median algorithm instead of the k-means algorithm is that the k-means algorithm is sensitive to extreme values, which cannot be avoided in most economic data. This study adopts Wang and Song's (2011) optimal k-median clustering method for unidimensional data. The optimal number of clusters will be determined by the algorithm with the Bayesian information criterion (BIC) (Song & Zhong, 2020). Song and Zhong (2020) suggested that the maximized BIC indicates the optimal number of the clustering under their algorithm. Therefore, different financial profitability ratios can obtain various numbers of clustering based on the internal structure of the data.

2.6 Psychometric Models

Measure development progresses under a cycle: creating a construct map, item development, collecting item scores, using a measurement model to analyze item scores, and reflecting back to a construct map (Wilson, 2005, p. 18-19). Psychometric models have been created to analyze item scores.

Before introducing modern advanced psychometric models, such as Rasch models, it is helpful to review a widely used classic psychometric model, Classical Test Theory (CTT). CTT is a traditional psychometric method that includes two parts: the theoretical constant true score and the random error (Bandalos, 2018, p. 156; de Ayala, 2009, p. 6;

Raykov & Marcoulides, 2011, p. 117), which is demonstrated in formula 2.17, where X is the observed score, T is the true score, and E is the error.

$$X = T + E \tag{2.17}$$

However, “An important limitation of CTT is that it does not place routinely in the center of its concerns how individuals at different levels of the construct studied (ability, trait, attribute) perform on the components, or items, of an instrument aimed at measuring that underlying latent dimension” (Raykov & Marcoulides, 2010, p. 247).

Rasch model builds a relationship between a person's latent trait and each item based on a mathematic model. Shaw (1991) described the Rasch model as a prescriptive probabilistic measurement model, which is estimated based on the person's raw score and item-total raw scores. This means that the measures in the Rasch framework are “item-free (item-distribution-free)” and “person-free (person-distribution-free)” (Linacre & Wright, 1993, p. 34; Stemler & Naples, 2021; Wright & Stone, 1979).

The Rasch model focuses more on the measurement side instead of the algorithms, which is a rigorous approach to creating measures by constraining the model algorithms (Mead, 2008). Thus, the Rasch model creates a standardized ruler (measurement) to measure a latent construct. Sick (2010) summarized the assumptions of the Rasch model as “unidimensionality,” “equal item discrimination,” and “no error due to guessing.” Furthermore, the Rasch model requires data to fit the model (Linacre, 2005). Finally, Mead (2008) mentioned that it is necessary to focus on theory-driven with a solid substantive theory and data-verification when using the Rasch model to create measures.

2.6.1 Dichotomous Rasch Models

Georg Rasch (1960) first brought the Rasch model to researchers, and the model was then promoted by Benjamin D. Wright (as cited in Linacre, 2005). The Rasch Model allows the item difficulty variation or item location variation and restricts the discrimination parameter α to be constant at 1. This makes the Rasch model with the fewest components: ability parameter for each person and difficulty parameter for each item (Wright, 1977). Formula 2.18 shows the mathematical definition of the dichotomous Rasch model. The term $x_j = 0$ represents a person giving a wrong answer on item j, whereas $x_j = 1$ represents a person giving the right answer on item j. The θ represents person ability or person location with a theoretical range from negative infinity to positive infinity. The δ_j represents item j's difficulty or item j's location with a theoretical range from negative infinity to positive infinity. The P $(x_j=1|\theta,\delta_j)$ represents the probability of person *n* giving the right answer on item j.

$$P(x_j|\theta,\delta_j) = \frac{e^{(\theta-\delta_j)}}{1+e^{(\theta-\delta_j)}} \tag{2.18}$$

2.6.2 Polytomous Psychometric Models

Master's (1982) Partial Credit (PC) model is a Rasch model approach for analyzing ordered polytomous data (de Ayala, 2019, p. 163; Hays et al., 2000), which can be either a Likert-type scale (e.g., Strongly Disagree, Disagree, Agree, and Strongly Agree) or a grading system (e.g., 0 credit, 1, credit, 2 credit, and 3 credit). Formula 2.19 shows the mathematical definition of the PC model. The θ represents person ability or person location with a theoretical range from negative infinity to positive infinity. The δ_{jh} represents the item *j*'s difficulty or item *j*'s location with a theoretical range from negative infinity to positive infinity. The P (x_j=1|θ,δ_{jh}) represents the probability of person *n* giving a person location with x_j. δ_{jh} is a parameter of transition location. Meanwhile, by definition, the mean of the difficulty of the thresholds is the item difficulty for each item (Linacre, 2022, p. 686).

$$P(x_j|\theta, \delta_{jh}) = \frac{\exp\left[\sum_{h=0}^{x_j}\left(\theta - \delta_{jh}\right)\right]}{\sum_{k=0}^{m_j} \exp\left[\sum_{h=0}^{k}\left(\theta - \delta_{jh}\right)\right]} \tag{2.19}$$

In this study, the person means firm. In order to keep the terminology consistent with the literature, this study used person in Chapter 2. In the following chapters, this study will use the form of the firm (person), where the firm indicates the participants in this study, and (person) in the bracket means the name in the literature and software.

CHAPTER 3. METHODS

3.1 Introduction

This study explored how prescriptive psychometric models (i.e., Dichotomous Rasch Model, & Partial Credit Model) perform in conducting comprehensive analyses to create the financial profitability index to rank firms' financial profitability performance by extracting information from financial statements through financial profitability ratios based on data discretization methods. Firstly, financial profitability ratios were discretized based on industry medians at the sector level, quartiles, deciles, and k-median algorithms, respectively. Secondly, the dichotomous Rasch model was used on data that has been discretized based on the industry median at the sector level, and the Partial Credit model was used on the data that has been discretized based on quartiles, deciles, and the k-median algorithm (see Table 3.1). It is worth noting that, in order to ensure the assumption of the Rasch model, some items need to be reversed to ensure monotonicity in the same direction, such as Effective Tax Rate.

Table 3.1 The Rasch Models and Four Data Discretization Methods

	Median	Quartiles	Deciles	K-Median
Dichotomous Rasch Model	X			
Partial Credit Model		X	X	X

3.2 Overview of the Dataset

3.2.1 Data of the Financial Ratios Items

Fifteen financial profitability ratios have been selected for this initial study: Effective Tax Rate, Gross Profit/Total Assets, After-tax Return on Average Common

Equity, After-tax Return on Total Stockholders' Equity, After-tax Return on Invested Capital, Gross Profit Margin, Net Profit Margin, Operating Profit Margin After Depreciation, Operating Profit Margin Before Depreciation, Pre-tax Return on Total Earning Assets, Pre-tax return on Net Operating Assets, Pre-tax Profit Margin, Return on Assets, Return on Capital Employed, and Return on Equity.

3.2.2 Data of the Firms

This study focused on the January 2018 firms' financial ratios which were downloaded from the Wharton Research Data Services (WRDS) database. In addition, this study adopted the following strategies to select firms: (1) The Global Industry Classification Standard (GICS) has been used as an industry classification in the study along with the GICS-20 industrial sector. This means that firms that do not have the GICS-20 sector number in the database have been excluded from this study. (2) Firms have been excluded from this study if they do not have a score in the S&P Quality Ranking to ensure the firms are commonly used and widely studied.

3.3 Data Discretization

The 15 financial profitability ratios were discretized based on industry medians at the sector level, quartiles, deciles, and k-median algorithm, respectively. The 15 financial profitability ratios were discretized using R (Version 4.1.3) into

- 0 and 1 based on the industry medians at the sector level,
- 1 to 4 based on sample quartiles,
- 1 to 10 based on the sample deciles, and
- Several clusters based on the k-median algorithm (Wang and Song's 2011, Song & Zhong, 2020) using package Ckmeans.1d.dp (Version4.3.4)

Last, the items were renamed after being transformed from ratio-type data to discretized data (see Table 3.2).

Table 3.2 Items' Names and Financial Profitability Ratios

Item Name	Financial Profitability Ratio
R1_EFFTA	Effective Tax Rate
R2_GPROF	Gross Profit/Total Assets
R3_AFTRE	After-tax Return on Average Common Equity
R4_AFTRE	After-tax Return on Total Stockholders' Equity
R5_AFTRE	After-tax Return on Invested Capital
R6_GPM	Gross Profit Margin
R7_NPM	Net Profit Margin
R8_OPMAD	Operating Profit Margin After Depreciation
R9_OPMBD	Operating Profit Margin Before Depreciation
R10_PRET	Pre-tax Return on Total Earning Assets
R11_PRET	Pre-tax return on Net Operating Assets
R12_PTPM	Pre-tax Profit Margin
R13_ROA	Return on Assets
R14_ROCE	Return on Capital Employed, and
R15_ROE	Return on Equity

3.4 Scale Development

The scale development process of financial profitability used Winsteps (Version 5.1.2, Linacre, 2022), and there were two main stages of development: the initial model and the final model. The initial model, the original Rasch model, used all 15 items. The model was constructed by removing any items from the model that violated the local independency assumption. Items were then removed one at a time according to the highest poor fit value, recalibrating after each removal until all items met the fit criterion. Scale validation indices within the Rasch framework include: dimensionality, item local independence, firm (person) separation, firm (person) reliability, item separation, item reliability, global fit, infit, and outfit. Then, the final model was produced. The final model was also examined based on the scale validation indices. In order to unify the standard, all Rasch models will use unified criteria to ensure the model assumptions and item qualities which is an advantage of Rasch models.

3.4.1 Analysis of Dimensionality

The Rasch model was designed to analyze the unidimensional latent structure of the items, and different techniques are available to explore the number of dimensions in any dataset. This study adopted Linacre's principal component analysis (PCA) of the residuals approach to explore dimensionality within the Rasch framework (Linacre, 2022). Within this framework, the unidimensionality assumption is tenable if the eigenvalue of the first contrast of the residuals is less than 2 (Linacre & Wright, 1993, p. 412). However, conducting a PCA is necessary but not sufficient; it is also necessary to combine the PCA results with the psychometric evidence and item content.

3.4.2 Local Item Independence

The local item independence assumption requires that items should be only related to the latent structure, which requires no significant correlation between each pair of items in the residuals (Lord & Novick, 1968). The local item independence assumption was checked based on the residual correlations between items in Winsteps Table 23.99 (Version 5.1.2). The violation of the local item independence assumption needs to be considered when residual correlations are greater than +0.7 (Linacre & Wright, 1993, p. 423). Meanwhile, items were removed based on the violation of the local item independence to address the dimensionality issue.

3.4.3 Model Quality Indices

In the Rasch model, reliability is related to data reproducibility (Linacre, 2022, p. 738). The firm (person) reliability can be interpreted in the same way as a traditional test reliability, such as Cronbach's alpha (Linacre, 2022, p. 738). To achieve higher firm (person) reliability and higher item reliability, a sample with a wide range of firm (person) ability and item difficulty is needed (Linacre, 2022, p. 738). Firm (person) separation indicates whether or not the measure is sensitive enough to distinguish people with high ability and low ability, and if not, more items may be needed (Linacre, 2022, p. 738). Item separation indicates whether or not the sample is larger enough to verify the hierarchy of item difficulty, and if not, a more diverse sample is needed (Linacre, 2022, p. 738). The critical values for a good firm (person) separation and firm (person) reliability are larger than 2.0 and 0.8, respectively. The critical values for item separation and item reliability are larger than 3.0 and 0.9, respectively (Linacre & Wright, 1993, p. 706).

Log-likelihood chi-squared was used to check the model for approximate global fit. Approximate global fit is tenable when the Log-likelihood chi-squared is non-significant at the .05 level (Linacre & Wright, 1993, p. 465). Item fit indices show how well the observed data fit the Rasch model at the item level (Bond & Fox, 2015; Linacre, 2002a). The item weighted (infit) and item unweighted (outfit) mean square values (MNSQ) were used to determine the single item fit. The infit is sensitive to inliers and the outfit is unweighted and therefore more sensitive to outliers (Linacre, 2002a). Items can be considered for removal based on the commonly used critical range for infit and outfit, which is the same for both, between 0.5 and 1.5 (Linacre, 2002a). Item characteristic curves (ICCs) also were used to visually inspect fit for each item, and the Wright map was used to observe the distribution of item difficulty and firm (person) ability.

CHAPTER 4. RESULTS

4.1 Results of Median-based model

4.1.1 Initial Model

Results of the dimensionality analysis (see Table 4.1) shows that 37.7% of the total raw variance has been explained by the initial median-based dichotomous Rasch model. Furthermore, 36.7% of the raw variance was explained by firms, and 1.1% was explained by items. The eigenvalue of the first contrast was approximately 3.59, which was larger than the criterion value of 2.00 (Linacre & Wright, 1993, p. 412).

There were two pairs of items that violated the item local independency assumption with a criterion value of +.7 for the residual correlations (Linacre & Wright, 1993, p. 423): item R3_AFTRE and item R4_AFTRE with a residual correlation of .97, and item R7_NPM and item R12_PTPM with a residual correlation of .87.

The initial median-based dichotomous Rasch model had an item separation of 1.00 with item reliability of .50, which were both lower than the criterion values. In addition, the model had a firm (person) separation of 1.91 with firm (person) reliability of .78, which were both slightly lower than the criterion values.

The global fit of the initial median-based dichotomous Rasch model was tenable according to the non-statistically significant result of the log-likelihood chi-square test ($p = .53$) (Linacre & Wright, 1993, p. 465). At the item level, the infit range was from 0.56 to 2.10, and the outfit range was from 0.42 to 2.87.

4.1.2 Final Model

In the development of final median-based dichotomous Rasch model, R3_AFTRE and R12_PTPM were removed according to violations of the item local independence assumption. Subsequently, R1_EFFTA, R2_GPROF, R6_GPM, R11_PRET, and R9_OPMBD were removed one at a time according to the highest poor fit value after each recalibration until all items met the fit criterion. Finally, the final median-based dichotomous Rasch model included eight items.

Results of the dimensionality analysis (see Table 4.1) showed that 28.9% of the total raw variance has been explained by the final median-based dichotomous Rasch model. Furthermore, 28.4% of the raw variance was explained by firms, and 0.5% was explained by items. The eigenvalue of the first contrast was approximately 2.30, which was slightly larger than the criterion value of 2.00 (Linacre & Wright, 1993, p. 412).

The final median-based dichotomous Rasch model had an item separation of 0.00 with item reliability of .00, which were both lower than the criterion values. In addition, the model had a firm (person) separation of 1.37 with firm (person) reliability of .65, which were both lower than the criterion values.

The global fit of the final median-based dichotomous Rasch model was tenable according to the non-statistically significant result of the log-likelihood chi-square test ($p = .50$) (Linacre & Wright, 1993, p. 465). At the item level, the infit range was from 0.67 to 1.39, and the outfit range was from 0.57 to 1.42, which both fell into the criterion range from 0.5 to 1.5 (Linacre, 2002a). The results of the infit and outfit values also aligned with each ICC (see Figure 4.2). The Wright map shows that the eight items crowded together around the item difficulty of 0 logits (see Figure 4.1 and Table 4.9).

Table 4.1 Results of Median-based Dichotomous Rasch Models

	Initial	Final
Sample Size	412	412
No. of Items	15	8
Firm (Person) Separation	1.91	1.37
Firm (Person) Reliability	.78	.65
Item Separation	1.00	0.00
Item Reliability	.50	.00
Eigenvalue of Model	9.09	3.25
	37.7%	28.9%
Eigenvalue of Firms	8.84	3.19
(Persons)	36.7%	28.4%
Eigenvalue of Items	0.26	0.05
	1.1%	0.5%
Eigenvalue of 1st Contrast	3.59	2.30
	14.9%	20.4%
Global Fit	.53	.50
Infit Range	[0.56, 2.10]	[0.67, 1.39]
Outfit Range	[0.42, 2.87]	[0.57, 1.42]

Note. The percentage under each eigenvalue shows us how many variances have been explained by it. Global Fit = p-value of Log-likelihood chi-squared.

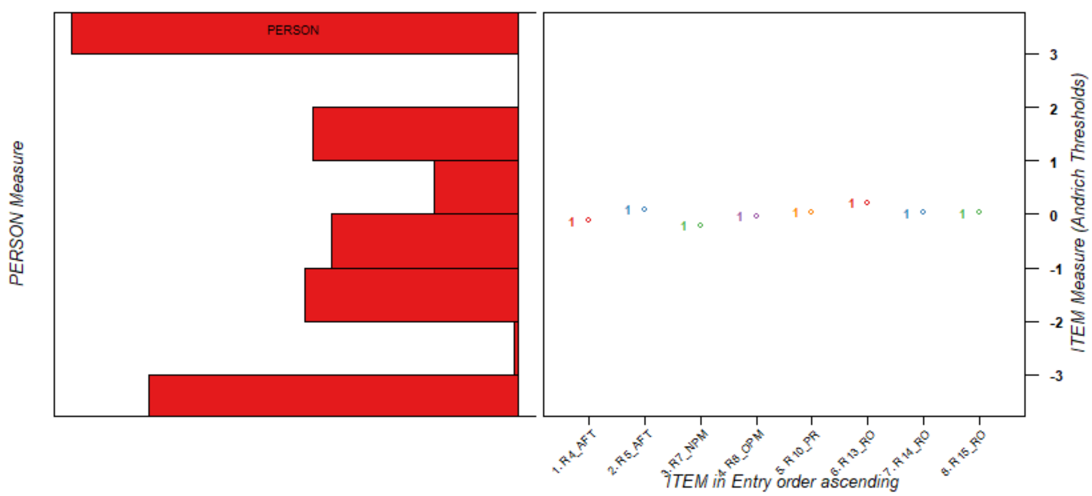

Figure 4.1 Wright Map of Final Median-based Dichotomous Rasch Model

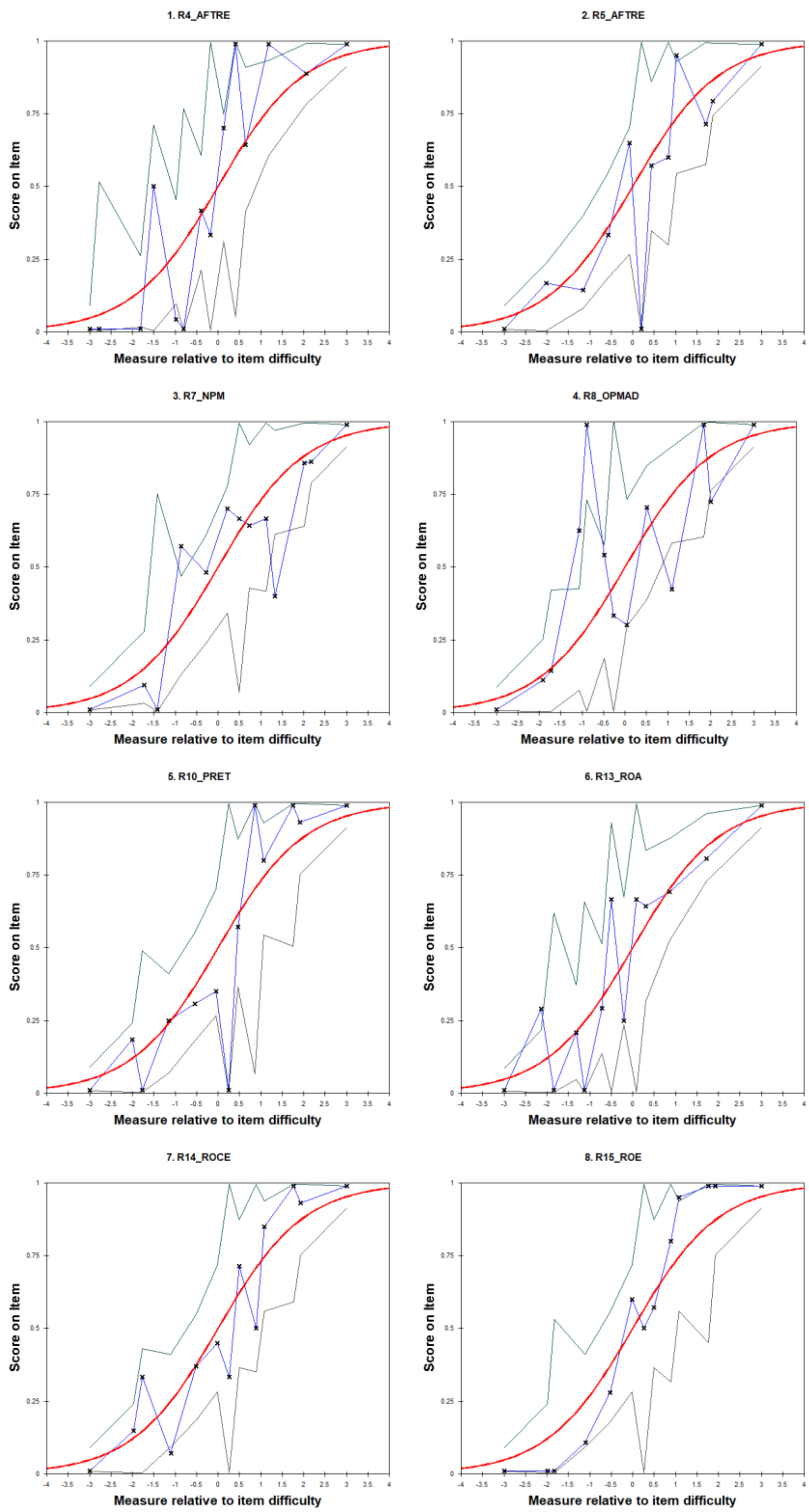

Figure 4.2 Item Characteristic Curves of Final Median-based Dichotomous Rasch Model

4.2 Results of Quartiles-Based model

4.2.1 Initial Model

Results of the dimensionality analysis (see Table 4.2) shows that 50.1% of the total raw variance has been explained by the initial quartiles-based Partial Credit model. Furthermore, 49.4% of the raw variance was explained by firms, and 0.6% was explained by items. The eigenvalue of the first contrast was approximately 3.96, which was larger than the criterion value of 2.00 (Linacre & Wright, 1993, p. 412).

There were four pairs of items that violated the item local independency assumption with a criterion value of +.7 for the residual correlations (Linacre & Wright, 1993, p. 423): R3_AFTRE and R4_AFTRE with a residual correlation of .97; R7_NPM and R12_PTPM with a residual correlation of .87; R4_AFTRE and R15_ROE with a residual correlation of .77; and R3_AFTRE and R15_ROE with a residual correlation of .76.

The initial quartiles-based Partial Credit model had an item separation of 1.14 and item reliability of .57, which were both lower than the criterion values. In addition, the model had a firm (person) separation of 2.60 and firm (person) reliability of .87, which were both higher than the criterion values.

The global fit of the initial quartiles-based Partial Credit model was tenable according to the non-statistically significant result of the log-likelihood chi-square test ($p = .42$) (Linacre & Wright, 1993, p. 465). At the item level, the infit range was from 0.51 to 2.31, and the outfit range was from 0.47 to 3.20.

4.2.2 Final Model

In the development of the final quantiles-based Partial Credit model, R3_AFTRE, R12_PTPM, and R4_AFTRE were removed according to violations of the item local independence assumption. Subsequently, R1_EFFTA, R2_GPROF, R6_GPM, and

R9_OPMBD were removed one at a time according to the highest poor fit value after each recalibration until all items met the fit criterion. Finally, the final quartiles-based Partial Credit model included eight items.

Results of the dimensionality analysis (see Table 4.2) shows that 56.8% of the total raw variance has been explained by the final quartiles-based Partial Credit model. Furthermore, 37.9% of the raw variance was explained by firms, and 18.9% was explained by items. The eigenvalue of the first contrast was approximately 2.12, which was slightly larger than the criterion value of 2.00 (Linacre & Wright, 1993, p. 412).

The final quartiles-based Partial Credit model had an item separation of 0.00 and item reliability of .00, which were both lower than the criterion values. In addition, the model had a firm (person) separation of 2.52 and firm (person) reliability of .86, which were both higher than the criterion values.

The global fit of the final quartiles-based Partial Credit model was tenable according to the non-statistically significant result of the log-likelihood chi-square test ($p = .56$) (Linacre & Wright, 1993, p. 465). At the item level, the infit range was from 0.64 to 1.34, and the outfit range was from 0.63 to 1.38, which fell into the criterion range from 0.5 to 1.5 (Linacre, 2002a). Results of the infit and outfit values also aligned with each ICC (see Figure 4.4). The Wright map shows that all the eight items crowded together around the item difficulty of 0 logits (see Figure 4.3 and Table 4.9). In addition, the Wright map and the threshold table show that the thresholds were in the property order (see Figure 4.3 and Table 4.3), and the category probability curves also show that the order of the response categories and the thresholds were as intended (see Figure 4.5).

Table 4.2 Results of Quartiles-based Partial Credit Models

	Initial	Final
Sample Size	412	412
No. of Items	15	8
Firm (Person) Separation	2.60	2.52
Firm (Person) Reliability	.87	.86
Item Separation	1.14	0.00
Item Reliability	.57	.00
Eigenvalue of Model	15.03	10.52
	50.1%	56.8%
Eigenvalue of Firms (Persons)	14.85	7.02
	49.4%	37.9%
Eigenvalue of Items	0.19	3.50
	0.6%	18.9%
Eigenvalue of 1st Contrast	3.96	2.12
	13.2%	11.4%
Global Fit	.42	.56
Infit Range	[0.51, 2.31]	[0.64, 1.34]
Outfit Range	[0.47, 3.20]	[0.63, 1.38]

Note. The percentage under each eigenvalue shows us how many variances have been explained by it. Global Fit = p-value of Log-likelihood chi-squared.

Table 4.3 Thresholds (Observed Count) of Final Quartiles-based Partial Credit Model

	1	2	3	4
R5_AFTRE	NONE	-1.94	0.23	1.70
	(101)	(102)	(102)	(102)
R7_NPM	NONE	-1.90	0.22	1.69
	(102)	(103)	(103)	(103)
R8_OPMAD	NONE	-1.90	0.22	1.69
	(102)	(103)	(103)	(103)
R10_PRET	NONE	-1.94	0.23	1.74
	(98)	(98)	(98)	(99)
R11_PRET	NONE	-1.94	0.23	1.74
	(98)	(98)	(98)	(98)
R13_ROA	NONE	-1.93	0.22	1.69
	(102)	(103)	(103)	(103)
R14_ROCE	NONE	-1.93	0.22	1.69
	(102)	(103)	(103)	(103)
R15_ROE	NONE	-1.96	0.22	1.69
	(100)	(100)	(100)	(100)

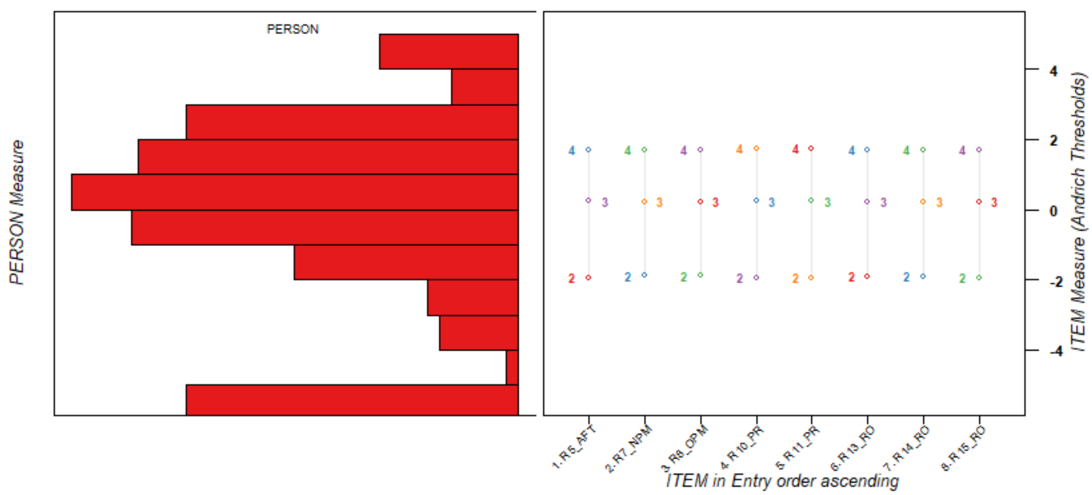

Figure 4.3 Wright Map of Final Quartiles-based Partial Credit Model

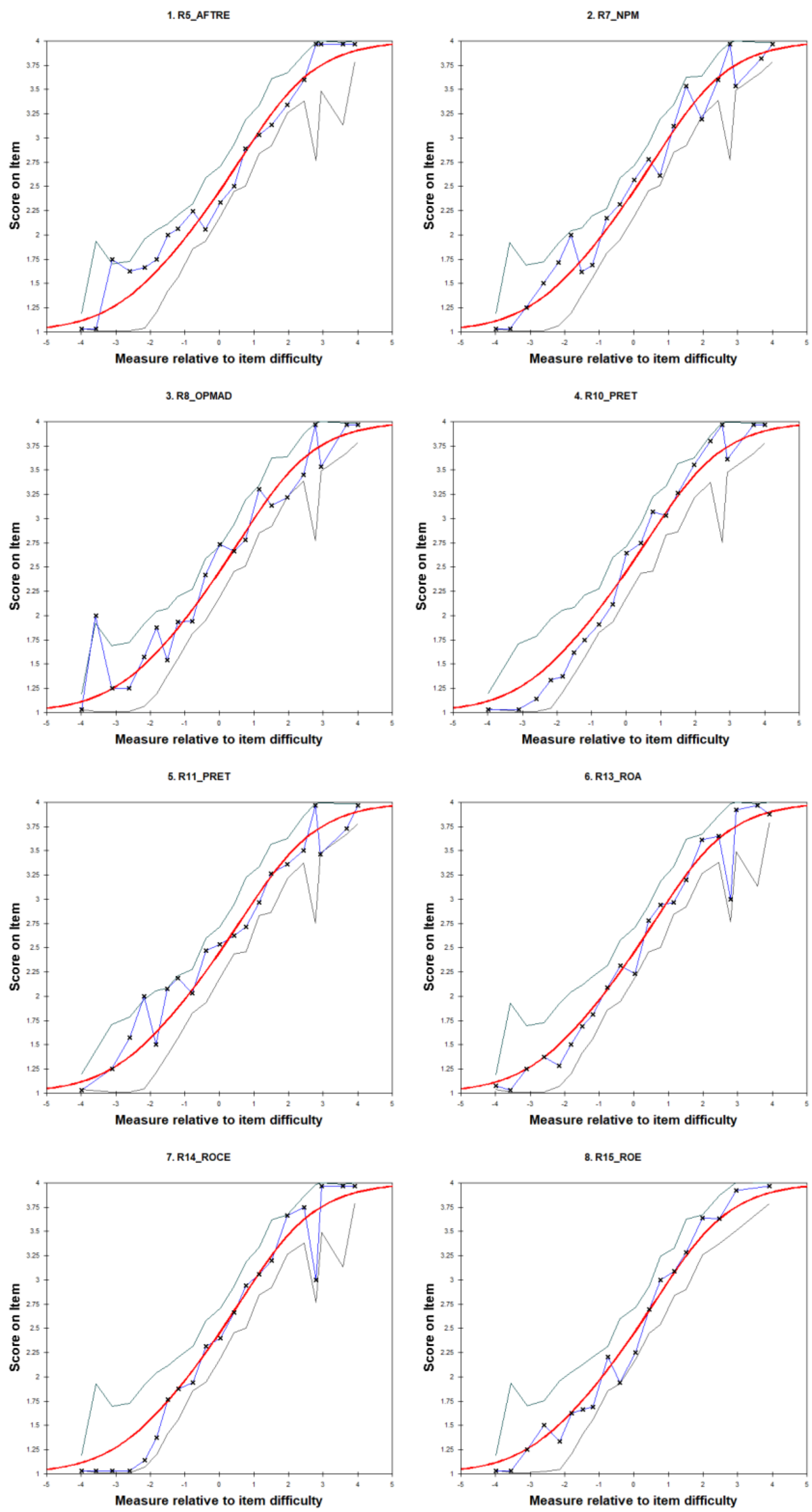

Figure 4.4 Item Characteristic Curves of Final Quartiles-based Partial Credit Model

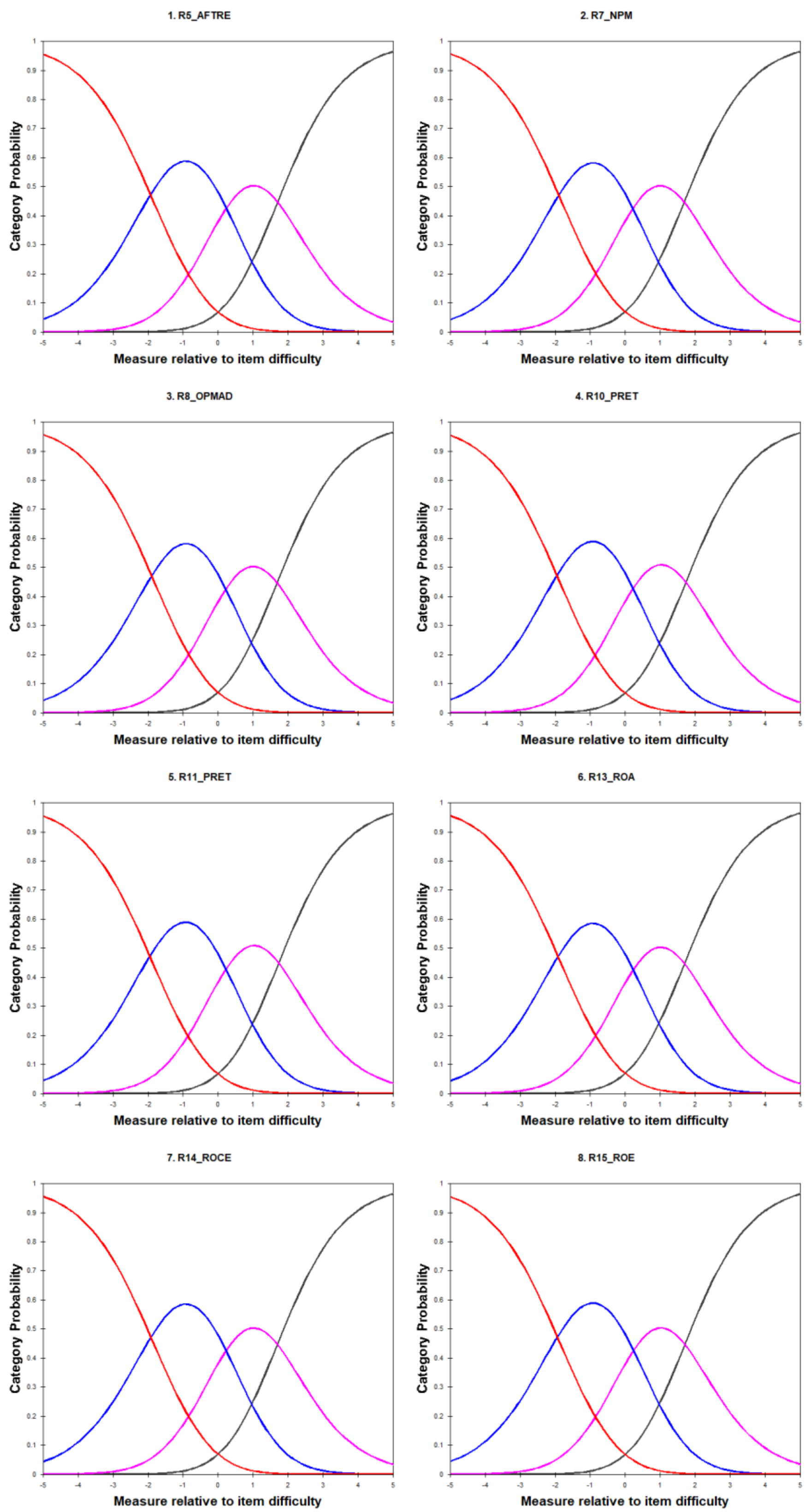

Figure 4.5 Category Probability Curves of Final Quartiles-based Partial Credit Model

4.3 Results of Deciles-Based model

4.3.1 Initial Model

Results of the dimensionality analysis (see Table 4.4) shows that 54.9% of the total raw variance has been explained by the initial deciles-based Partial Credit model. Furthermore, 53.5% of the raw variance was explained by firms, and 1.4% was explained by items. The eigenvalue of the first contrast was approximately 4.07, which was larger than the criterion value of 2.00 (Linacre & Wright, 1993, p. 412).

There were five pairs of items that violated the item local independency assumption with a criterion value of +.7 for the residual correlations (Linacre & Wright, 1993, p. 423): R3_AFTRE and R4_AFTRE with a residual correlation of 1.00; R7_NPM and R12_PTPM with a residual correlation of .91; R8_OPMAD and R9_OPMBD with a residual correlation of .77; R4_AFTRE and R15_ROE with a residual correlation of .74; and R3_AFTRE and R15_ROE with a residual correlation of .74.

The initial deciles-based Partial Credit model had an item separation of 1.80 and item reliability of .76, which were both lower than the criterion values. In addition, the model had a firm (person) separation of 2.71 and firm (person) reliability of .88, which were both higher than the criterion values.

The global fit of the initial deciles-based Partial Credit model was tenable according to the non-statistically significant result of the log-likelihood chi-square test ($p = .44$) (Linacre & Wright, 1993, p. 465). At the item level, the infit range was from 0.49 to 2.45, and the outfit range was from 0.48 to 3.27.

4.3.2 Final Model

In the development of the final deciles-based Partial Credit model, R3_AFTRE, R12_PTPM, R9_OPMBD, and R4_AFTRE were removed according to violations of the item local independence assumption. Subsequently, R1_EFFTA, R2_GPROF, R6_GPM, and R11_PRET were removed one at a time according to the highest poor fit value after each recalibration until all items met the fit criterion. Finally, the final deciles-based Partial Credit model included seven items.

Results of the dimensionality analysis (see Table 4.4) shows that 73.2% of the total raw variance has been explained by the final deciles-based Partial Credit model. Furthermore, 38.8% of the raw variance was explained by firms, and 34.4% was explained by items. The eigenvalue of the first contrast was approximately 2.26, which was slightly larger than the criterion value of 2.00 (Linacre & Wright, 1993, p. 412).

The deciles-based final Partial Credit model had an item separation of 0.00 and item reliability of .00, which were both lower than the criterion values. In addition, the model had a firm (person) separation of 3.26 and firm (person) reliability of .91, which were both higher than the criterion values.

The global fit the model of the final deciles-based Partial Credit model was tenable according to the non-statistically significant result of the log-likelihood chi-square test ($p = .37$) (Linacre & Wright, 1993, p. 465). At the item level, the infit range was from 0.70 to 1.30, and the outfit range was from 0.66 to 1.31, which fell into the criterion range from 0.5 to 1.5 (Linacre, 2002a). The results of the infit and outfit values also aligned with each ICC (see Figure 4.7). The Wright map shows that all the seven items crowded together around the item difficulty of 0 logits (see Figure 4.6 and Table 4.9). In addition, the Wright map and threshold table show that the thresholds were in the property order (see Figure 4.6 and Table 4.5), and the category probability curves also show that the order of the response categories and the thresholds were as intended (See Figure 4.8).

Table 4.4 Results of Deciles-based Partial Credit Models

	Initial	Final
Sample Size	412	412
No. of Items	15	7
Firm (Person) Separation	2.71	3.26
Firm (Person) Reliability	.88	.91
Item Separation	1.80	0.00
Item Reliability	.76	.00
Eigenvalue of Model	18.29	19.17
	54.9%	73.2%
Eigenvalue of Firms	17.81	10.16
(Persons)	53.5%	38.8%
Eigenvalue of Items	0.48	9.02
	1.4%	34.4%
Eigenvalue of 1st Contrast	4.07	2.26
	12.2%	8.6%
Global Fit	.44	.37
Infit Range	[0.49, 2.45]	[0.70, 1.30]
Outfit Range	[0.48, 3.27]	[0.66, 1.31]

Note. The percentage under each eigenvalue shows us how many variances have been explained by it. Global Fit = p-value of Log-likelihood chi-squared.

Table 4.5 Thresholds (Observed Count) of Final Deciles-based Partial Credit Model

	1	2	3	4	5	6	7	8	9	10
R5_AFTRE	NONE	-3.10	-1.06	-0.38	0.03	0.32	0.58	0.84	1.14	1.60
	(40)	(40)	(40)	(41)	(41)	(41)	(41)	(41)	(41)	(41)
R7_NPM	NONE	-3.02	-1.02	-0.34	0.04	0.32	0.58	0.84	1.14	1.57
	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(42)
R8_OPMAD	NONE	-3.02	-1.02	-0.34	0.04	0.32	0.58	0.84	1.14	1.57
	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(42)
R10_PRET	NONE	-3.22	-1.07	-0.37	0.02	0.32	0.59	0.84	1.17	1.62
	(39)	(39)	(39)	(39)	(39)	(39)	(39)	(40)	(40)	(40)
R13_ROA	NONE	-3.11	-1.04	-0.35	0.03	0.32	0.58	0.84	1.14	1.57
	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(42)
R14_ROCE	NONE	-3.11	-1.04	-0.35	0.03	0.32	0.58	0.84	1.14	1.57
	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(41)	(42)
R15_ROE	NONE	-3.10	-1.06	-0.36	0.03	0.32	0.58	0.84	1.13	1.57
	(40)	(40)	(40)	(40)	(40)	(40)	(40)	(40)	(40)	(40)

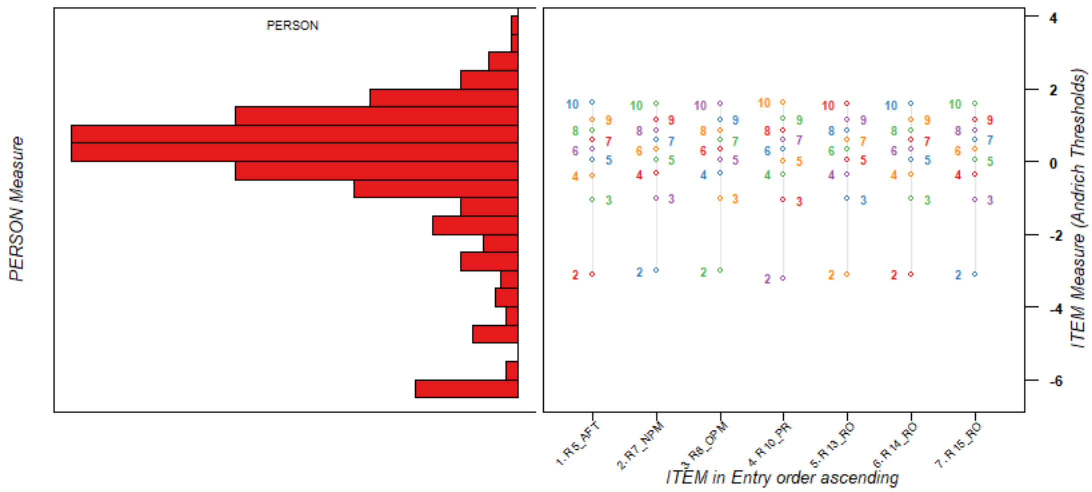

Figure 4.6 Wright Map of Final Deciles-based Partial Credit Model

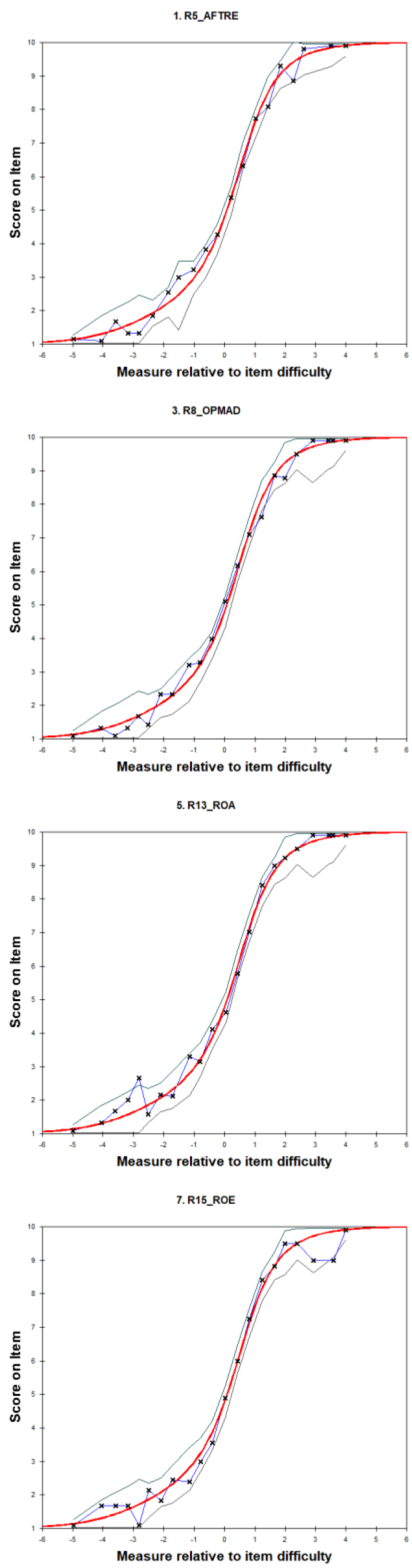

Figure 4.7 Item Characteristic Curves of Final Deciles-based Partial Credit Model

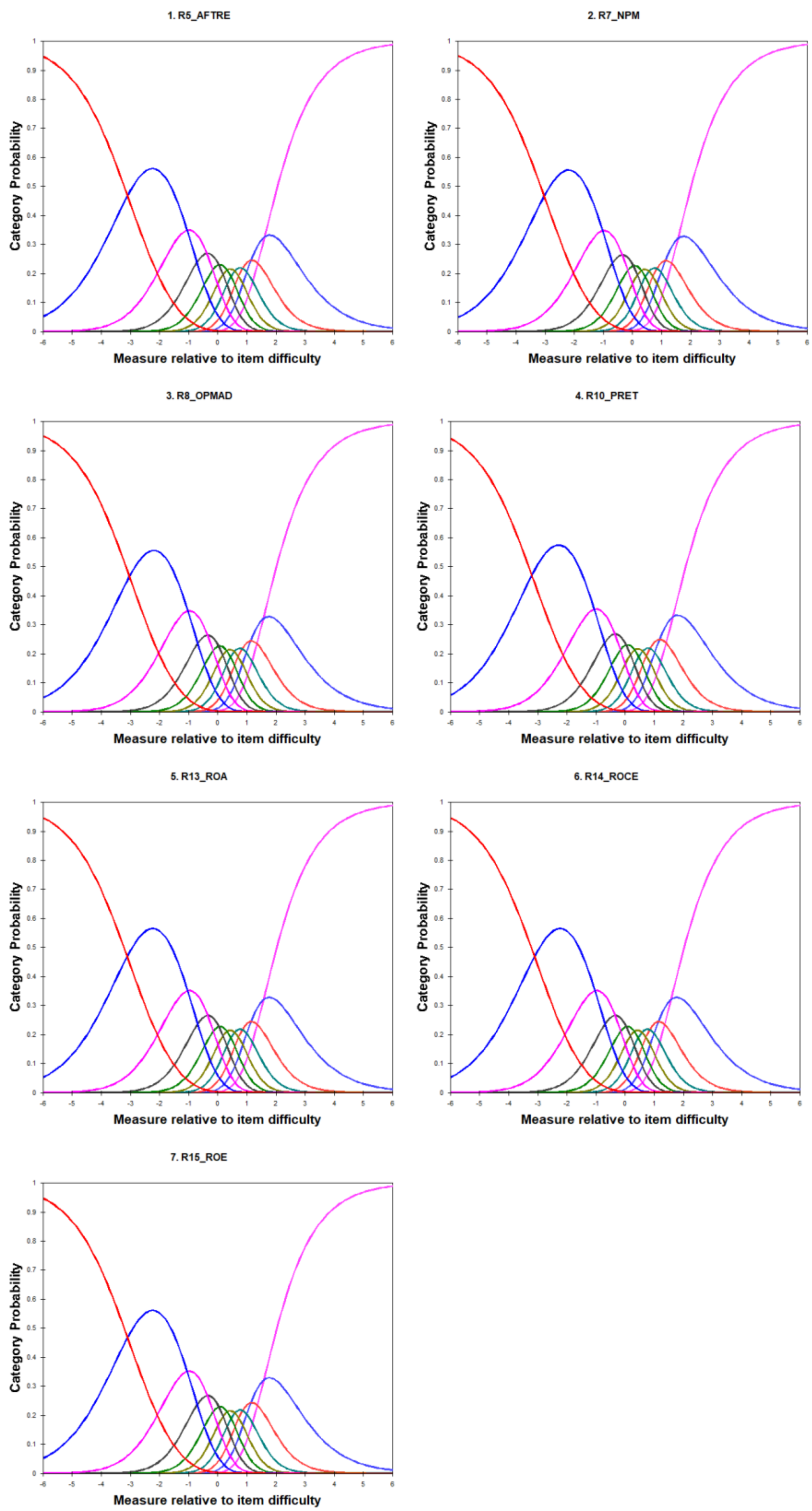

Figure 4.8 Category Probability Curves of Final Deciles-based Partial Credit Model

4.4 Results of K-median-based model

4.4.1 Initial Model

Results of the dimensionality analysis (see Table 4.6) shows that 63.4% of the total raw variance has been explained by the initial k-median-based Partial Credit model. Furthermore, 40.2% of the raw variance was explained by firms, and 23.2% was explained by items. The eigenvalue of the first contrast was approximately 3.80, which was larger than the criterion value of 2.00 (Linacre & Wright, 1993, p. 412).

There was one pair of items that violated the item local independency assumption with a criterion value of +.7 of the residual correlations (Linacre & Wright, 1993, p. 423): R3_AFTRE and R4_AFTRE with a residual correlation of .99.

The initial k-median-based Partial Credit model had an item separation of 13.62 and item reliability of .99, which were both higher than the criterion values. In addition, the model had a firm (person) separation of 3.22 and firm (person) reliability of .91, which were both higher than the criterion values.

The global fit of the initial k-median-based Partial Credit model was tenable according to the non-statistically significant log-likelihood chi-square test ($p = .49$) (Linacre & Wright, 1993, p. 465). At the item level, the infit range was from 0.61 to 2.25, and the outfit range was from 0.58. to 2.87.

4.4.2 Final Model

In the development of the final k-median-based Partial Credit model, R3_AFTRE was removed according to violations of the item local independence assumption. Subsequently, R2_GPROF, R6_GPM, R1_EFFTA, R11_PRET, R9_OPMBD, and R8_OPMAD were removed one at a time according to the highest poor fit value after each recalibration until all items met the fit criterion. Finally, the final k-median-based Partial Credit model included eight items.

Results of the dimensionality analysis (see Table 4.6) shows that 82.3% of the total raw variance has been explained by the final k-median-based Partial Credit model. Furthermore, 56.5% of the raw variance was explained by firms, and 25.8% was explained by items. The eigenvalue of the first contrast was approximately 2.23, which was slightly larger than the criterion value of 2.00 (Linacre & Wright, 1993, p. 412).

The final k-median-based Partial Credit model had an item separation of 19.68 and item reliability of 1.00, which were both higher than the criterion values. In addition, the model had a firm (person) separation of 4.23 and firm (person) reliability of .95, which were both higher than the criterion values.

The global fit of the final k-median-based Partial Credit model was tenable according to the non-statistically significant result of the log-likelihood chi-square test ($p = .74$) (Linacre & Wright, 1993, p. 465). At the item level, the infit range was from 0.62 to 1.13, and the outfit range was from 0.51 to 1.27 which all fell into the criterion range from 0.5 to 1.5 (Linacre, 2002a). The results of the infit and outfit values also aligned with each ICC (see Figure 4.10). The Wright map shows that the items spread out along the scale roughly from -4.5 logits to 3 logits (see Figure 4.9 and Table 4.9). Some items appeared to be disordered at some thresholds, such as R4_AFTRE, R12_PTPM, and R15_ROE (see Figure 4.9 and Table 4.7). The category probability curves also show that there were problems of disorder with some items, which aligned with the results of the Wright map (See Figure 4.11).

Table 4.6 Results of K-median-based Partial Credit Models

	Initial	Final
Sample Size	412	412
No. of Items	15	8
Firm (Person) Separation	3.22	4.23
Firm (Person) Reliability	.91	.95
Item Separation	13.62	19.68
Item Reliability	.99	1.00
Eigenvalue of Model	25.95	37.29
	63.4%	82.3%
Eigenvalue of Firms	16.44	25.60
(Persons)	40.2%	56.5%
Eigenvalue of Items	9.51	11.70
	23.2%	25.8%
Eigenvalue of 1st Contrast	3.80	2.23
	9.3%	4.9%
Global Fit	.49	.74
Infit Range	[0.61, 2.25]	[0.62, 1.13]
Outfit Range	[0.58, 2.87]	[0.51, 1.27]

Note. The percentage under each eigenvalue shows us how many variances have been explained by it. Global Fit = p-value of Log-likelihood chi-squared.

Table 4.7 Thresholds (Observed Count) of Final K-median-based Partial Credit Model

	1	2	3	4	5	6	7	8	9	10	11
R4_AFTRE	NONE (1)	-10.78 (10)	-8.24 (51)	-2.21 (226)	5.16 (112)	9.20 (7)	8.60 (2)	7.95 (2)			
R5_AFTRE	NONE (1)	-10.22 (5)	-9.17 (28)	-4.69 (75)	0.78 (145)	4.50 (113)	7.25 (33)	8.64 (7)			
R7_NPM	NONE (3)	-9.72 (13)	-7.41 (226)	4.12 (169)							
R10_PRET	NONE (25)	-5.89 (162)	3.20 (144)	6.49 (62)							
R12_PTPM	NONE (1)	-9.75 (2)	-8.06 (1)	-10.31 (7)	-7.72 (7)	-7.49 (52)	-1.45 (112)	2.95 (116)	5.31 (83)	7.45 (28)	9.94 (2)
R13_ROA	NONE (23)	-6.29 (220)	4.14 (168)								
R14_ROCE	NONE (16)	-7.36 (237)	4.36 (157)	11.52 (1)							
R15_ROE	NONE (9)	-8.16 (15)	-6.17 (67)	-0.14 (173)	4.93 (87)	6.65 (45)	9.74 (3)	8.91 (1)			

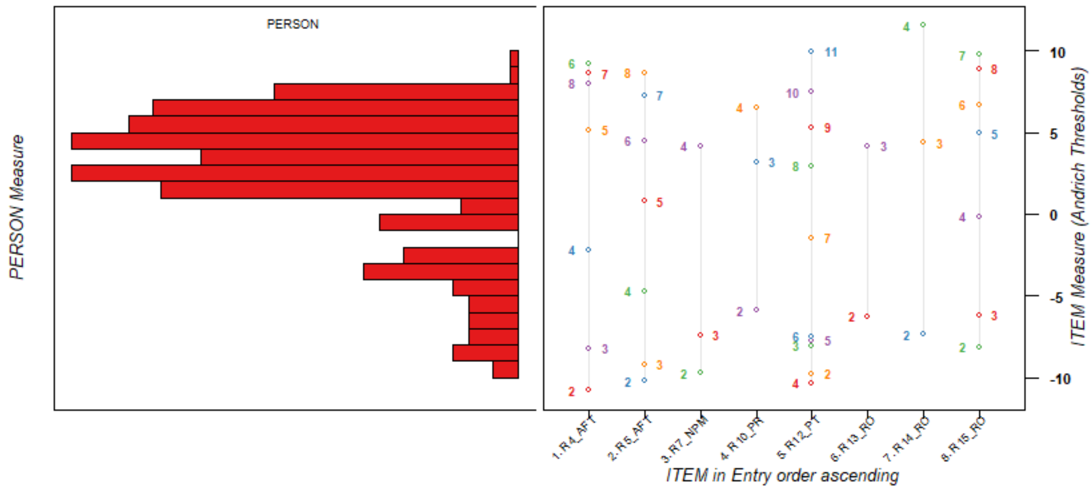

Figure 4.9 Wright Map of Final K-median-based Partial Credit Model

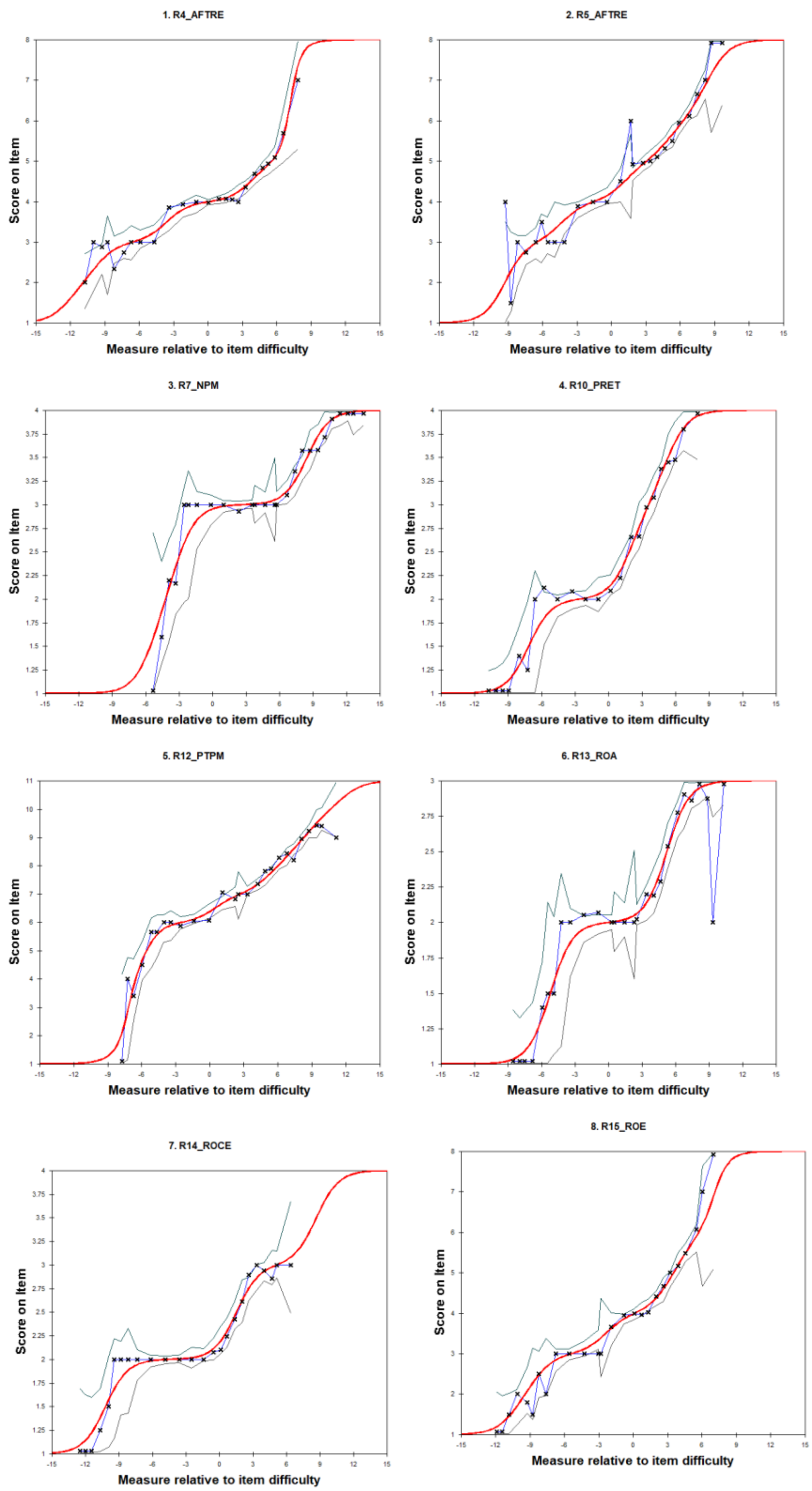

Figure 4.10 Item Characteristic Curves of Final K-median-based Partial Credit Model

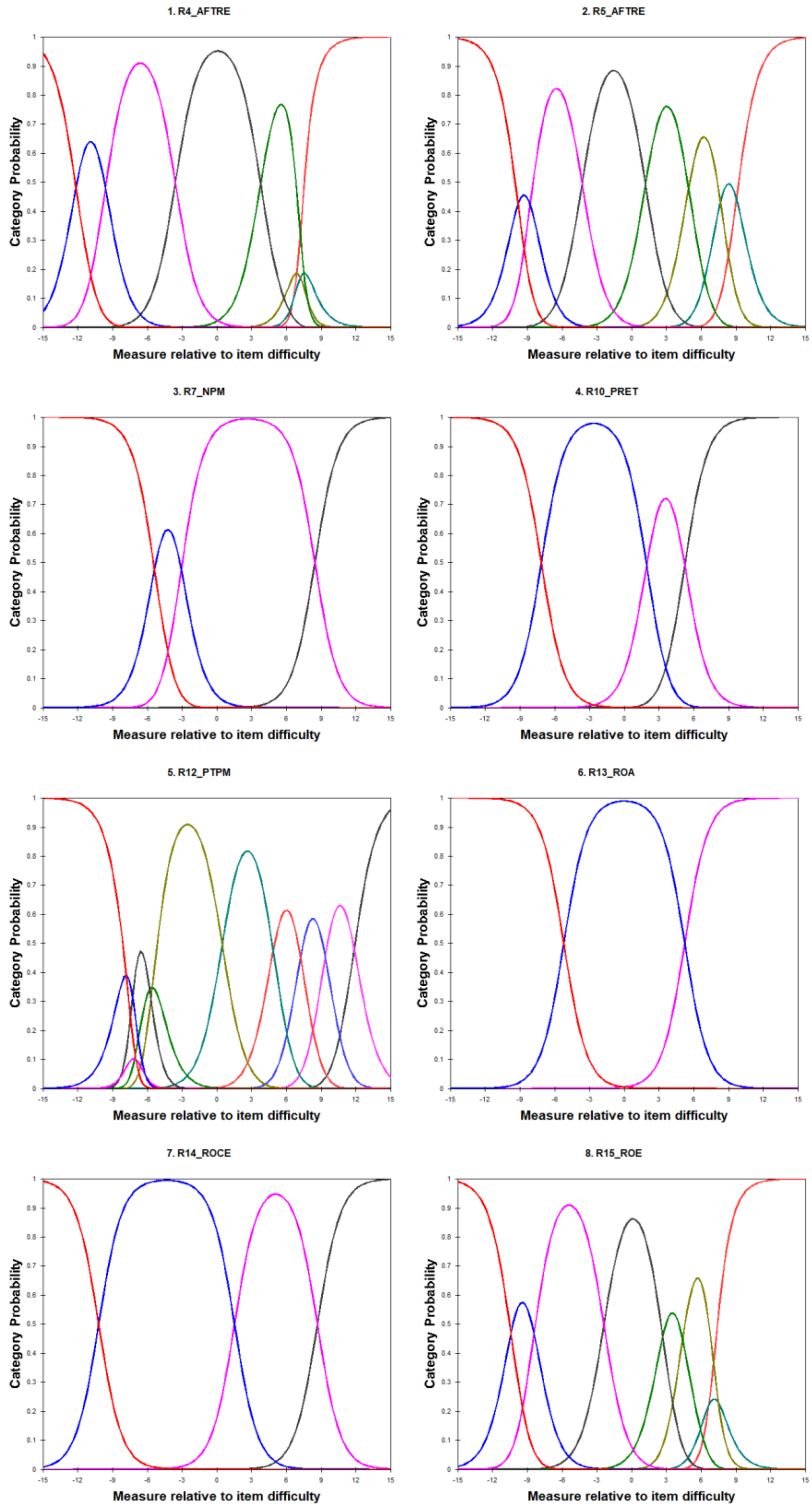

Figure 4.11 Category Probability Curves of Final K-median-based Partial Credit Model

4.5 Overall

Table 4.8 shows a summary of the Rasch model performance on each initial model and final model based on the financial profitability ratios in January 2018. There were 15 items in the initial dichotomous Rasch model using a median-based method, ending with 8 items in the final median-based dichotomous Rasch model. There were 15 items in the initial Partial Credit model using a quartiles-based discretization method, ending with 8 items in the final quartiles-based Partial Credit model. There were 15 items in the initial Partial Credit model using a deciles-based discretization method, ending with 7 items in the final deciles-based Partial Credit model. There were 15 items in the initial Partial Credit model using a k-median-based discretization method, ending with 8 items in the final k-median-based Partial Credit model.

In summary, the unidimensional assumption, item local dependence assumption, global fit, infit and outfit were all tenable for the final Rasch models. The median-based final dichotomous Rasch model did not meet the criteria of firm (person) separation (reliability) and item separation (reliability). The quartiles-based final Partial Credit model and deciles-based final Partial Credit model did not meet the criteria of item separation (reliability). The k-median-based Partial Credit model met all the criteria, however, some items showed disordered thresholds.

Table 4.8 Overall Results of Models (Data = January 2018)

	UD	ILD	GF	FS	FR	IS	IR	IF	OF
Median-15	✗	2	✓	✗	✗	✗	✗	✗	✗
Median-8	✓	✓	✓	✗	✗	✗	✗	✓	✓
Quartiles-15	✗	4	✓	✓	✓	✗	✗	✗	✗
Quartiles-8	✓	✓	✓	✓	✓	✗	✗	✓	✓
Deciles-15	✗	5	✓	✓	✓	✗	✗	✗	✗
Deciles-7	✓	✓	✓	✓	✓	✗	✗	✓	✓
K-Median-15	✗	1	✓	✓	✓	✓	✓	✗	✗
K-Median-8	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note. UD = Unidimensional, ILD = Item Local Dependence, GF = Global Fit, FS = Firm (Person) Separation, FR = Firm (Person) Reliability, IS = Item Separation, IR = Item Reliability, IF = Infit, and OF = Outfit. ✓ = holding, ✗ = out of acceptable range.

Table 4.9 shows a comparison of the item measures and fit indices of the final Rasch models under each data discretization method. Upon further examination, six items performed relatively well across all four Rasch models based on different data discretization methods: R5_AFTRE, R7_NPM, R10_PRET, R13_ROA, R14_ROCE, and R15_ROE. The distributions of item difficulty were very similar among the final median-based dichotomous Rasch model, final quartiles-based Partial Credit model, and final deciles-based Partial Credit model, which were all crowded together around 0 logits. The distribution of item difficulty was loosely distributed between -4.5 to 3 logits in the final K-median-based Partial Credit model.

Table 4.9 Item Measures and Fit Indices of Final Rasch Models

	Median		Quartiles		Deciles		K-median	
	Measure SE	Infit Outfit	Measure SE	Infit Outfit	Measure SE	Infit Outfit	Measure SE	Infit Outfit
R1_EFFTA	Exclude		Exclude		Exclude		Exclude	
R2_GPROF	Exclude		Exclude		Exclude		Exclude	
R3_AFTRE	Exclude		Exclude		Exclude		Exclude	
R4_AFTRE	-0.13 0.17	0.72 0.67	Exclude		Exclude		1.38 0.11	1.03 1.01
R5_AFTRE	0.08 0.17	1.01 1.07	0.00 0.08	1.17 1.18	0.00 0.04	1.24 1.31	-0.42 0.09	1.00 0.92
R6_GPM	Exclude		Exclude		Exclude		Exclude	
R7_NPM	-0.22 0.17	1.22 1.23	0.00 0.08	1.18 1.17	0.01 0.04	1.20 1.12	-4.34 0.13	0.99 1.08
R8_OPMAD	-0.04 0.17	1.39 1.42	0.00 0.08	1.19 1.20	0.01 0.04	1.30 1.20	Exclude	
R9_OPMBD	Exclude		Exclude		Exclude		Exclude	
R10_PRET	0.05 0.18	0.94 0.94	0.01 0.09	0.66 0.65	-0.01 0.04	0.86 0.79	1.26 0.11	1.09 1.03
R11_PRET	Exclude		0.01 0.09	1.34 1.38	Exclude		Exclude	
R12_PTPM	Exclude		Exclude		Exclude		-1.91 0.08	1.13 1.05
R13_ROA	0.20 0.17	1.13 1.26	-0.01 0.08	0.99 1.09	0.00 0.04	1.03 1.08	-1.07 0.14	1.00 1.27
R14_ROCE	0.02 0.17	0.88 0.87	-0.01 0.08	0.64 0.63	0.00 0.04	0.70 0.66	2.84 0.14	0.74 0.54
R15_ROE	0.02 0.18	0.67 0.57	-0.02 0.09	0.76 0.75	-0.01 0.04	0.76 0.77	2.25 0.09	0.62 0.51

Note. Measure = Item Difficulty. SE = Standard Error. Infit and Outfit based on the MNSQ.

Figure 4.12 shows the Pearson correlations of firms' financial profitability abilities among four models: median-based dichotomous Rasch model, Quartiles-based Partial Credit model, deciles-based Partial Credit model, and k-median-based Partial Credit model, which was represented as MB_Theta, QB_Theta, DB_Theta, and KMB_Theta, respectively. The Pearson correlations were all higher than +.07 among the four models.

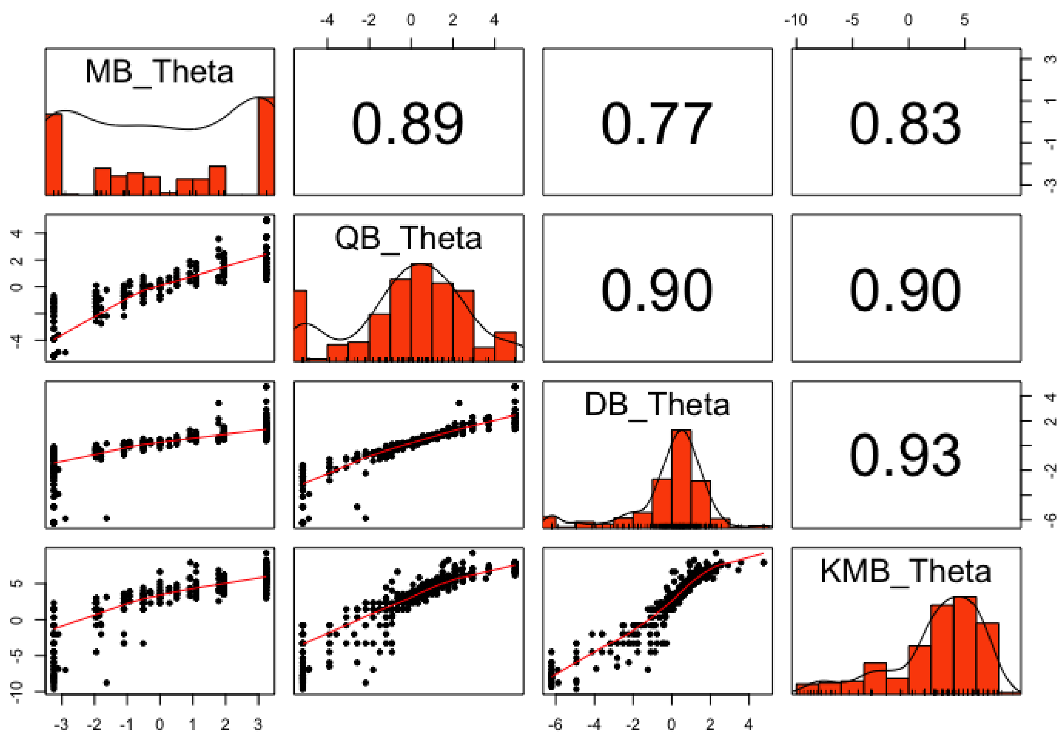

Figure 4.12 Plot Matrix of Correlations among Four Models

CHAPTER 5. DISCUSSION

5.1 Model Performance

The data discretization methods performed differently in regard to separation and reliability under the dichotomous Rasch and Partial Credit models. Among the median-based, quartiles-based, and deciles-based models, firm (person) separation, firm (person) reliability, item separation, and item reliability increased as the number of categories increased. This is because more information is retained as the number of categories increases, allowing the model to better distinguish the items' difficulty and distinguish among participants based on their ability.

Between the initial models and final models, firm (person) separation and firm (person) reliability dropped slightly when the number of items decreased due to model assumptions and item qualities in the median-based dichotomous Rasch model and quartiles-based Partial Credit model. However, between the initial models and final models, firm (person) separation and firm (person) reliability increased slightly in the deciles-based Partial Credit model and k-median-based Partial Credit model.

Item separation and item reliability of the final median-based, quartiles-based, and deciles-based models dropped to zero from the initial models' values because most items crowded together around an extremely similar item difficulty logit of the initial models, and only a few items were spread out on the scale away from the majority item group. Therefore, the item separation and item reliability of the final models show a dramatic drop from the initial models if these few items have to be removed based on the model criteria of the assumptions and item qualities because these few items occupied the main contribution of the item separation (reliability). This caused the number and variance of firms around the test items to become insufficient to separate items when the items all crowded together in the final model. This created a phenomenon of local insufficient sample size, even though the sample size was globally sufficient. To address this issue, researchers can adjust benchmarks if there is practical evidence that can support it, or add more items that can be used as a potential solution.

The k-median-based model did not show a local insufficient sample size after items from the initial model were removed due to the criteria of model assumptions and item qualities issues. Both initial and final k-median-based models showed good performance in firm (person) separation, firm (person) reliability, item separation, and item reliability compared to the median-based, quartiles-based, and deciles-based models. Also, the performance of firm (person) separation, firm (person) reliability, item separation, and item reliability were improved after all items that did not satisfy the criteria of the model assumptions and item qualities were removed. This demonstrated that using the dynamic clustering method for each item individually can better extract information and present the data pattern in preparation for building psychometric models. The model benefitted from the k-median clustering algorithm, allowing each item to have a different number of categories, compared to the median-based, quartiles-based, and deciles-based models, which fixed the number of categories same across all items.

According to the criterion of the item local independence assumption, the correlation of the residuals should be no greater than +0.7 between each pair of items (Linacre & Wright, 1993, p. 423). Based on this criterion, different data discretization methods influenced items' performance on the item local independence assumption. There were two pairs of items that violated the item local independence assumption in the initial median-based dichotomous Rasch model, four pairs of items that violated the item local independence assumption in the initial quartiles-based Partial Credit model, five pairs of items that violated the item local independence assumption in the initial deciles-based Partial Credit model, and one pair of items that violated the item local independence assumption in the initial k-median-based Partial Credit model. This study used the data-driven approach to detect the violations of the item local independence assumption. In practice, the violations of the item local independence can be conceptually detected at the early stages of item development. The choice of which item in the pair should be removed can be determined based on the theoretical foundations and practical experience.

According to the PCA on the residuals in the final models, all eigenvalues of the first contrast of the residuals were close to the criterion value of 2 (Linacre & Wright, 1993, p. 412) across all four data discretization conditions, which assumed that the unidimensional assumption as tenable. In the final median-based dichotomous Rasch model, 28.9% of total variance has been explained by the dichotomous Rasch model, and 20.4% of variance has been explained by the first contrast of the residuals, which is very high. In the final quartiles-based Partial Credit model, 56.8% of the variance has been explained by the Partial Credit model, and 11.4% of the variance has been explained by the first contrast of the residuals, which is much lower than the portion of the variance explained by the Partial Credit model. In the final deciles-based Partial Credit model, 73.2% of the variance has been explained by the Partial Credit model, and 8.6% of the variance has been explained by the first contrast of the residuals, which is much lower than the portion of the variance explained by the Partial Credit model. In the final k-median-based Partial Credit model, 82.3% of the variance has been explained by the Partial Credit model, and 4.9% of the variance has been explained by the first contrast of the residuals, which is much lower than the portion of the variance explained by the Partial Credit model.

Among the median-based, quartiles-based, and the deciles-based models, a higher portion of variance was explained by the Rasch model, and a lower portion of variance was explained by the first contrast of the residuals as the number of categories increased. However, it is not recommended to pursue maximum information retention by blindly increasing the number of categories. Linacre (2002b) suggested that the minimum number per category should be 10 to ensure the high quality of estimation. Therefore, the maximum number of the categories is restricted by the sample size and discreteness of the data. The dynamic data clustering method, k-median, can further improve the portion of the variance explained by the model and reduce the portion of the variance explained by the first contrast of the residuals.

Using the ICCs to compare the dichotomous Rasch model and the Partial Credit model, results showed that the data fit the Partial Credit model better than the dichotomous Rasch model. This is due to more information retained with the Partial Credit model, and the natural advantage of the Partial Credit model, which allows items to have their own thresholds. The ICCs also showed that between the final quartiles-based Partial Credit model and final deciles-based Partial Credit model, the data fit the model better as the number of categories increased. The ICCs of the final k-median-based Partial Credit model performed differently and more flexibility among different items again allowing different items to have different numbers of categories based on the data patterns.

Andrich threshold disorder was not detected in the final quartiles-based Partial Credit model or the final deciles-based Partial Credit model; however, results showed the threshold disorder in the final k-median-based Partial Credit model. This issue of threshold disorder happened on both ends: the maximum and minimum. Through further examination of the data, a common cause was found that the number of the data at both ends is prone to be small under the k-median algorithm, especially when there are extreme values in the data. Linacre (2022) suggested "Andrich thresholds: disordered thresholds are no problem for the formulation of polytomous Rasch models, nor for estimating Rasch measures, nor do they cause misfit to the Rasch model. They are only a problem if the Andrich thresholds are conceptualized as the category boundaries on the latent variable." Another potential common solution is to consider collapsing the categories to ensure the minimum requirement of 10 items in each category to achieve valid estimation (Linacre, 2002b). After that, the further collapsing process can be considered again if there are still threshold disorders.

In the final model, six out of seven or eight items are the common items in the final model across the four data discretization methods (i.e., R5_AFTRE, R7_NPM, R10_PRET, R13_ROA, R14_ROCE, R15_ROE). Results showed that most items performed consistently under the data-driven item selection approach based on model assumptions and item qualities across the four data discretization methods.

5.2 Implications

The Rasch model family provides a set of psychometric tools to assist researchers to develop the scales: (1) researchers can use the dichotomous Rasch model to build the scale on dichotomous data; (2) researchers can use the rating scale model to build the scale on polytomous data when all the items share the same thresholds, and (3) research can use the partial credit mode to build the scale on polytomous data when each item has their own thresholds. Rasch model has the fewest components: an ability parameter for each person and a difficulty parameter for each item in the dichotomous Rasch model (Wright, 1977), or an ability parameter for each person and thresholds for each item in the polytomous Rasch model. Meanwhile, as a prescriptive probabilistic measurement model (Shaw, 1991), the Rasch model can assist researchers to create a ruler-type scale that is “item-free (item-distribution-free)” and “person-free (person-distribution-free)” (Linacre & Wright, 1993, p. 34; Stemler & Naples, 2021; Wright & Stone, 1979). Furthermore, all Rasch models can use the same criterion to ensure model assumptions and item qualities during scale development.

This study used financial profitability ratios as a demonstration to show the process of scale development, and explored how different data discretization methods affect scale development. Therefore, the focus of this study was to explore the performance of the dichotomous and polytomous Rasch models. Under this strategy, the study adopted the data-driven approach in the process of scale development, which simply removed items based on the criteria of model assumptions and item qualities. In practical scale development, the tradeoff strategy between theory-driven and data-driven approaches should be considered. It is necessary to systematically select items based on theory and practice instead of being overly immersed in the pursuit of extreme model performance, which is very significant in the practical process of scale development.

Furthermore, it is also meaningful to adopt the logic of grading to cluster each individual ratio-type data into each discretized item. In the traditional examination and survey analysis, the grading system commonly has been pre-designed and fixed. However, we can have more flexibility when analyzing the existing ratio-type data, allowing us to select the grading system or answer key based on what is needed in the practice. For example, Dorsey (2004), the director of the stock analysis in Morningstar, suggested that generally, a firm is in the right position when it can generalize the net margin above 15% (p. 23). Therefore, we can use 15% as a benchmark for net profit margin discretization. However, it is worth noting that benchmark figures always need to be checked against calculation formulas before adopting the suggested benchmarks from other resources because there is no uniform industry standard for the naming and calculation formula of financial ratios. This lack of standard is reflected in difference between analysts and databases (CFA, 2020, p. 244).

The Rasch theta value can be used as a comprehensive score to conduct a ranking or comparison analysis across firms within the same industry. Namely, a higher Rasch theta score indicates a higher performance in financial profitability. A lower Rasch theta score indicates a lower performance in financial profitability. In the context of traditional comparisons, the original financial profitability ratios have been used for a while for different purposes in practice. Therefore, original financial profitability ratios can be associated with the firms' ability (comprehensive or theta score) to remedy the loss of information due to the data transformation process. The psychometric model provides an efficient ranking solution for the comparison of firms' financial profitability performances or positions in the same industry.

The Rasch beta score can be used to detect which financial profitability ratios are easier for firms to achieve and which financial ratios are harder for firms to achieve during a time period. A financial profitability ratio with a higher beta score indicates it is harder for firms to achieve during the period and vice versa.

In conclusion, this study systematically conducted scale development on ratio-type data under the Rasch model and explored the effect of different data discretization methods on scale development by using financial profitability ratios as a demonstration. In the narrow sense, this study provided psychometric evidence to support using the Rasch model to create scales to measure the firms' financial profitability based on different kinds of data discretization methods. Therefore, it is meaningful to use the logic of examination to extract a comprehensive score from ratio-type data under the dichotomous or polytomous Rasch model. This can compensate for the incomplete coverage of a single item.

The results showed that retaining more categories can benefit the Rasch modeling thus keeping more information in the model. Using dynamic clustering algorithms, k-median, is better for extracting characteristic patterns of the ratio-type data and preparing the data for the Rasch model. However, this k-median algorithm may lead to insufficient sample sizes at both ends of the categories due to the presence of extreme values. In the Rasch model, some items have a threshold disorder issue, and future research can adopt category collapsing as a solution according to specific needs. Therefore, this study advocates that there is no single best solution for data discretization method for ratio-type data under the Rasch model. It is more reasonable to use the traditional algorithms if each item has characteristic benchmark/benchmarks, and harder and easier items need to be created. If there is a lack of benchmark information, the k-median clustering algorithm can achieve good modeling results.

In the broad sense, this study provided the psychometric evidence to support that using Rasch model to extract information from ratio-type data under the four discretization methods is appropriate. Therefore, the Rasch model can be adopted as a solution to extract information from ratio-type data when the research meets the two following requirements: (1) the logic of data discretization is consistent with how people process information through ratio-type data comparisons during decision-making to ensure data discretization is meaningful, and (2) many ratios measure one latent variable to ensure the unidimensionality assumption can be held conceptually.

In the more broader sense, the Rasch model can be used as a tool to extract information from the mass of quantitative and qualitative data when data discretization is practically meaningfully and all the data can conceptually measure the same latent variable. However, more studies need to be conducted to answer those questions.

5.3 Limitations and Future Research

This study adopted the purely data-driven approach to conducting the scale development instead of using the theoretical evidence and practical needs. In the practical setting of scale development, it is necessary to develop the scale by considering all psychometric evidence, theoretical evidence, and practical needs. In traditional examinations and questionnaires, the setting and grading of items are pre-tailored to the measurement population and measurement purpose. However, compared with traditional examinations and questionnaires, in this study, more attention needed to be paid to the existed ratio-type data. For example, different accounting standards and methods may affect the quality of financial ratios, such as the use of First-In, First-Out (FIFO) and the Last-In, First-Out (LIFO) may influence the inventory on the balance sheet, and the cost of goods sold on the income statement, such as inventory turnover (Holdren, 1964). Different industries have different ratio measurement tendencies. Therefore, in the practice of scale development, these factors need to be fully considered.

There is always a need of balancing tradeoffs in modeling. This study sacrificed some information in the data to extract features of each item to fit the requirements of the data format of the Rasch model. Fortunately, this information transformation fit the logic of human decision-making. Therefore, it is necessary to consider the meaning of the data transformation process when adopting the study's approach to extracting information from a dataset.

Another potential limitation is that many firms involve multiple types of businesses. For cross-industry firms, it is difficult to consider all the industries the firm is involved in when running the comparison in a single industry. In other words, we need to assume that the firm’s main industry sector is its only involved sector.

REFERENCES

Bandalos. D.L. (2018). *Measurement theory and applications for the social sciences*. The Guilford Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: springer.

Bond, T., & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences, third edition* (3rd ed.). Routledge. https://doi.org/10.4324/9781315814698

CDC (2022, March 03). *Calculating BMI Using the Metric System.* Centers for Disease Control and Prevention. https://www.cdc.gov/nccdphp/dnpao/growthcharts/training/bmiage/page5_1.html

CFA Institute. (2020). *CFA program curriculum: Level I, Volume 3, 2020*. Boston, Mass: CFA Institute.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

Dorsey, P. (2004). *The five rules for successful stock investing: Morningstar's guide to building wealth and winning in the market (1st ed.).* Wiley.

Gori, E., & Gori, G. (2018). Credit ratings: A new objective method using the Rasch model: The case of consumer discretionary. *International Journal of Business and Management Studies*, *07*(01), 53-84.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item Response Theory and Health Outcomes Measurement in the 21st Century. *Medical Care*, *38*(9), II28–II42. https://doi.org/10.1097/00005650-200009002-00007

Holdren, G. C. (1964). LIFO and Ratio Analysis. *The Accounting Review*, *39*(1), 70–85. JSTOR.

Horrigan, J. O. (1968). A Short History of Financial Ratio Analysis. *The Accounting Review*, *43*(2), 284–294. JSTOR.

Jambulingam, T., Schellhorn, C., & Sharma, R. (2016). Using a Rasch Model to Rank Big Pharmaceutical Firms by Financial Performance. *Journal of Commercial Biotechnology*, *22*(1), 49–60. https://doi.org/10.5912/jcb734

Koh, H. C., & Killough, L. N. (1990). The Use of Multiple Discriminant Analysis in the Assessment of the Going-Concern Status of an Audit Client. *Journal of Business Finance & Accounting*, *17*(2), 179–192. https://doi.org/10.1111/j.1468-5957.1990.tb00556.x

Linacre, J. M. (2002b). Understanding Rasch measurement: Optimizing rating scale category_effectiveness. Journal of Applied Measurement, 3, 85–106.

Linacre, J. M. (2002a). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), p.878. https://www.rasch.org/rmt/rmt162f.htm

Linacre J.M. (2005). Rasch dichotomous model vs. One-parameter Logistic Model. *Rasch Measurement Transactions,* 19:3, 1032. https://www.rasch.org/rmt/rmt193h.htm

Linacre, J. M. (2022). A user's guide to Winsteps and Ministeps Rasch-model computer programs program manual (Version 5.2.1) [Computer software]. Winsteps.com. https://www.winsteps.com/winsteps.html

Linacre, J. M., & Wright, B. D. (1993). *A user's guide to BIGSTEPS: Rasch-model computer program*. Mesa Press.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Addison-Wesley.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. https://doi.org/10.1007/BF02296272

Mead, R. (2008). *A Rasch Primer: The Measurement Theory of Georg Rasch*. Data Recognition Corporation.

MSCI. (2020). *Global industry classification standard (GICS®) methodology: Guiding Principles and Methodology for GICS*. MSCI Inc.

Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. *Nielsen & Lydiche.*

Raykov,T., & Marcoulides, G. A. (2011). Introduction to psychometric theory. Routledge. https://doi.org/10.4324/9780203841624

Ridzak, T. (2011). *Are Some Banks More Lenient in the Implementation of Placement Classification Rules?* Zagren: Croatian National Bank

Ross, S., Westerfield, R., Jaffe, J., & Jordan, B. (2018). *Fundamentals of Corporate Finance* (12th Edition). McGraw-Hill Education.

Schellhorn, C., & Sharma, R. (2013). Using the Rasch model to rank firms by managerial ability. *Managerial Finance, 39*(3), 306–319. https://doi.org/10.1108/03074351311302818

Shaw, F. (1991). Descriptive IRT vs. Prescriptive Rasch. *Rasch Measurement Transactions, 5*(1), 131. https://www.rasch.org/rmt/rmt51f.htm

Sick, J. (2010). Rasch measurement in language education Part 5: Assumptions and requirements of Rasch measurement. *SHIKEN: JALT Testing & Evaluation SIG Newsletter*, *14*(2). https://hosted.jalt.org/test/sic_5.htm

Song, M., & Zhong, H. (2020). Efficient weighted univariate clustering maps outstanding dysregulated genomic zones in human cancers. *Bioinformatics*, *36*(20), 5027–5036. https://doi.org/10.1093/bioinformatics/btaa613

Stemler, S. E., & Naples, A. (2021). Rasch measurement vs. item response theory: Knowing when to cross the line. *Practical Assessment, Research & Evaluation*, 26(11). Available online: https://scholarworks.umass.edu/pare/vol26/iss1/11 /

U.S. Securities and Exchange Commission (SEC) (2021). *How to Read a 10-K*. Retrieved February 21, 2022, from https://www.sec.gov/fast-answers/answersreada10khtm.html.

Wang, H., & Song, M. (2011). Ckmeans.1d.dp: Optimal k-means Clustering in One Dimension by Dynamic Programming. *The R Journal*, *3*(2), 29. https://doi.org/10.32614/RJ-2011-015

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410611697

WRDS Research Team (2016). WRDS Industry Financial Ratio.

Wright, B. D. (1977). Solving Measurement Problems with the Rasch Model. *Journal of Educational Measurement*, *14*(2), 97–116.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.

VITA

Chen Qiu

EDUCATION

M.S., Finance — May 2020

Gatton College of Business and Economics, University of Kentucky

Graduate Certificate, Applied Statistics — May 2020

College of Arts & Sciences, University of Kentucky

M.A., Social Science — May 2016

Morehead State University

B.S., Applied Psychology — Jun 2012

Tianjin Normal University

APPOINTMENTS

Research Assistant — Aug 2020 - Aug 2022

College of Education Evaluation Center, University of Kentucky

Research Assistant — Aug 2019 - Jun 2020

College of Health Sciences, University of Kentucky

Graduate Teaching Assistant (Applied Statistics) — Aug 2018 - May 2020

College of Education, University of Kentucky

Research Assistant — Sep 2017 - Aug 2019

Student and Academic Life, University of Kentucky

PUBLICATIONS

Lingat, J. M., Toland, M. D., **Qiu, C.,** Chen, X.-Y., Han, J., Shen, L., Ford, C. J., Love, A. M. A., Li, C., Blevins, J., Clement-Montgomery, T., White, C. J., Jeter, K., Kerhwald, N. P., & Williams, C. M. (2021). Optimizing the measurement of resident advisors' self-efficacy using the Rasch model. *Journal of College and University Student Housing.*

Dueber, D. M., Toland, M. D., Lingat, J. E., Love, A. M. A., **Qiu, C.,** Wu, R., Brown, A. V. (2021). To reverse item orientation or not to reverse item orientation, that is the question. *Assessment.* https://doi.org/10.1177/10731911211017635

AWARDS

Student Academic Travel Funding Award from College of Education (UKY)	2021
The Association of Test Publishers (ATP) Graduate Student Research Award	2021
Student Professional Development Award from College of Education (UKY)	2020
Student Academic Travel Funding Award from College of Education (UKY)	2019
Student Academic Travel Funding Award from College of Education (UKY)	2018
International Student Tuition Scholarship (UKY)	2018
The Dermontti F. and Regina M. Dawson Graduate Fellowship in Education (UKY)	2018