

LONGITUDINAL EXAMINATION OF FIRM-LEVEL SUPPLY CHAIN SUSTAINABILITY

By

Ming Li

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Business Administration — Logistics — Doctor of Philosophy

2022

ABSTRACT

LONGITUDINAL EXAMINATION OF FIRM-LEVEL SUPPLY CHAIN SUSTAINABILITY

By

Ming Li

Supply chain sustainability is a topic of immense importance for the press, political activists, managers, analysts, investors, shareholders, and stakeholders (e.g., local communities). One aspect of a firm's sustainability concerns the actions taken by members of its supply chain, such as the use of child labor and the dumping of toxic emissions. While there have been several attempts in measuring firms' sustainability as it pertains to their supply chains, these approaches suffer from numerous methodological weaknesses. This has limited our ability to answer important questions such as how firms' sustainability performance as it pertains to their supply chain evolved over time, and what factors affect this evolution? This question constitutes my primary research interest and has motivated my three-essay dissertation that:

Essay one: develops a new approach for measuring firm-level corporate social responsibility (CSR) and corporate social irresponsibility (CSI) supply chain practices using log-logistic item response theory models;

Essay two: studies how firm-level CSR and CSI supply chain practices have evolved over time using piecewise latent growth curve models; and

Essay three: examines the dynamic inter-relationships between firm-level CSR and CSI supply chain practices using dynamic panel models.

To answer these questions, I use panel data from 2003 – 2018 for hundreds of publicly-traded manufacturers, wholesalers, and retailers from KLD, which I merge with financial data from COMPUSTAT, market concentration from US Census Bureau Economic Indicators, and

Upstreamness measuring of the average distance from final use from American Economic Review. My results improve our theoretical understanding of how sustainable supply chain practices can be measured and how they evolve over time.

Keywords: supply chain sustainability, longitudinal examination, item response theory (IRT) model, KLD data, corporate social responsibility (CSR), corporate social irresponsibility (CSI), developing scales and measuring sustainability.

ACKNOWLEDGEMENTS

I would like to express my gratitude to everyone who helped me on this journey. First, my truly special thanks go to my advisors, Dr. Jason Miller and Dr. Yemisi Bolumole, as well as my mentor, Dr. Simone Peinkofer. From the moment when I met them at the doctoral program, they have always been supportive and thoughtful. I genuinely appreciate their guidance in helping me accomplish this memorable goal in my life. I am indebted to their expertise in this field in developing this dissertation. They are undoubtedly the greatest mentors, advisors, and role models in my academic voyage.

I also would like to thank my committee members: Dr. Sriram Narayanan and Dr. Steven Melnyk. Everyone on the committee provided me with invaluable academic insights in developing this dissertation. Dr. Melnyk's papers and workshop helped me develop the theoretical basis for the essays. I would like to give my special thanks to Dr. Narayanan. His support was instrumental in improving the dissertation, and I am also grateful for his advice which motivated me to become a better researcher. Dr. Narayanan's academic work ethic has encouraged me a lot to push myself harder.

Next, I must also thank all the faculty members at Michigan State University who helped me become an independent academic researcher. I am very grateful for all their advice and support during my doctoral study. Finally, I owe an enormous amount of gratitude to my family. Each of my family members has always supported me in pursuing this career. Without their support, I would not have been able to continue this journey.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER 1 - Measuring Firm-Level Supply Chain Sustainability Performance: A Longitudinal Multidimensional Item Response Theory Scaling Approach	1
1.1 Introduction	1
1.2 Measuring Sustainability	4
1.3 Item Response Theory (IRT) Models	9
1.3.1 A Brief Visit to IRT Models	9
1.3.2 The Connection between IRT Models and Factor Analysis Frameworks	11
1.3.3 Challenge of Dimensionality	12
1.3.4 Challenges in the Application of Standard IRT Models to Measure Sustainability	13
1.3.5 Present Research: Meeting the Challenges	14
1.4 Research Setting	16
1.4.1 Data descriptions	16
1.4.2 Sampling	17
1.4.3 Multidimensional Unipolar IRT Analytic Methods	18
1.5 Analysis and Results	19
1.5.1 Item Calibration (i.e., Item Parameter Estimates)	19
1.5.2 DIF Testing	22
1.5.3 Multidimensional IRT-scaled Scoring across All Items	32
1.5.4 How Firm Size Affects CSR and CSI	34
1.6 Discussion	38
1.6.1 Contributions	38
1.6.2 Limitations and Suggestions for Future Research	41
APPENDIX	42
REFERENCES	46
CHAPTER 2 - Longitudinal Examination of Firm-Level Supply Chain Sustainability	55
2.1 Introduction	55
2.2 Background Literature	60
2.2.1 Supply Chain Sustainability	60
2.2.2 Driving Factors of Firm-level Supply Chain Sustainability	61
2.2.3 Contributions to the Literature	63
2.3 Theory & Hypotheses Development	64
2.3.1 Longitudinal Development of Firm-level CSR in Supply Chain Practices	64
2.3.2 Longitudinal Development of Firm-level CSI in Supply Chain Practices	66
2.3.3 Moderation Effect of Firm Size on Rates of Change regarding CSR	68
2.3.4 Moderation Effect of Firm Size on Rates of Change regarding CSI	69
2.4 Research Setting and Data	71
2.4.1 Research Setting	71
2.4.2 Data Sources	72

2.4.3 Sample	72
2.4.4 Definition of Variables	73
2.5 Methods and Results	75
2.5.1 Methods	75
2.5.2 Results.....	77
2.6 Discussion.....	85
2.6.1 Theoretical and Empirical Contributions.....	85
2.6.2 Limitations and Suggestions for Further Research.....	88
REFERENCES	90

CHAPTER 3 - Developing and Testing the Dynamic Inter-relationships between Corporate

Social Responsibility and Irresponsibility Supply Chain Practices	101
3.1 Introduction.....	101
3.2 Background Literature	105
3.3 Theory and Hypotheses Development.....	108
3.3.1 Cross-lagged Effects between CSR and CSI	108
3.3.2 Autoregressive Effects of CSR and CSI	111
3.3.3 Moving Average Effects of CSR and CSI	113
3.4 Research Setting and Data	114
3.4.1 Research Setting	114
3.4.2 Data Sources	115
3.4.3 Measure Description	116
3.5 Methods and Results	117
3.5.1 Methods	117
3.5.2 Results.....	120
3.5.3 Cross-Validating VARMA(1,1) Model	124
3.6 Discussion.....	125
3.6.1 Theoretical and Empirical Contributions.....	125
3.6.2 Managerial Implications	128
3.6.3 Limitations and Suggestions for Future Research	131
REFERENCES	135

LIST OF TABLES

Table 1.1 Datasets and their Sources	17
Table 1.2 Variable Definitions and Summary Statistics.....	18
Table 1.3 Item Parameter Estimates and Standard Errors	22
Table 1.4 Candidate DIF Items Analyses Outputs between Manufacturers and Retailers	26
Table 1.5 Academic Attempts to Measure Firms' Sustainability Performance via Using KLD Rating Data	43
Table 2.1 Datasets and their Sources	72
Table 2.2 Sample Statistics and Correlation Matrix	73
Table 2.3 Metrics Coding for Measurement Occasions of CSR/CSI	74
Table 2.4 Results for Testing H1 – H2 for Firm-level CSR/CSI.....	78
Table 2.5 Results from Testing of Moderating Effects of <i>Size</i> on Firms' Rates of Change in Occasion_1 and Occasion_2 for CSR and CSI.....	83
Table 3.1 Datasets and their Sources	115
Table 3.2 Correlations, Means, and Standard Deviations for all Measures on all Occasions	117
Table 3.3 Results from Fitting the VARMA(1,1) Model to the Multivariate Time Series regarding the Measures of CSR and CSI	121

LIST OF FIGURES

Figure 1.1 Test Response Curves of CSR for Manufacturers and Retailers.....	24
Figure 1.2 Test Response Curves of CSI for Manufacturers and Retailers	24
Figure 1.3 Trace Lines for DIV_str_C by Industry	27
Figure 1.4 Trace Lines for EMP_str_D by Industry	27
Figure 1.5 Trace Lines for EMP_str_X by Industry	28
Figure 1.6 Trace Lines for EMP_str_G by Industry	28
Figure 1.7 Trace Lines for ENV_con_F by Industry	30
Figure 1.8 Trace Lines for EMP_con_X by Industry	30
Figure 1.9 Trace Lines for PRO_con_E by Industry	31
Figure 1.10 Trace Lines for EMP_con_B by Industry	31
Figure 1.11 Compare Log-logistic IRT Approach with Prior Summing-up Approach	34
Figure 1.12 Plot of the Moderating Effects of Size for Rate of Change regarding CSR	37
Figure 1.13 Plot of the Moderating Effects of Size for Rate of Change regarding CSI	37
Figure 2.1 Quasi-likelihood Estimated Means and Model-implied Means for Each Measurement Occasion for the Model with the Lowest -2 Log Likelihood for CSR	80
Figure 2.2 Quasi-likelihood Estimated Means and Model-implied Means for Each Measurement Occasion for the Model with the Lowest -2 Log Likelihood for CSI.....	80
Figure 2.3 Plot of the Moderating Effects of Size for Each Measurement Occasion for the Model with the Lowest -2 Log Likelihood for CSR	84
Figure 2.4 Plot of the Moderating Effects of Size for Each Measurement Occasion for the Model with the Lowest -2 Log Likelihood for CSI.....	84
Figure 3.1 Mechanisms that Propose the Relationship between CSR and CSI	109
Figure 3.2 Simplified Version of the VARMA(1,1) Model	119

Figure 3.3 Plot of CSR Implied by the VARMA Model Contingent on CSI at the First Measurement Occasion	123
Figure 3.4 Plot of CSI Implied by the VARMA Model Contingent on CSR at the First Measurement Occasion	124

CHAPTER 1 - Measuring Firm-Level Supply Chain Sustainability Performance: A Longitudinal Multidimensional Item Response Theory Scaling Approach

1.1 Introduction

Sustainability is growing increasingly important among the press, political activists, managers, analysts, investors, shareholders, and a company's other stakeholders, such as customers, employees, and local communities (Fahimnia, Sarkis, and Talluri 2019; Negri et al. 2021; Sarkis, Gonzalez-Torre, and Adenso-Diaz 2010). A company must manage the supply chain practices aligned with the sustainability expectations of its stakeholders to achieve long-term success, as failure to meet such expectations can result in various risks, such as consumers' negative perception of the firm, reputation and brand damage, labor disputes, market value reduce, or increased pressures from the community, government, and other social groups (Busse, Kach, and Bode 2016; Chowdhury and Quaddus 2021). Specifically, as firms' socially responsible supply chain practices are increasingly valued by stakeholders, sustainable investing has seen rapid growth and mutual funds that invest based on sustainability ratings are experiencing record inflows (Hartzmark and Sussman 2019). More and more investors expect companies to pay close attention to sustainability issues (Krueger, Sautner, and Starks 2020) and monitor socially irresponsible activities taken by members of their supply chain (Dyck, Lins, Roth, and Wagner 2019).

Meanwhile, rapid growth has been seen in academic studies that rely on sustainability ratings for empirical analysis (e.g., Albuquerque, Koskines, and Zhang 2019; Flammer 2015; Liang and Renneboog 2017; Lins, Servaes, and Tamayo 2017; Servaes and Tamayo 2013). As a result, sustainability ratings have been receiving increasing attention, and measuring firms' sustainability performance is undoubtedly an important topic (Berg, Koelbel, and Rigobon 2019).

Kinder, Lydenberg, Domini Research & Analytics (KLD) social and environmental index is one of the most well-known data sets to measure firms' sustainability performance (Berg, Koelbel, and Rigobon 2019; Fernandes and Bornia 2019; Ladygina 2021). In particular, KLD data has been regarded as "the largest multidimensional CSP [corporate social performance] database available to the public" (Deckop, Merriman, and Gupta 2006, p. 334). According to Chatterji, Levine, and Toffel (2009), the KLD database has been "the oldest and most influential and, by far, the most widely analyzed by academics" (p. 127). This occurs because KLD data has significant advantages in 1) rating on multiple attributes, 2) using objective measures, 3) having a large sample, and 4) emphasizing independent analysis (Graves and Waddock 1994; Hart and Sharfman 2015).

Although a tremendous amount of time and effort has been and continues to be devoted, gaps exist regarding how to aggregate the information in KLD data to measure sustainability. In particular, the existing measurement approaches in the literature have been inconsistent and often contradictory (Chatterji, Levine, and Toffel 2009; Eccles, Lee, Strohle 2020; Mattingly and Berman 2006). For instance, the most common approach is to sum or average the strengths and then subtract from the sum or average of weaknesses (e.g., Becchetti and Ciciretti 2011; Castillo et al. 2018; Sharfman 1996; Statman and Glushkov 2009). Such an approach has been prone to criticism for several unrealistic assumptions (Chatterji, Levine, and Toffel 2009). This is because sustainability inherently can not be observed directly, which makes it difficult to be measured (Carroll, Primo, and Richter 2016; Godfrey and Hill 1995). Considering this latent trait, the most recent studies have applied factor analysis to measures of firms' sustainability performance (e.g., Carroll, Primo, and Richter 2016; Nicolosi et al. 2014). However, these existing factor analysis approaches also suffer from several methodological weaknesses. For instance, these latent trait

approaches posit that there is a normal distribution for these latent traits, which is named bipolar trait assumption by Lucke's (2013; 2015). The assumption of bipolarity, however, creates several issues for measures, such as sustainability. This is because the absence of doing good things in a firm's supply chain doesn't imply that it would have done bad things. Similarly, the absence of negative issues doesn't imply improving positive ratings. For example, Nike has invested substantial resources in helping suppliers improve production processes yet has one or more suppliers engage in bad behavior (Distelhorst, Hainmueller, and Locke 2017). As such, the appropriateness of using standard latent trait approaches in prior studies with KLD data should be questioned.

This manuscript makes contributions in several ways. First, this research provides a new approach to measuring sustainability using state-of-the-art psychometric techniques and focusing on the set of sustainability metrics that pertains to the supply chain, which addresses measurement problems encountered with prior approaches. This is critical as the existing measurement approaches have been inconsistent and often contradictory (Chatterji, Levine, and Toffel 2009; Eccles, Lee, Strohle 2020; Mattingly and Berman 2006), which poses a major challenge for building cumulative knowledge (Shaver 2020) and hampers replication and extension (Pagell 2021). Specifically, this lack of valid measurement has handicapped the ability to answer important questions such as how has firms' sustainability performance as it pertains to their supply chain evolved over time, and what factors affect this evolution. Second, it applies a specific model, the log-logistic IRT model (Lucke 2013, 2015), which provides several improvements in the social and behavioral literature using categorical data (e.g., KLD data) that is characterized by very low base rates of the indicator and highly skewed total scores (Reise, Du, Wong, and Hubbard 2021). As the normality assumption of the underlying latent trait in the

population is violated, parameter estimates via the traditional approaches can be highly inaccurate (Kirischi, Hsu, and Yu 2001; Reise, Rodriguez, Spritzer, and Hays 2018; Wall, Park, and Moustaki 2015). The advantage of our approach is that it can be formulated to more realistically represent the underlying traits of specific types of unipolar constructs (Lucke 2013, 2015). Third, it evaluates the between-group differences in test functioning by comparing item parameters for the manufacturing industry and retailing industry. Finally, this study highlights the importance of separately calculating scores for firms' strengths through corporate social responsibility (CSR) initiatives and weaknesses through corporate social irresponsibility (CSI) incidents pertaining to supply chain sustainability.

This essay is structured in several sections. The next section provides a quick review of the existing approaches that have done scale development or measurement creation using KLD data in the literature. Section 3 provides a brief visit to item response theory (IRT) models, offers some reflective thoughts regarding the challenges of standard IRT approaches, and explains how the present research addresses the challenges. Section 4 describes our research setting, sample design, and our model formulation. Section 5 presents the results, and summarizes the implications of our findings. We conclude in section 6 by discussing the contributions and how this research can be extended.

1.2 Measuring Sustainability

A tremendous amount of time and effort has been and continues to be devoted to developing scales and measuring sustainability, especially with a wide array of existing archival data sources (Berg, Koelbel, and Rigobon 2019; Fernandes and Bornia 2019; Ladygina 2021). The KLD ESG ratings have been regarded as “the largest multidimensional CSP [corporate social performance] database available to the public” (Deckop, Merriman, and Gupta 2006, p. 334). Specifically, the

KLD data has been referred to as “the de facto [sustainability] research standard at the moment” (Waddock 2003, p. 369). According to Chatterji, Levine, and Toffel (2009), KLD’s social and environmental ratings have been “the oldest and most influential and, by far, the most widely analyzed by academics” (p. 127). This occurs because KLD datasets have significant advantages in 1) rating on multiple attributes, 2) using objective measures, 3) having a large sample, and 4) emphasizing independent analysis (Graves and Waddock 1994; Hart and Sharfman 2015).

Although “most of the academic literature to date relies on KLD data” (Berg, Koelbel, and Rigobon 2019, p. 8), there is no agreement regarding how to aggregate the information to measure firms’ sustainable performance. Some of the recent works in the supply chain literature have done scale development or measurement creation to frame up a picture. For example, Castillo et al. (2018) adopt a technique that aggregates a total score by subtracting the sum of weakness indicators from the sum of strength indicators. However, simply using such aggregate scores could discard valuable information, because an aggregated score of zero can be confusing for being reflective of either a firm that has an equal number of strength indicators and weakness indicators in a given year or a firm that simply was not rated on that indicators. To fix such an issue, Castillo et al. (2018) also set up two additional measures: total strength score (i.e., the sum of strength indicators), and total weakness score (i.e., the sum of weakness indicators). More approaches based on KLD data have emerged in the corporate social responsibility literature. To exemplify the diversity of those approaches, we provide a brief review of the methodology in Table 1.5 (see APPENDIX). For example, a diverse set of aggregate measures have been conducted by combining the strength and weakness scores to form an overall score as a research variable score (e.g., Becchetti and Ciciretti 2009; David, Bloom, and Hillman 2007; Graves and Waddock 1994; Ruf, Muralidhar, and Paul 1998; Sharfman 1996). One of the most common

approaches is to sum or average the strengths and then subtract from the sum or average of weaknesses (e.g., Becchetti and Ciciretti 2011; Sharfman 1996; Statman and Glushkov 2009). This simple approach of using KLD ratings has been prone to criticism for several unreasonable assumptions (Chatterji, Levine, and Toffel 2009). The first issue of this approach is that summing scoring assumes a parallel structure, which requires very strict constraints (McNeish and Wolf 2018). This is because the unstandardized loadings and error variances of a parallel structure are assumed identical across items (Graham 2006; McNeish and Wolf 2018). The second issue of this measurement is that it assumes the strength indicators are negatively related to weakness indicators. If the strengths and weaknesses are negatively correlated, the difference score of strengths and weaknesses will have limited variance¹. As a result, numerous assumptions warrant empirical testing (Kempf and Osthoff 2007; Nicolosi, Grassi, and Stanghellini 2014).

Some scholars attempt to be more theoretically grounded by combining the strength and weakness scores on the basis of weighting schemes (e.g., Graves and Waddock 1994; Ruf, Muralidhar, and Paul 1998). Then, the issue arises, such that is how to choose weights. For example, Ruf, Muralidhar, and Paul (1998) develop weights that take account of the preferences of various stakeholders, whereas Waddock and Graves (1997) build a weighting scheme that relies on experts' opinions. As such, the linear weighting schemes have been criticized for subjective choices of weights (Bird, Hall, Momentè, and Reggiani 2007). Moreover, the application of such weighting schemes can be challenged when the dataset has more than one period of time or when incorporating different datasets (Rowley and Berman 2000). Specifically,

¹ The variance of the difference of two correlated random variables can be shown as $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2 \times \text{Cov}[X,Y]$. Note if the covariance is negative, this means the difference score will have limited variance.

there is no way to verify the comparisons across dimensions (Eccles, Lee, Stroeble 2020; Mattingly and Berman 2006). Instead, a model-based weighting approach would be better.

The core challenge is that, unlike the objective characteristics in the physical sciences, sustainability cannot be observed directly, which makes it difficult to measure (Carroll, Primo, and Richter 2016; Godfrey and Hill 1995). As such, a few studies have performed an exploratory factor analysis (EFA) on the KLD indicators (e.g., Johnson and Greening 1999; Mattingly and Berman 2006; Waldman, Siegel, and Javidan 2006). However, the EFA approach applies factor analysis by assuming that the variables being factor analyzed are continuous (Browne 2001; Cudeck and Harring 2007). This is clearly not the case because all of the KLD indicators are binary. The EPA approach has to fit the model to the tetrachoric correlation matrix of the poly correlation matrix (Browne 2001), which has not been done regarding the existing studies that applied the EPA approach with KLD indicators. Another issue exposed by this stream of research is that none of the existing studies have considered that there is a huge problem with missing data.

Furthermore, it gets scholars to another approach, which is applying IRT models. Nicolosi, Grassi, and Stanghellini (2014) provide innovative insights into the measure based on the KLD data by adopting the polytomous item response models. Later, Carroll, Primo, and Richter (2016) conduct a new measure of the KLD indicators by using the Bayesian dynamic item-response model. A prominent advantage of these IRT-scaled approaches is the ability to create sustainability scores on the set of metrics across time periods and across various dimensions. Unfortunately, these studies have several weaknesses when applying the IRT approach. For instance, the challenge is that Nicolosi et al. (2014) and Carroll et al. (2016) applied a unidimensional IRT measurement model for more than 80 separate indicators to form

an overall score. This assumes a bifactor structure, which indicates that negative items load more strongly with each other, and the same with positive items. Such an assumption is against the essential aspects of the correlation between strengths and weaknesses. Specifically, the standard IRT approaches that are adopted by Nicolosi et al. (2014) and Carroll et al. (2016) assume that the underlying latent trait is unidimensionality, local independence, and monotonicity, such that the latent trait is normally distributed in the population (Reise, Rodriguez, Spritzer, and Hays 2018). However, KLD ESG ratings are characterized by very low base rates of the indicator and highly skewed total scores (Berg, Koelbel, and Rigobon 2019). As a result, the normality assumption of the underlying latent trait is violated. As such, parameter estimates via the standard IRT approaches can be highly inaccurate (Kirischi, Hsu, and Yu 2001; Reise, Rodriguez, Spritzer, and Hays 2018; Wall, Park, and Moustaki 2015).

To summarize, although a tremendous amount of time and effort has been devoted to measuring sustainability based on KLD data, there is no agreement regarding how one should aggregate the information to do scale development or measurement creation in existing studies (Eccles, Lee, Strohle 2020). In particular, the existing measurement approaches have been inconsistent and often contradictory (Chatterji, Levine, and Toffel 2009; Eccles, Lee, Strohle 2020; Mattingly and Berman 2006), which poses a major challenge for building cumulative knowledge (Shaver 2020) and hampers replication and extension (Pagell 2021). Specifically, this lack of valid measurement has handicapped the ability to answer important questions such as how has firms' sustainability performance as it pertains to their supply chain evolved over time, and what factors affect this evolution. Thus, there is a need to develop a credible way of measuring firm-level sustainability by using the KLD data.

1.3 Item Response Theory (IRT) Models

1.3.1 A Brief Visit to IRT Models

IRT frameworks² were specifically developed for categorical responses (Wirth and Edwards 2007). Our discussion here focuses on the two-parameter logistic (2PL) model, given that the 2PL model was created for dichotomously scored items and has been one of the most widely used forms of the IRT framework (Birnbaum 1968; Flora et al. 2008). This model can be expressed as follows:

$$P(y_{jti} = 1|\theta_{ti}) = \frac{1}{1 + e^{-Da_{jt}(\theta_{ti}-b_{jt})}} \quad (1)$$

where P is the probability of endorsing a given item, y_{jti} presents the observed response for item j at time t for subject i , θ_{ti} represents the latent construct hypothesized to be measured by the underlying trait of observed item response patterns, a_{jt} is the slope (or the discrimination parameter) for item j at time t , while b_{jt} is the intercept (or severity parameter) indicating how much of the latent construct subject i must possess to have a 50% probability of endorsing item j at time t , and D is a constant that scales the logistic model to that of the normal ogive. In fact, the a and b parameters are very similar to the factor loading and threshold parameters in traditional dichotomous factor analysis, as in many cases these are isomorphic (Flora et al. 2008; Takane and de Leeuw 1987).

For organizational researchers, IRT models hold methodological advantages relative to alternative methods (Nye et al. 2019; Foster, Min, and Zickar 2017). For example, the IRT framework provides a powerful alternative methodology to the most common approach, which is

² IRT settings are conceptually different from traditional factor analysis modeling frameworks, but they are closely related (Flora et al. 2008; Takane and de Leeuw 1987; Wirth and Edwards 2007). We will explain their connections in the following subsection.

simply summing up scorings (Flora et al. 2008). This methodology works particularly well when evaluating the quality of existing measures or developing new measures for assessing organizational constructs more effectively (Foster, Min, and Zickar 2017). The reason is that IRT models have unique advantages in scale construction over other techniques (Nye et al. 2019). For example, the IRT approach can effectively examine the quality of a scale regardless of the sample that is used to validate it. This is because the quality of items can be tested through the item parameters estimated from IRT by incorporating the item severity (i.e., how much of the latent construct a subject must possess to endorse the item correctly) and item discrimination (i.e., the strength of the relationship between the latent construct and the endorsed item). Specifically, the IRT framework assumes that the “parameters are invariant across different subpopulations of examinees, meaning that they can be readily compared after a linear transformation that puts them on a common scale” (Nye et al. 2019, p. 458). However, this condition can not be met by any traditional techniques based on classical test theory (CTT), because CTT approaches rely on alternative parameter estimations, such as estimating item severity through the proportion of subjects endorsing an item and estimating item discrimination through item-total correlations (Bauer and Hussong 2009). As a consequence, “a set of items in a measure can appear highly discriminating and include a range of difficulties (i.e., indicating a high-quality assessment) in one sample but then appear more or less discriminating and severity in another sample using CTT techniques” (Nye et al. 2019, p. 458).

Moreover, the IRT approach can be applied for detecting items (i.e., test bias) through differential item functioning (DIF), which is particularly useful for organizational research (Tay, Meade, and Cao 2015). Traditional techniques rely on mean differences to identify and test biases in a measure, however, observed differences across groups do not necessarily suggest test

bias. In other words, we cannot conclude that there is test bias in a measure because of the significant mean-level differences. The reason is that conclusions about observed differences can be confounded by any other bias under traditional approaches (Stark, Chernyshenko, and Drasgow 2004). In particular, “bias in a measure can inflate mean differences or even reverse the direction of these differences in organizational research” (Nye et al. 2019, p. 461). IRT techniques adopt DIF to identify the item or test bias, which is a more appropriate way in comparison with approaches depending on mean differences to test bias in a measure (Nye et al. 2019).

1.3.2 The Connection between IRT Models and Factor Analysis Frameworks

Although factor analysis and IRT are different modeling frameworks, they are closely related (Flora et al. 2008; Takane and de Leeuw 1987; Wirth and Edwards 2007). There's a strong relationship between discrimination parameters and slopes and then between the severity parameter and the intercept. Actually, Lord (1952) and Lord and Novick (1968) have demonstrated the analytic relationship between a one-factor categorical confirmatory factor analysis model and the unidimensional 2PL IRT model. Transforming the parameter estimates between these frameworks is pretty straightforward (Takane and de Leeuw 1987). We don't demonstrate these transformations in this study; instead, we do highlight that the underlying trait is assumed to be (1) continuous, and (2) of latent nature in both factor analysis models and IRT models (Raykov and Marcoulides 2011).

In addition, traditionally with factor analysis, the mean structure is always discarded, as there's no interest in that. However, the IRT setting highlights the role of unique variability in the relationship between the latent construct and the probability of endorsing an item, meaning that this approach can have items that all strongly discriminate on the scale and different severity. As

such, the IRT setting captures overall test information, which allows us to get information across the whole construct, whereas most factor analysis frameworks only care about the factor loadings or the discrimination parameters (Wirth and Edwards 2007). In particular, IRT models are specifically developed for dichotomous or categorical items (e.g. binary or binary scored items, or alternatively with Likert-type items), which is non-linear.

1.3.3 Challenge of Dimensionality

It is important to note that the challenge of dimensionality continues to pose when implementing the IRT approach, as parameter estimation typically requires integration (Raykov and Marcoulides 2011). Over the past four decades or so, a limited body of evidence has been accumulated that under some conditions IRT models can be somewhat robust to their assumptions, in particular, that of unidimensionality (Raykov and Marcoulides 2011). With serious violations of unidimensionality, IRT models can be misleading. Thus, it is readily realized that the actual interactions of examined subjects and multi-component measuring instruments (in particular, with their individual items/elements/components), are oftentimes involving more than one trait. As such, to accommodate the likely violations of unidimensionality, a researcher may hypothesize that studied subjects vary on more than merely a single trait (e.g., the one presumably evaluated by a used measuring instrument), with two or more constructs possibly responsible for their performance on subsets of the administered items. Hence, an extension of unidimensional IRT is needed, regarding providing a more adequate representation and modeling of the complexity underlying the used measuring instruments then. Such an extension is furnished by multidimensional IRT.

1.3.4 Challenges in the Application of Standard IRT Models to Measure Sustainability

Measurement models from IRT models have been applied with increasing frequency to social and behavioral research (Embretson and Reise 2000; Reise, Rodriguez, Spritzer, and Hays 2018). Standard IRT models assume that the underlying latent trait follows a density, not necessarily the standard normally distributed, but with the range of probability from negative infinity to positive infinity. Such an assumption has been named the continuous “bipolar trait” assumption by Lucke’s (2015, p. 273). This bipolarity is appropriate for “the traits of ability, achievement, or attitude for which everyone can be assigned a score, positive or negative, relative to an anchor at zero, representing the average level of the trait” (Lucke 2013, p. 199). However, this assumption of bipolarity can result in several problems for measures, such as sustainability. It makes no sense to assert that a firm has a below- or above-average level of socially responsible or irresponsible supply chain practices. This is because the absence of good things in a firm’s supply chain doesn’t imply that it would have bad things. Similarly, the absence of negative issues doesn’t imply positive ratings. For example, Nike has invested substantial resources in helping suppliers improve production processes yet has one or more suppliers engage in bad behavior (Distelhorst, Hainmueller, and Locke 2017). As such, the assumption of bipolarity is not appropriate for the latent construct that is applied to measure sustainability.

When the assumption of bipolarity is violated, estimation of IRT model parameters can be highly inaccurate (Kirischi, Hsu, and Yu 2001; Reise, Rodriguez, Spritzer, and Hays 2018; Wall, Park, and Moustaki 2015). For example, Woods and Thissen (2006) highlight the consequences of the misspecified normal distribution: “There is fairly consistent evidence that, when normality of $g(\theta)$ is assumed, MML estimates [by standard latent trait approaches] of more extreme item parameters (e.g., thresholds around $-(+) 2$) are nontrivially biased when the true

population distribution is platykurtic or skewed, and if is skewed, the bias increases as the skewness increases” (p. 283). Non-normal latent trait distributions posit particular challenges when applying standard latent trait approaches to measures in research that assuming normal distribution in a general population may not be tenable (Lucke 2013, 2015). In such cases, the appropriateness of applying standard latent trait approaches should be questioned, and an alternative latent trait approach should be developed to estimate the non-normal latent trait distributions (Reise et al. 2018).

1.3.5 Present Research: Meeting the Challenges

The objective of this essay is to develop an alternative IRT-related model to measure firms’ sustainability performance by using KLD data. We apply a unipolar IRT model (Lucke 2013, 2015), which allows us to address weaknesses of the existing approaches that attempted to measure sustainability via utilizing KLD data. This is because KLD indicators are characterized by low base rates of implementation and highly skewed total scores. As such, the assumption of underlying normal distributed latent trait scores is violated. Thus, the appropriateness of using standard latent trait approaches with KLD data should be questioned because the estimation of these model parameters can be highly inaccurate (Kirischi, Hsu, and Yu 2001; Reise et al. 2018; Wall, Park, and Moustaki 2015). Specifically, the interpretation of a level of sustainability via KLD indicators does not make sense. This is because the absence of doing good things in the supply chain doesn't imply that the firm has done bad things under the KLD indicators. Similarly, the absence of doing bad things doesn't imply doing good things within the firm’s supply chain. For example, Nike has invested substantial resources in helping suppliers improve production processes yet has one or more suppliers engage in bad behavior (Distelhorst,

Hainmueller, and Locke 2017). In other words, the underlying latent trait of KLD indicators involves unipolar constructs rather than bipolar constructs (Lucke 2013, 2015).

As such, this essay applies an alternative log-logistic IRT model, also named as unipolar IRT model by Lucke's (2015), to essentially deal with (1) postulating models about the relationship between the probability of adopting a set of indicators and the latent traits (i.e., θ); (2) estimating the models using the KLD dataset; (3) evaluating the (relative) fit of the models utilized and carrying out model choice; and (4) based on all that, after a model is selected, estimating individual subject values for θ , which are positioned on the same 'scale' (underlying dimension).

In addition, we adopt a multidimensional construct to more realistically represent the underlying constructs that explain firms' response to KLD indicators. This is because an underlying multidimensional construct enables a vector of traits to explain the interaction of examined subjects with a given set of items (measuring instrument) (Bonify, Reise, Scheines, and Meijer 2015). This emphasizes that the multidimensional construct can be used to model complex reality, which allows us to access useful and more adequate models of complex phenomena that cannot be obtained within the framework of unidimensional constructs (Raykov and Marcoulides 2011; Reckase 2009).

Summing it up, based on KLD data, this essay focuses on developing a whole new set of sustainability metrics by using a flexible multidimensional unipolar IRT (Lucke 2013, 2015) model to calculate scores for firms' CSR and CRI separately, which addresses measurement problems encountered with prior approaches. This is because positive and negative social behaviors require separate measurements (Strike, Gao, and Bansal 2006). Specifically, our approach is consistent with calls to look at a firm's ratings in strengths through CSR and

weaknesses through CSI independently, because combining positive and negative social practices in empirical research could obscure the countervailing effects of each indicator on the integrated total score (Mattingly and Berman 2006).

1.4 Research Setting

1.4.1 Data descriptions

KLD socially responsible ratings provide the longest time series of firms' environmental and social sustainability information (Chatterji, Levine, and Toffel 2009). KLD ratings for measuring firms' environmental, social, and governance performance were assigned based on the company's CSR reports and other relevant public information, which is released yearly (KLD 2018). Corporate behaviors are rated across seven dimensions: governance, community, diversity, employee relations, environment, human rights, and product quality. For each dimension, KLD ratings consist of paired items. Each such a paired item has both a strength and concern indicator, which is binary taking values 0 or 1. A score of 1 in a strength indicator indicates that the firm has a positive behavior in complying with the social responsibility standards, whereas a score of 1 in a concern indicator indicates the firm has a negative activity (i.e., social irresponsibility practice) that can be considered as a weakness to meet the standards of social responsibility.

Between 1991 and 2000, KLD ratings focused on the largest 650 publicly-traded companies that belonged to the Domini 400 Social Index³ and/or to the S&P500 Index. In 2001, KLD ratings extended the coverage to consist of the top 1000 publicly-traded US companies by market capitalization. Since 2003, the coverage of KLD ratings expanded to include the top 3000 publicly-traded US companies by market capitalization, and this coverage went on to include the

top 4,000 later on. Considering that, limited companies have been covered before 2003. We focus on the panel data from 2003 to 2018 for hundreds of publicly traded manufacturing, wholesale, and retailing firms included in KLD ratings. Once we obtained the KLD data, we merged it with financial data from COMPUSTAT, and market concentration from US Census Bureau Economic Indicators. To provide a better view of the datasets used in this manuscript, Table 1.1 provides all the datasets, using variables from them, and their sources.

Table 1.1 Datasets and their Sources

Data	Using Variables	Source
KLD data	ESG performance indicators	MSCI ESG KLD 2003-2018
Financial data	the current value of investment, cost of income, cost of investment, revenue, cost of goods sold, total assets, R&D expenditure, capital expenses	Standard & Poor's Compustat 2003-2018
Market concentration	HHI	US Census Bureau Economic Indicators 2003-2018

1.4.2 Sampling

KLD data has many missing scores in their indicators (Chatterji, Levine, and Toffel 2009; Chen and Ho 2019). We set up a few filters to handle the missing scoring issues. Our first filter is to eliminate the indicators that have more than 40% missing scores, because the literature suggests that a variable with more than 40% missing data can introduce serious bias which could potentially invalidate the entire analysis (Pritikin et al. 2018). Only 28 indicators have no more than 40% missing data. Thus, we limit ourselves to these 28 indicators. Second, we checked these indicators' definitions year over year. Then, we selected the indicators with the most consistent definitions from 2003 to 2018. This process reduced the universe of measures to 10 indicators in strengths and 10 indicators in weaknesses. Table 1.2 is a summary of these indicators left after a few filters have been implied.

Table 1.2 Variable Definitions and Summary Statistics

Variable	Definition	Obs.	Mean	St.Dev.
ENV_str_B	Pollution & waste – toxic emissions and waste	14,646	0.04	0.20
ENV_str_D	Climate change - carbon emissions	14,803	0.10	0.30
ENV_str_X	Superior commitment to management systems, voluntary programs, or other proactive activities	10,233	0.05	0.21
EMP_str_C	Cash profit sharing	10,142	0.08	0.28
EMP_str_D	Employee involvement	11,769	0.10	0.30
EMP_str_G	Employee health & safety	14,140	0.05	0.23
EMP_str_X	Human capital - is designed to capture best-in-class management performance the area of human capital	10,940	0.07	0.25
DIV_str_B	Representation - at least one woman among the executive management team	9,249	0.28	0.45
DIV_str_C	Board diversity - strong gender diversity on their board of directors	13,423	0.09	0.29
PRO_str_A	Product safety and quality	12,116	0.10	0.30
ENV_con_B	Recently paid substantial fines or civil penalties for violations of major environmental regulations	11,537	0.05	0.21
ENV_con_D	Toxic emissions and waste - controversies related to a firm's operational non-ghg emissions	14,288	0.05	0.21
ENV_con_F	Energy & climate change - controversies related to climate change and energy-related impacts	15,993	0.02	0.14
EMP_con_B	Health & safety - controversies related to the health and safety of a firm's employees	15,991	0.08	0.26
EMP_con_X	Labor rights & supply chain – other concerns	15,992	0.04	0.18
DIV_con_A	Discrimination & workforce diversity - controversies related to a firm's workforce diversity	15,993	0.03	0.16
COM_con_B	Impact on community - controversies related to a firm's interactions with related communities	15,993	0.03	0.16
PRO_con_A	Product quality & safety - controversies due to the quality/safety of a firm's products/services	15,088	0.07	0.25
PRO_con_D	Marketing & advertising - controversies related to a firm's marketing and advertising practices	15,646	0.05	0.21
PRO_con_E	Anticompetitive practices - controversies related to a firm's anti-competitive business practices	15,993	0.04	0.19

1.4.3 Multidimensional Unipolar IRT Analytic Methods

As we showed in the literature review, the existing approaches have numerous weaknesses and assume a tremendous amount of things, such as a parallel structure and negative relationship between strength and weakness indicators assumptions (e.g., Deckop, Merriman, and Gupta 2006; Mattingly and Berman 2006; Chatterji, Levine, and Toffel 2009), and the unidimensional assumption (Carroll et al. 2016). We expect to get out of these unrealistic assumptions by applying a multidimensional construct to access a useful and more adequate model of complex

phenomena that cannot be obtained within the framework of unidimensionality (Bonify, Reise, Scheines, and Meijer 2015; Reckase 2009). As such, we used flexMIRT software to conduct our item analysis and test scoring, because flexMIRT is the “most advanced IRT software available” (Houts and Cai 2020). As such, flexMIRT allows us to efficiently estimate high-dimensional models. Specifically, flexMIRT software is powerful in handling missing scorings due to its rich psychometric and statistical features (Cai and Monroe 2013; Cai and Houts 2019; Houts and Cai 2020).

1.5 Analysis and Results

Our analyses proceeded according to three stages: item calibration (i.e., parameter estimation), Differential item functioning (DIF) testing, and subsequent scale scoring across all items.

1.5.1 Item Calibration (i.e., Item Parameter Estimates)

The existing literature indicates that positive and negative social behaviors require separate measurements (Strike, Gao, and Bansal 2006), because combining “strength” and “concern” indicators in empirical research can result in obscuring the countervailing effects of each indicator on the integrated total score (Mattingly and Berman 2006). Moreover, we explored the structure of the 20-item set of interest in this study and allowed the latent factors to correlate (Reckase 2009). For each calibration model, we computed several model-data fit indices: simulated loglikelihood of the fitted model, the root mean squared error of approximation (RMSEA), AIC, and BIC. All model-data fit indices were computed in each replication using the estimated item and parameter. A scree test suggested that there were two dominant (or 2-trait) factors. The RMSEA statistic for the 2-factor model was .053, further suggesting a reasonable fit of a 2-factor model (e.g., Browne and Cudeck 1992). The eigenvalues of the tetrachoric correlation matrix of the 20 items in this study, also suggested 2 traits rather than 1 or 3 as

underlying the examined firms' sustainability performance on the 20 items. BIC and AIC are much smaller for this 2-trait model than for the single-trait model, suggesting the superiority of the 2-trait in a model comparison sense (Raykov and Marcoulides 2008). Taken together, the empirical evidence strongly suggests that there are two dimensions underlying the 20 items.

After confirming the dimensionality of the set of dichotomous items, we next fitted a confirmatory two-factor model to estimate firms' CSR through strength indicators and CSI through concern indicators. In this stage, we estimate the discrimination and severity parameters for each item by implementing a multidimensional IRT model, which is the phase of the analysis named item calibration (Flora et al. 2008). The pattern of the estimated loadings suggests that the items were assigned to the factors as we expected, with certain loadings restricted to 0.00 and had no associated SE. The correlation matrix of the latent variable (i.e., the correlation between the two standardized factors) is estimated to be 0.57, with a SE of 0.01. As such, the findings of such a correlation matrix indicate that the prior approaches that integrate ratings across strengths and weaknesses can be inaccurate (Mattingly and Berman 2006). This is because these two series are measuring different things. For instance, if we use the standard summing of the strength indicators to subtract the summing of the weakness indicators, we won't get any good points since these two series are not measuring the same thing because of the correlation of 0.57. The following table (i.e., Table 1.3) reports the estimated item parameters. The results suggest clear differences in both the severity and the discrimination parameters across the set of items. CSR (i.e., θ_1) captures the underlying latent trait of all the strength indicators, whereas CSI (i.e., θ_2) shows the underlying latent trait of all the weakness indicators. According to Curran et al. (2008), "[t]his alone is a substantial improvement over the proportion score method of scoring in which the relations of all items of the underlying construct are treated identically" (p. 371).

For example, under CSR, the item ENV_str_D “strengths in climate change”, shows the highest discrimination ($a = 3.82$), whereas the item DIV_str_B “strengths in representation of the board regarding gender diversity” shows the lowest discrimination ($a = 0.70$). This implies the indicator ENV_str_D “strengths in climate change” is more strongly reflective of latent strengths trait compared with the indicator DIV_str_B “strengths in representation of the board regarding gender diversity”. This demonstrates that the former is more likely to discriminate companies' performance in strength than the latter. This is relatively intuitive, because taking efforts to improve environmental performance costs much more in comparison with adopting gender diversity on the board of directors. Further, the item EMP_str_D “strengths in employee involvement” shows the highest severity, whereas the item DIV_str_B “strengths in representation of the board regarding gender diversity”, shows the lowest severity. This indicates that the item DIV_str_B “strengths in representation of the board regarding gender diversity” has the greatest probability of being reported, whereas the item ENV_str_B “strengths in employee involvement” has the lowest probability of being reported.

For comparison, under CSI, the item COM_con_B “weaknesses in impact on community regarding controversies” shows the highest discrimination ($a = 2.65$), whereas the item ENV_con_F “weaknesses in climate change and energy-related impacts” displays the lowest discrimination ($a = 1.17$). This implies the indicator the former is more strongly reflective of latent construct relative to the latter, which demonstrates that the former is more likely to discriminate companies' performance in weaknesses than the latter. Further, the item ENV_con_F “weaknesses in climate change and energy-related impacts” shows the highest severity, whereas the item EMP_con_B “weaknesses in employees' health and safety” shows the lowest severity. This indicates that the item ENV_con_F “weaknesses in climate change and

energy-related impacts” has the greatest probability of being reported, whereas the item EMP_con_B “weaknesses in employees’ health and safety” has the lowest probability of being reported.

Table 1.3 Item Parameter Estimates and Standard Errors

Variables	θ_1 regarding CSR				θ_2 regarding CSI			
	Discrimination		Severity		Discrimination		Severity	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
ENV_str_B	2.88***	0.11	2.01***	0.04	0.00	----	0.00	----
ENV_str_D	3.82***	0.13	1.43***	0.02	0.00	----	0.00	----
ENV_str_X	2.45***	0.10	1.87***	0.05	0.00	----	0.00	----
EMP_str_C	1.07***	0.05	2.52***	0.10	0.00	----	0.00	----
EMP_str_D	0.87***	0.04	2.68***	0.11	0.00	----	0.00	----
EMP_str_G	2.05***	0.07	2.25 ***	0.06	0.00	----	0.00	----
EMP_str_X	2.29***	0.09	1.76***	0.04	0.00	----	0.00	----
DIV_str_B	0.70***	0.03	1.25***	0.06	0.00	----	0.00	----
DIV_str_C	1.12***	0.04	2.57***	0.09	0.00	----	0.00	----
PRO_str_A	1.44***	0.05	1.85***	0.05	0.00	----	0.00	----
ENV_con_B	0.00	----	0.00	----	2.12***	0.09	2.20***	0.05
ENV_con_D	0.00	----	0.00	----	2.14***	0.08	2.14***	0.05
ENV_con_F	0.00	----	0.00	----	1.14 ***	0.07	3.60 ***	0.17
COM_con_B	0.00	----	0.00	----	2.65 ***	0.11	2.31***	0.05
EMP_con_B	0.00	----	0.00	----	1.76***	0.05	1.95***	0.04
EMP_con_X	0.00	----	0.00	----	1.27***	0.06	3.03***	0.12
DIV_con_A	0.00	----	0.00	----	1.86***	0.08	2.63***	0.08
PRO_con_A	0.00	----	0.00	----	1.74***	0.06	2.20***	0.06
PRO_con_D	0.00	----	0.00	----	1.68***	0.06	2.47***	0.07
PRO_con_E	0.00	----	0.00	----	1.84***	0.07	2.61***	0.08
Factor	Factor Correlations							
θ_1	1							
θ_2	0.57							

Notes: * = $p < 0.10$; ** = $p < 0.05$; *** = $p < 0.01$ (two-tailed).

Item parameter estimates and standard errors are estimated by a confirmatory multidimensional IRT model with correlated Factors via flexMIRT.

1.5.2 DIF Testing

Another advantage of the IRT approach is that “it provides an elegant framework for the evaluation of whether an item or test is measuring the same construct in the same way for two or more different groups” (Reise and Rodriguez 2016, p. 2034). A simple way to get a sense of the

DIF test is to check whether the test response curves (TRC)³ are equivalent when the item parameters are estimated separately for two or more groups (Reise and Rodriguez 2016). In our case, Figures 1.1 & 1.2 show the TRCs of CSR and CSI when the item parameters are estimated separately for manufacturers and retailers. As can be seen, it seems there are no clear differences in *CSR* between manufacturers and retailers, while there are clear differences in *CSI* between these two industries. We focus on manufacturers and retailers for DIF testing, because the number of wholesalers is far smaller than manufacturers and retailers. As a result, the small number of wholesalers in the sample size is less than the advised size for DIF testing (Cai 2013).

³ A TRC is the sum of the item characteristic curves (ICCs), which reflects how the latent trait is related to the expected total test score (Reise and Rodriguez 2016).

Figure 1.1 Test Response Curves of CSR for Manufacturers and Retailers

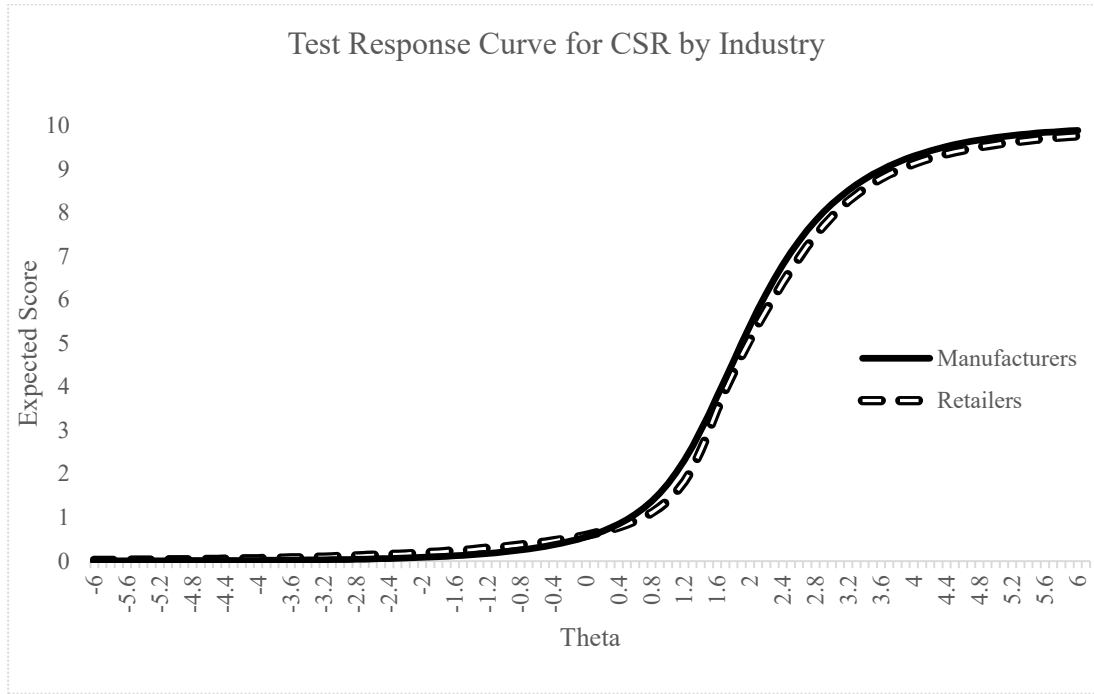
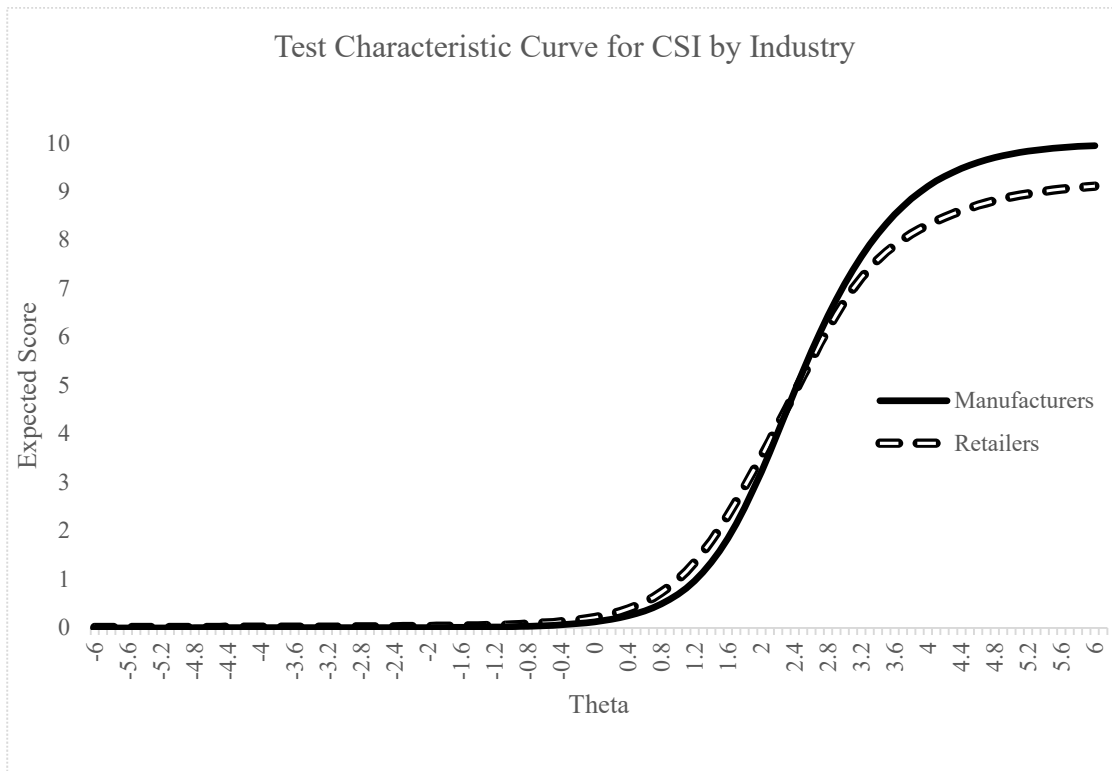


Figure 1.2 Test Response Curves of CSI for Manufacturers and Retailers



Furthermore, we apply Thissen, Steinberg, and Wainer's (1993) likelihood ratio method for testing DIF between manufacturers and retailers, given that "this approach has good statistic power and Type I error control" (Flora et al. 2008, p. 685). This DIF testing conducts by comparing the fit of a model where the discrimination and severity parameters are free across groups with the fit of a model where these parameters are constrained to be equal across groups. As such, the likelihood ratio statistic can be calculated by

$$\chi^2(d.f.) = -2(ll_{Model\ 1} - ll_{Model\ 2}). \quad (2)$$

Where $ll_{Model\ 1}$ is the log-likelihood value of the model with item parameters constrained to be equal across manufacturers and retailers and $ll_{Model\ 2}$ is the log-likelihood value of the model with item parameters set to be free across manufacturers and retailers.

There are two general types of DIF testing via flexMIRT: (a) a DIF sweep for all items, and (b) a more focused examination by testing candidate DIF items with designated anchor items (Houts and Cai 2020). First, we conduct the general DIF sweep to test all items across manufacturers and retailers, which allows us to identify the items that are obviously different across groups. This enables us to do a more focused examination. Second, we thus estimate candidate DIF items with designated anchor items. To do this, we actually free estimate every one of the items that have χ^2 differences greater than 20 across the two groups in the general DIF sweep for all items, and constrain the other items to be equal across the groups. The advantage of this approach is the blatant indicators (i.e., the items that are obviously different across groups) are still on the same scale because we have some common indicators (i.e., the anchor items) that are invariant. We report the likelihood ratio statistic for candidate DIF items testing between groups of manufacturers and retailers in Table 1.4.

Table 1.4 Candidate DIF Items Analyses Outputs between Manufacturers and Retailers

Variables	Total			Discrimination			Severity		
	χ^2	d. f.	p	χ^2	d. f.	p	χ^2	d. f.	p
EMP_str_C	35.8	2	0.000	1.3	1	0.208	34.5	1	0.000
EMP_str_D	44.8	2	0.000	44.4	1	0.000	0.4	1	0.523
EMP_str_G	31.5	2	0.000	3.3	1	0.070	28.2	1	0.000
EMP_str_X	58.0	2	0.000	0.3	1	0.575	57.7	1	0.000
DIV_str_B	69.0	2	0.000	4.3	1	0.038	64.7	1	0.000
DIV_str_C	328.0	2	0.000	64.5	1	0.000	263.5	1	0.000
PRO_str_A	30.5	2	0.000	20.4	1	0.001	10.0	1	0.000
ENV_con_B	49.0	2	0.000	2.2	1	0.143	46.9	1	0.000
ENV_con_D	79.6	2	0.000	0.7	1	0.388	78.9	1	0.000
ENV_con_F	39.5	2	0.000	27.3	1	0.000	12.1	1	0.001
EMP_con_B	19.1	2	0.000	5.1	1	0.024	14.0	1	0.000
EMP_con_X	92.1	2	0.000	3.5	1	0.063	88.7	1	0.000
PRO_con_A	44.9	2	0.000	11.1	1	0.001	338	1	0.000
PRO_con_E	25.5	2	0.000	1.2	1	0.272	24.3	1	0.000

Notes: Specifying this candidate DIF items testing, we are looking at variables listed in the table for possible DIF between groups of manufacturers and retailers. To do this, we set the anchor items (ENV_str_B, ENV_str_D, ENV_str_X, COM_con_B, DIV_con_A, PRO_con_D) equal across groups, and free estimate every one of the items listed in the table.

Given that our sample size is quite large, we have incredible power. As such, some CSR indicators are statistically significant and practically significant, while some are statistically significant but not practically significant. We use some trace lines to show the comparisons. For example, DIV_str_C, “strengths in board diversity” and EMP_str_D, “strengths in employee involvement” are both statistically significant and practically significant (see Figure 1.3 and Figure 1.4). However, EMP_str_X “strengths in employee health & safety” and EMP_str_G “strengths in human capital” are not practically significant (see Figure 1.5 and Figure 1.6). For discussion, we focus on those items that are practically significant. As expected of DIV_str_C “strengths in board diversity”, there is a stronger relationship between this indicator and the underlying latent trait among manufacturers than among retailers. In contrast, EMP_str_D “strengths in employee involvement” indicates employee involvement has a stronger relationship with the underlying latent trait among retailers than among manufacturers.

Figure 1.3 Trace Lines for DIV_str_C by Industry

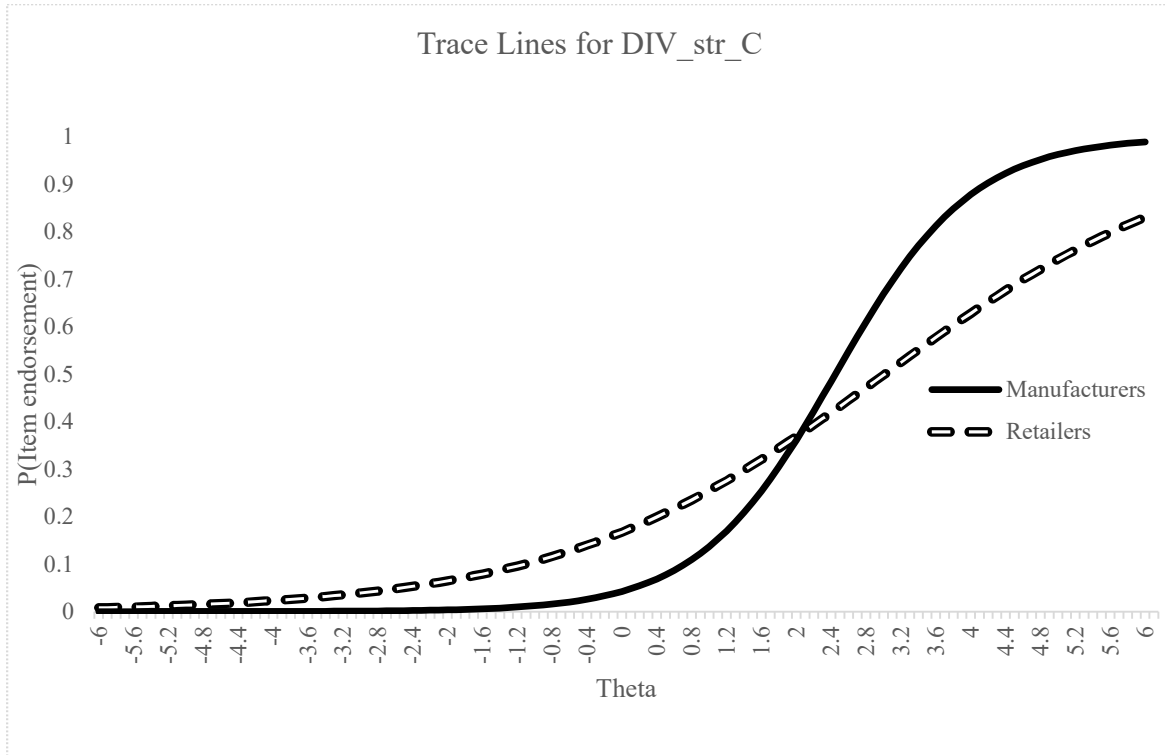


Figure 1.4 Trace Lines for EMP_str_D by Industry

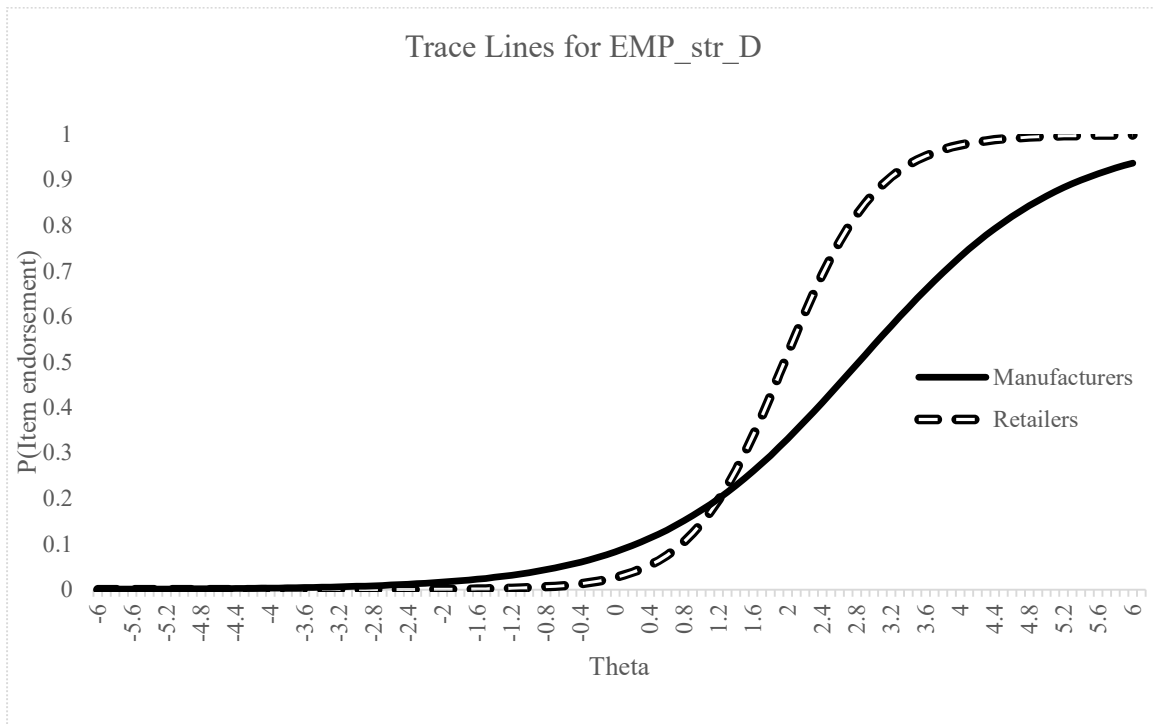


Figure 1.5 Trace Lines for EMP_str_X by Industry

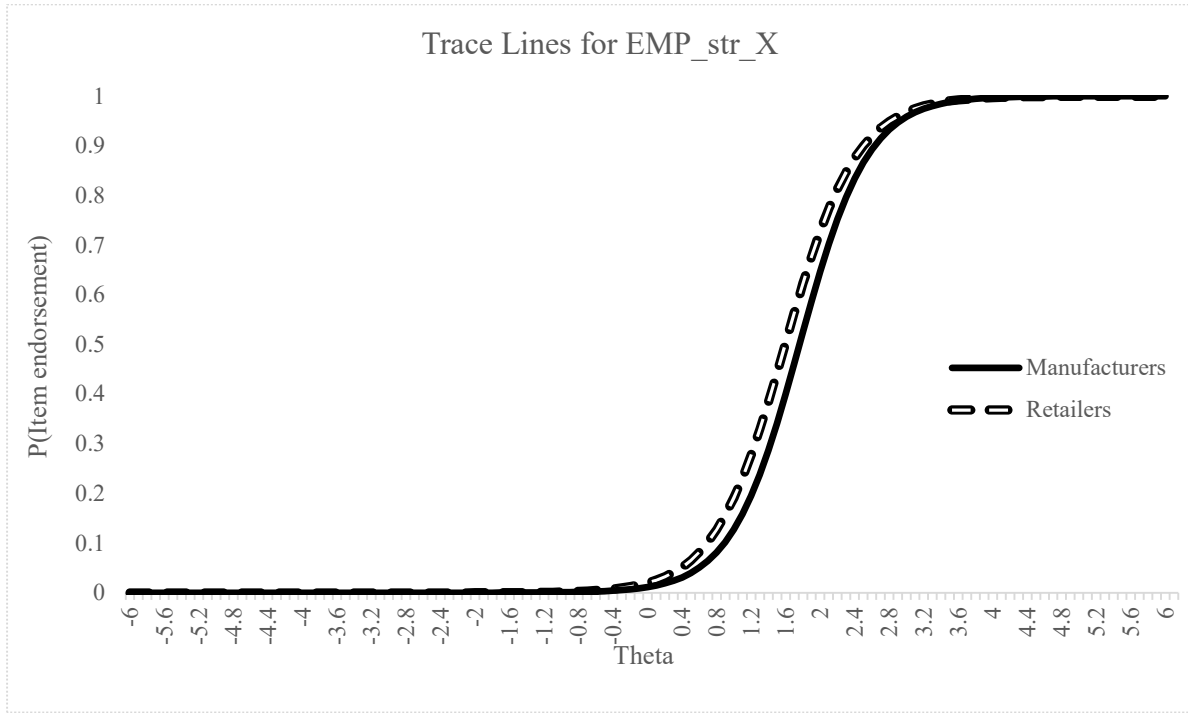
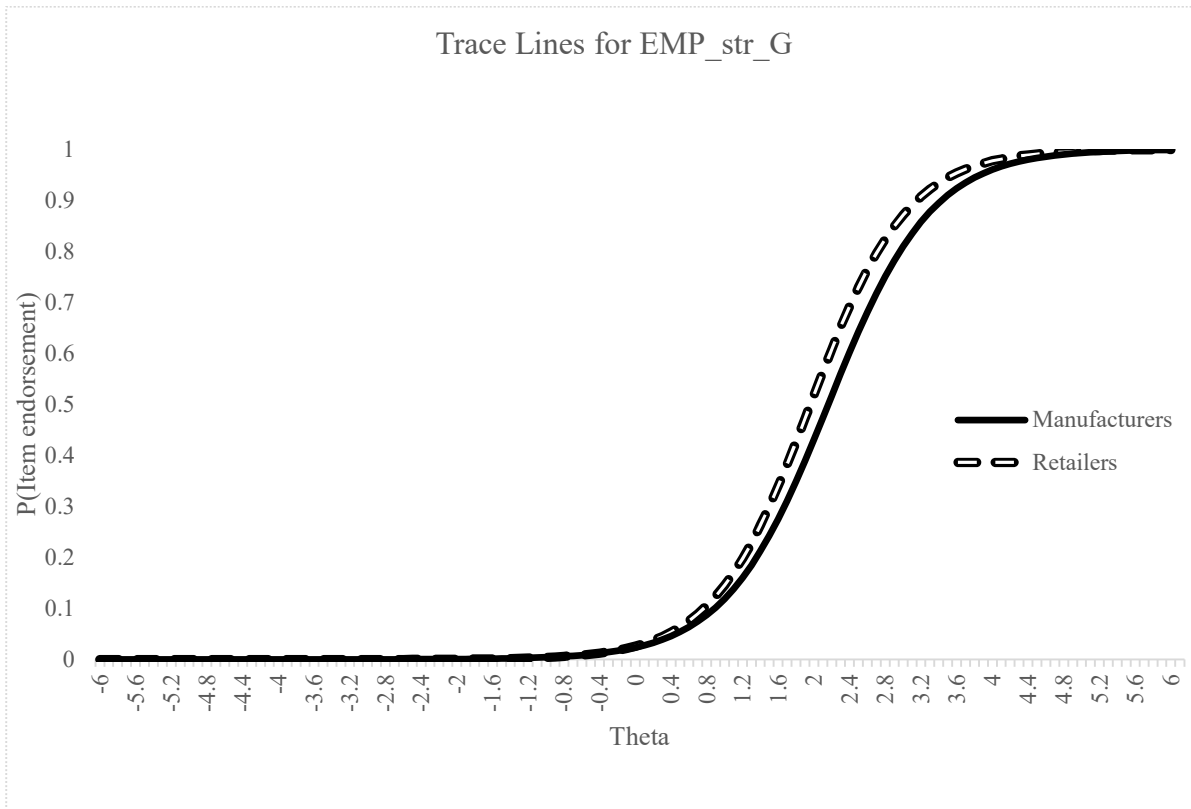


Figure 1.6 Trace Lines for EMP_str_G by Industry



Similarly, some CSI indicators are statistically significant and practically significant, while some are statistically significant but not practically significant. We use some trace lines to show the comparisons. For example, ENV_con_F “weaknesses in energy & climate change” and EMP_con_X “weaknesses in labor rights & supply chain” are both statistically significant and practically significant (see Figure 1.7 and Figure 1.8). However, PRO_con_E “weaknesses in anti-competitive practices” and EMP_con_B “weaknesses in employee health & safety” are not practically significant (see Figure 1.9 and Figure 1.10). For discussion, we focus on those items that are practically significant. With the expectation of CSI indicators, a stronger relation has been exposed between weaknesses in energy & climate change and the underlying construct of CSI among manufacturers than among retailers, whereas weaknesses in labor rights and supply chain have marked a stronger relationship with the underlying internalizing construct of CSI among retailers than among manufacturers.

Figure 1.7 Trace lines for ENV_con_F by industry

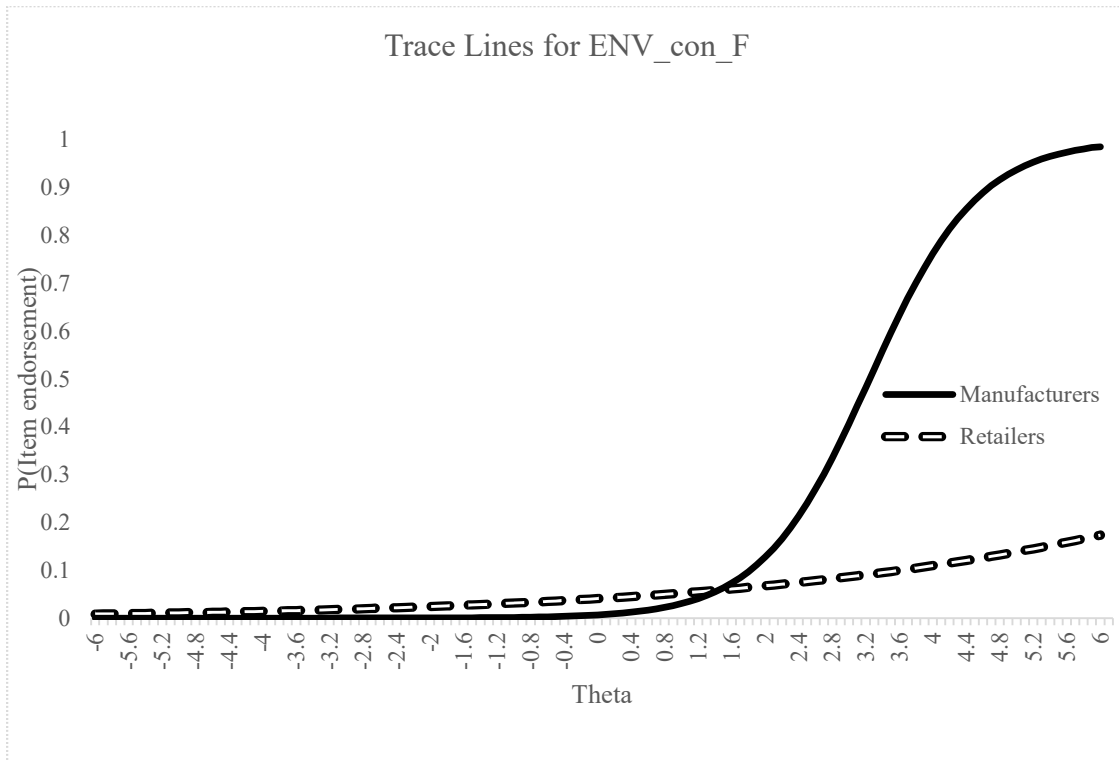


Figure 1.8 Trace lines for EMP_con_X by industry

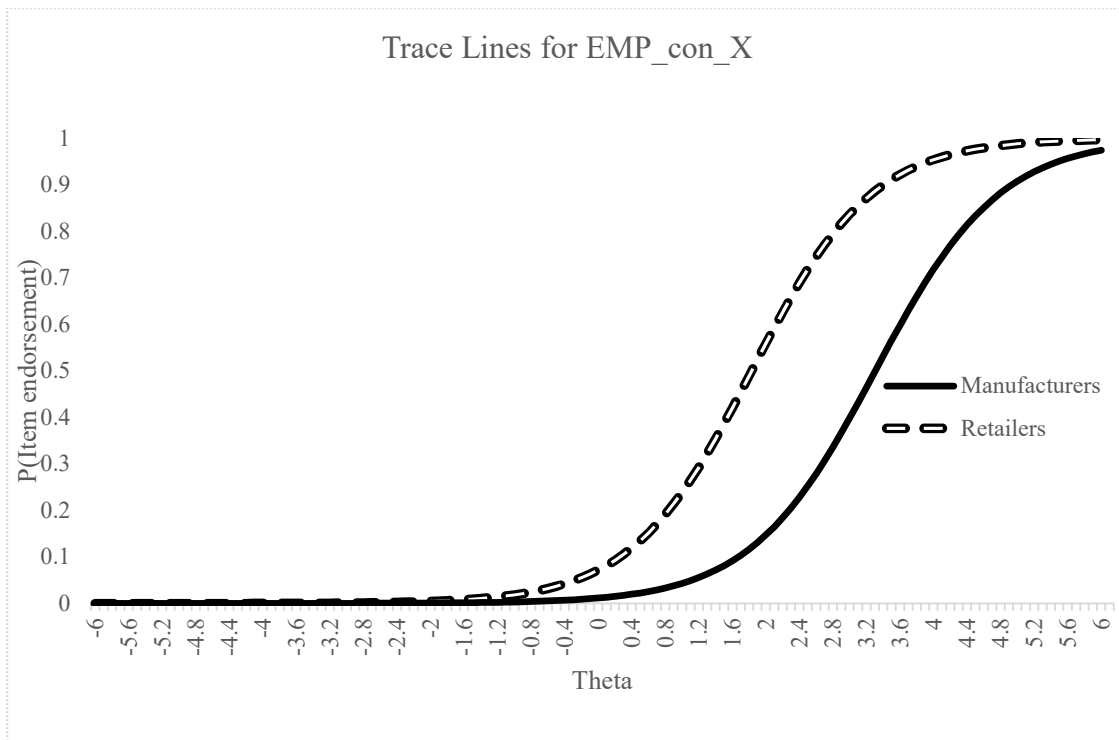


Figure 1.9 Trace lines for PRO_con_E by industry

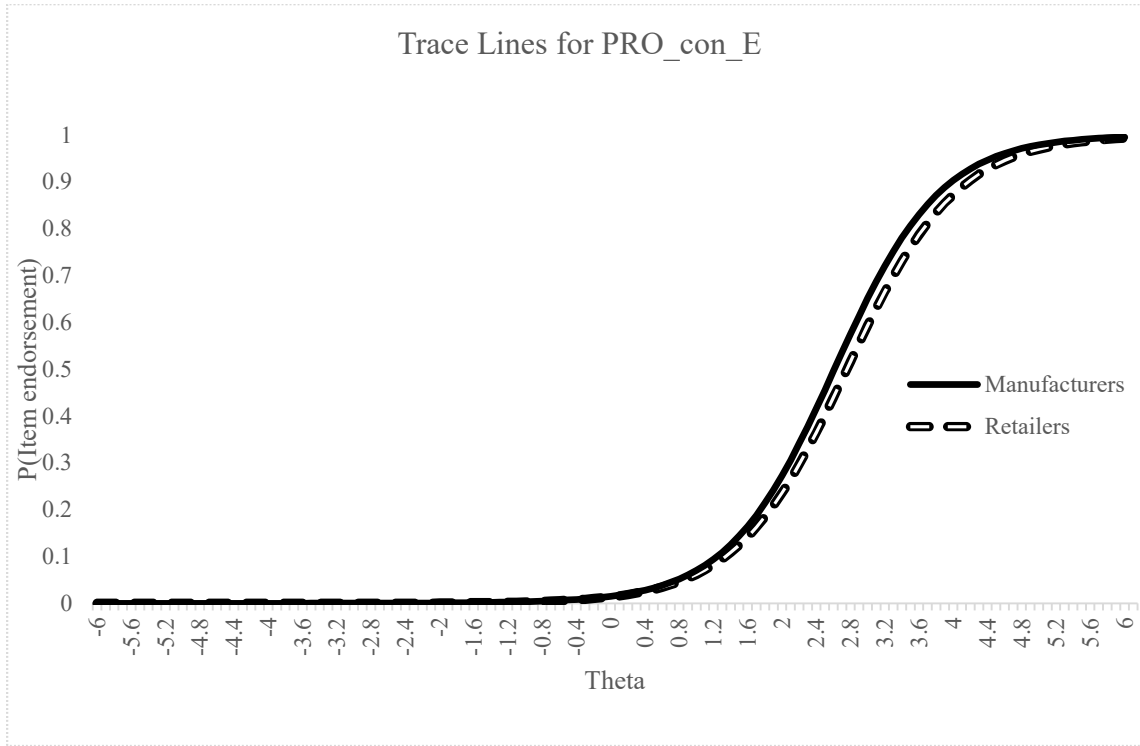
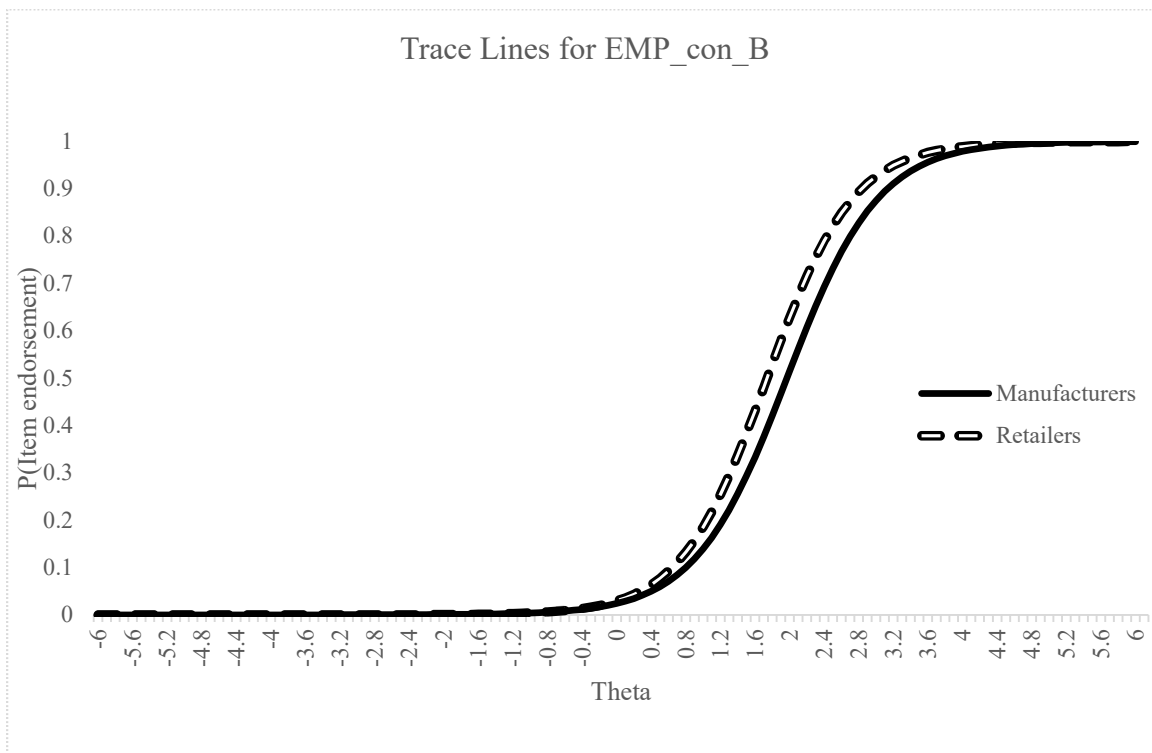


Figure 1.10 Trace lines for EMP_con_B by industry



1.5.3 Multidimensional IRT-scaled Scoring across All Items

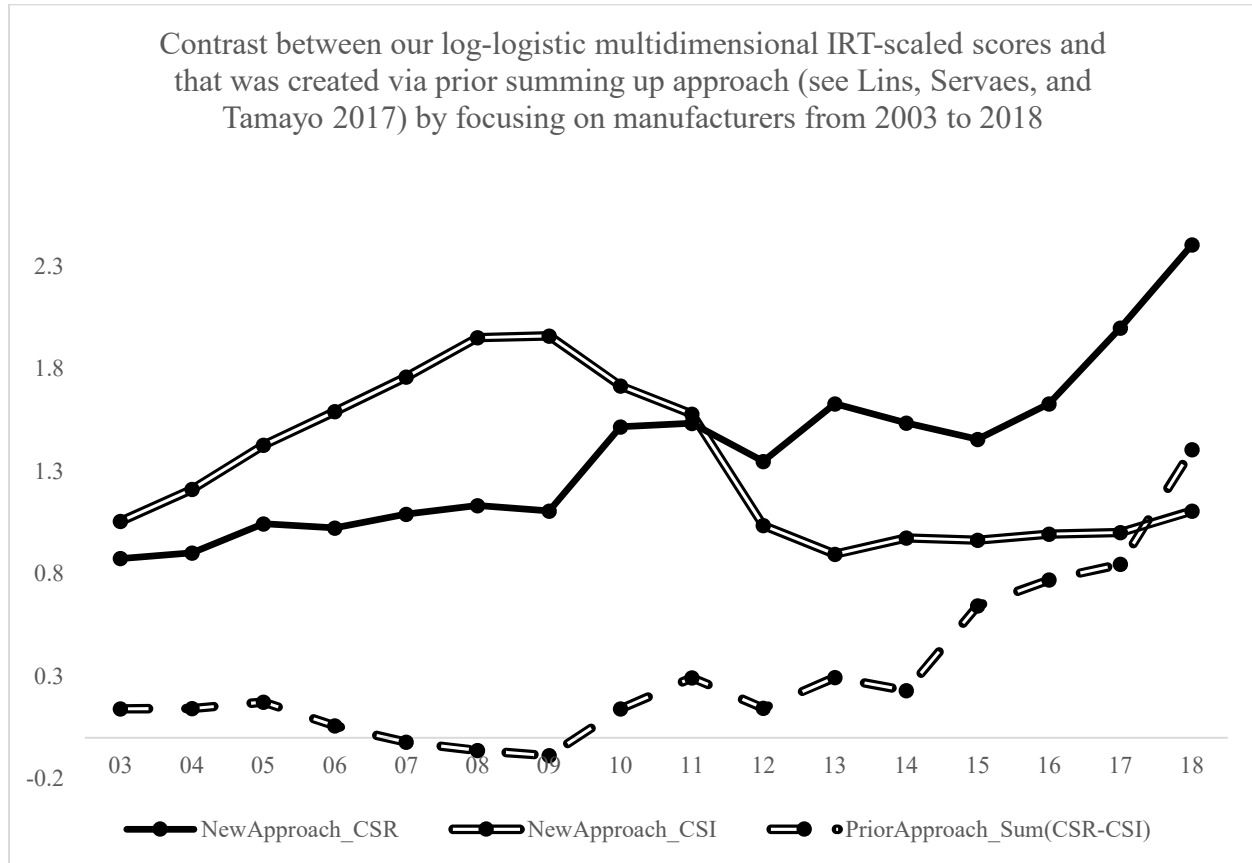
Finally, we conducted a log-logistic multidimensional IRT scaled score estimation for the set of 20 items that we focused on (Lucke 2013, 2015). This is because KLD indicators are characterized by low base rates of implementation and highly skewed total scores. The assumption of underlying normal distributed latent trait scores is violated. As such, the appropriateness of using standard IRT-scaled approaches should be questioned because the estimation of these model parameters can be highly inaccurate (Kirischi, Hsu, and Yu 2001; Reise et al. 2018; Wall, Park, and Moustaki 2015). Under this log-logistic IRT approach, latent trait (i.e., θ) is assumed log-normal, which is anchored at 0 on the low end and positive infinity on the high end. The discrimination parameters are defined exactly the same in the log-logistic IRT models as in the standard IRT models (Reise et al. 2021). The other parameters are exponential transformations of IRT parameters. That is to say, severity parameters are the ones that provide evidence for all practical purposes, as discrimination parameters have been treated as the same set of parameters when calculating the factor scores. A log-logistic IRT model can be regarded as an exponential transformation of a standard IRT model. As such, the log-logistic multidimensional IRT scaled scores can be estimated by the exponential transformation of the IRT scaled scores.

IRT-scaled scores can be estimated through either the mean or the mode of the posterior distribution of the latent construct for each firm according to its observed item responses and the model parameter estimations (Skrondal and Rabe-Hesketh 2004). For continuous items, the mean and mode of the posterior estimate can be addressed by the regression method (Knott and Bartholomew 1991). For binary or ordinal items, however, the mean and mode are unequal while being highly correlated (Knott and Bartholomew 1999). The mean approach is usually referred to

as the expected a posteriori (EAP) estimate, while the mode approach is usually referred to as the modal a posteriori (MAP) estimate (Thissen and Orlando 2001). Based on the recommendations of Thissen and Orlando (2001), we used the EAP method to obtain scores of CSR and CRI separately for each firm at each time point in this study.

To evaluate our approach, we then compared our log-logistic multidimensional IRT-scaled scores with those that were created via prior approaches by focusing on manufacturers from 2003 to 2018. As seen in Figures 1.11, firms improved their CSR rapidly during 2003 – 2018, while firms' CSI scoring was increasing from 2003 to 2009 and started to decrease since 2009, which indicates that firms' CSI performance became worse but get started to improve later on. However, it is hard to find much information by examining the average of firm performance conducted by using the sum of CSR minus the sum of CSI, which is the most common approach to measure firms' sustainability in the existing literature. As such, by comparing our new approach with prior approaches, we can conclude that our new approach provides much richer information regarding how firms' sustainability performance evolved over time compared to the existing approaches. Therefore, separately analyzing firms' CSR and CSI allows us to investigate the longitudinal evolution of corporate social action, while prior empirical research that adopted measurement with numerous weaknesses limited our ability to look at how firms' sustainability performance evolved over time.

Figure 1.11 Compare Log-logistic IRT Approach with Prior Summing-up Approach



1.5.4 How Firm Size Affects CSR and CSI

Finally, we take a peek at how firm size affects CSR and CSI in the case of the manufacturing industry. We plot the effects of firm size on CSR and CSI for small (10th percentile), medium (50th percentile), and large (90th percentile) manufacturing firms in Figures 1.12 & 1.13. Consistent with our expectations, Figure 1.12 reveals that larger firms' CSR activities were improving consistently, and the rate of improvement was more rapid during the early stage, while smaller firms with poor initial CSR scores have a more pronounced rate of increase later on. Figures 1.13 displays that larger firms' CSI scoring was increasing during the early stage, but got started to decrease rapidly as time passes. However, smaller firms' CSI behaviors have not changed much. This indicates that larger firms exhibited worse performance in CSI during

the early stage given they are more exposed, but improve rapidly as time elapses, while smaller ones have no incentive to improve their CSI practices as they are less visible. Indeed, Figures 1.12 & 1.13 show that firm size can affect how CSR and CSI evolved over time.

A few potential explanations for the effect of firm size on CSR and CSI come to mind. For CSR, Larger firms compared to smaller peers increase more rapidly during the early stage, however, as time elapses, larger firms display less pronounced rates of increase in CSR practices than smaller firms. This is because larger firms are more likely to benefit from socially responsible practices given that CSR can generate “moral capital from stakeholders in the form of customer credibility, brand faith, employee affective commitment, community legitimacy, supplier trust, and shareholder reliability” (Price and Sun 2017, p. 86). Thus, larger firms will have more incentive to increase socially responsible practices than their smaller peers at first. In particular, larger firms are more likely to have the resource capabilities to take action rapidly, such as hiring specialists. As time passes, however, larger firms have less room to increase CSR practices because they eventually run out of opportunities, while smaller firms with the worst initial CSR have low marginal cost opportunities for increasing CSR. As such, larger firms will display less pronounced rates of change than smaller peers regarding CSR, as time elapsed.

However, the effect of size on CSI can be very different from that effect on CSR. Larger firms expose higher increased rates of CSI activities than smaller firms at first, however, as time elapses, larger firms exhibit more rapid reduced rates of CSI than smaller firms. On the one hand, companies don't know what they were doing regarding CSI activities at first, as in the case of looking at how surprisingly quickly firms offshored in the early 2000s (Pierce and Schott 2016). On the other hand, larger firms tend to be more visible than smaller peers, and thus socially irresponsible behaviors of larger firms are more likely to be exposed compared to those

of smaller peers. Moreover, larger firms are more likely to be subjected to greater scrutiny and pressure from outside than smaller peers (Baron, Mittman, and Newman 1991; Goodstein 1994; Ingram and Simons 1995). Consequently, larger firms' CSI behaviors, thus are more likely to be exposed than smaller firms' (Kagan, Gunningham, and Thornton 2011; Ji and Weil 2015). Therefore, larger firms will have more pronounced increased CSI behaviors than smaller peers during the early stage. However, as time passes, larger firms will be more aware of reducing CSI behaviors and thus display more rapidly decreased CSI than smaller firms. This is because once larger firms recognize that the increased CSI brings more risks "through customer and shareholder lawsuits, community proceedings, and supplier mistrust" (Price and Sun 2017, p. 86), larger firms will put effort to reduce CSI behaviors and they have the resource capability to generate better knowledge that allows them to make improvements rapidly. Yet, smaller firms have fewer incentives to reduce CSI activities as they are less visible. Taken together, larger firms compared to smaller ones show more rapid rates of increase regarding CSI activities during the early stage, however, as time elapses, larger firms display more pronounced decreased CSI behaviors than smaller firms. Although we only take a peek at how firm size affects CSR and CSI without testing, this can shed light on the questions: how CSR and CSI evolve over time, and how firm size moderates this evolution. Further research should examine these questions.

Figure 1.12 Plot of the Moderating Effects of Size for Rate of Change regarding CSR

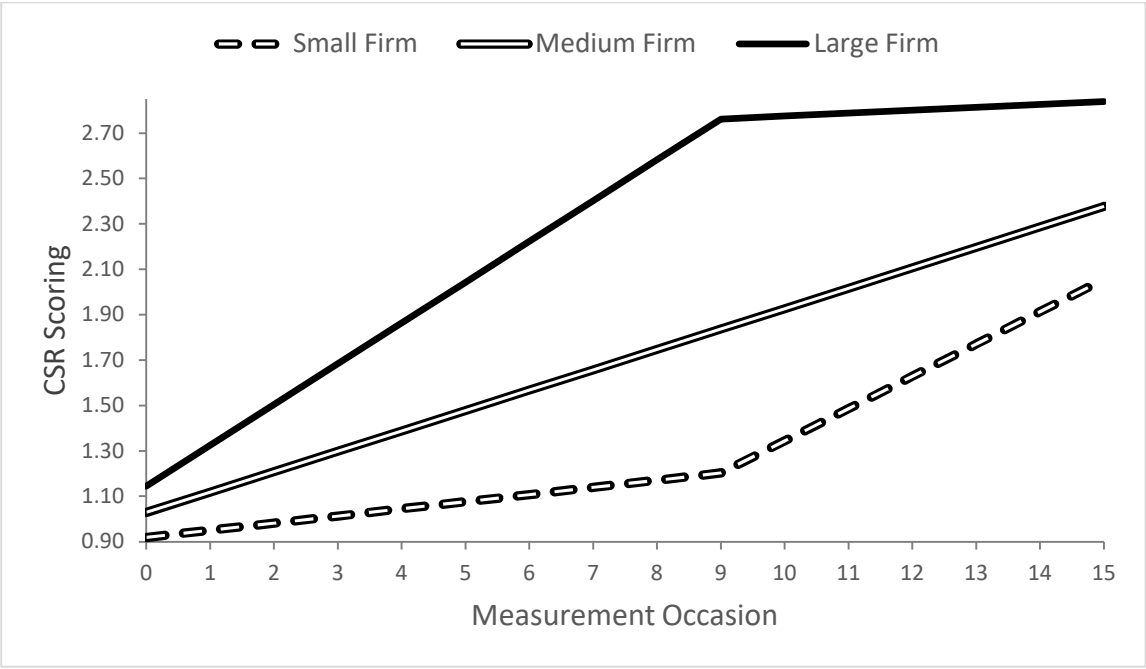
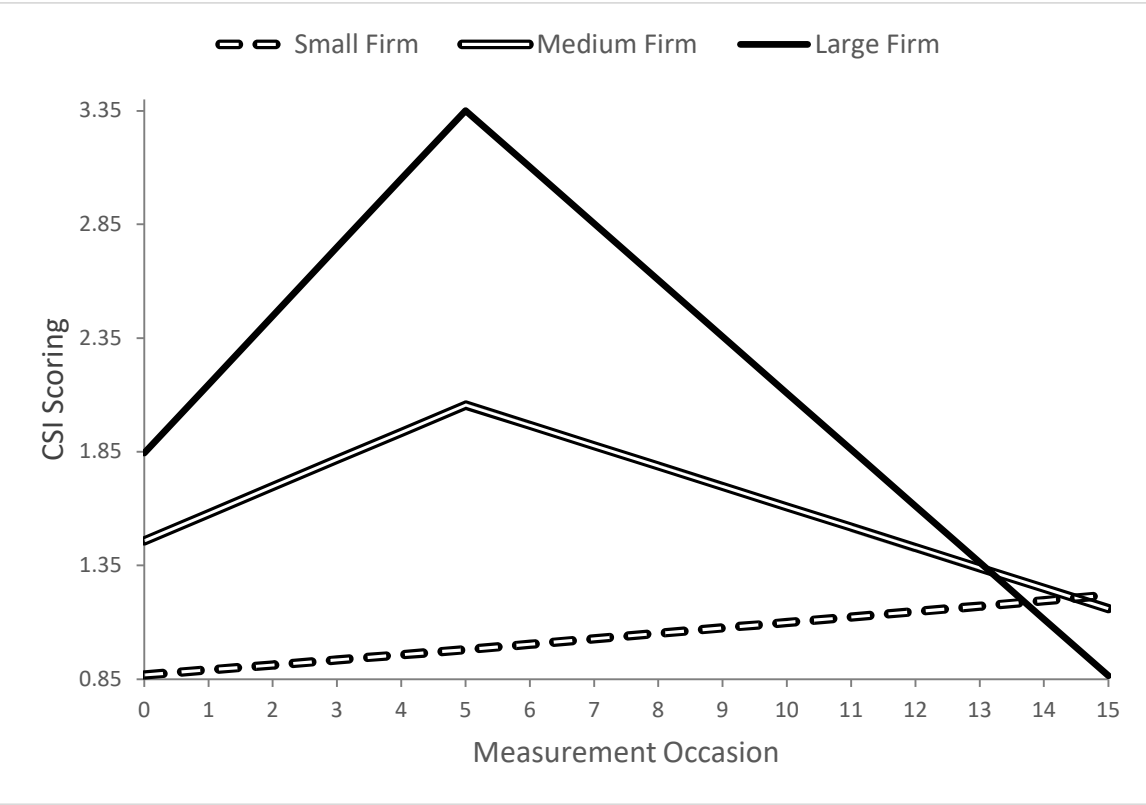


Figure 1.13 Plot of the Moderating Effects of Size for Rate of Change regarding CSI



1.6 Discussion

1.6.1 Contributions

Our work makes contributions to the existing literature in several ways. First, this manuscript focuses on developing a whole new set of metrics to measure firm-level sustainability by separating strengths via CSR and weaknesses via CSI from archival data, which no one in our field does. Such a new approach using state-of-the-art psychometric techniques and focusing on the set of sustainability metrics that pertains to the supply chain allows us to address measurement problems encountered with prior approaches and examine how firms' sustainability performance as it pertains to their supply chain evolved over time. This is critical as the existing measurement approaches have been inconsistent and often contradictory (Chatterji, Levine, and Toffel 2009; Eccles, Lee, Strohle 2020; Mattingly and Berman 2006), which poses a major challenge for building cumulative knowledge (Shaver 2020) and hampers replication and extension (Pagell 2021). Specifically, this lack of valid measurement has handicapped the ability to answer important questions such as how has firms' sustainability performance as it pertains to their supply chain evolved over time, and what factors affect this evolution.

Second, it applies a specific model, the log-logistic IRT model (Lucke 2013, 2015), which provides several improvements in the social and behavioral literature using categorical data (e.g., KLD data) that is characterized by very low base rates of the indicator and highly skewed total scores (Reise, Du, Wong, and Hubbard 2021). The advantage of our approach is that it can be formulated to more realistically represent the underlying processes of specific types of unipolar constructs (Lucke 2013, 2015). This is because parameter estimates via the traditional approaches can be highly inaccurate, when the normality assumption of the underlying latent construct is violated (Kirischi, Hsu, and Yu 2001; Reise, Rodriguez, Spritzer,

and Hays 2018; Wall, Park, and Moustaki 2015). In addition, our approach also rejects some other unrealistic assumptions under the prior approaches, such as the parallel structure and the negative relationship between strengths and weaknesses (e.g., Chatterji, Levine, and Toffel 2009; Chiu and Sharfman 2011; Deckop, Merriman, and Gupta 2006), as well as the unidimensionality (e.g., Carroll et al. 2016). We instead let data speak and identify the empirical relationship between CSR and CSI, which addresses measurement problems encountered with prior approaches. Collectively, our work echoes the calls of Miller and Kulpa (2022) for studying a phenomenon using archival data. Such an approach is particularly appropriate “in the early stages of the study of a phenomenon, when neither theory nor knowledge about correlates of the phenomenon is well developed” (Menard 1995, p. 42). Specifically, our scale development and empirical testing can be used to support the nascent development of concepts and relationships in the supply chain discipline (Rabinovich and Cheon 2011).

Third, our application of the multidimensional IRT approach to examine how firms respond to CSR and CSI has many implications for empirical inquiry in both firms’ sustainability domain and broader work seeking to understand how firms respond to institutional pressure. This is because our approach improves on existing state-of-the-art psychometric methods by developing measures of sustainability based on a richer, theory-driven analysis of how latent constructs of the respondents are reflected in proxies (Carroll et al. 2016; Embretson and Reise 2000). Another advantage of our approach is that the parameters of respondents are not dependent on any items and the parameters of each item are also not dependent on the latent constructs (Hambleton, Swaminathan, and Rogers 1991), which allows us to figure out the most appropriate items for measuring latent constructs (Stata Press 2015). In particular, our approach is capable of developing a measurement scale where the items and firms are positioned jointly,

which facilitates the interpretation of the scale (De Ayala 2013; Fernandes and Bornia 2019). As such, our scale development has several methodological advantages over other techniques in the literature (e.g., Carroll, Primo, and Richter 2016; Chiu and Sharfiman 2011; Graves and Waddock 1994; Mattingly and Berman 2006; Ruf, Muralidhar, and Paul 1998), regarding developing new measures to address organizational sustainability constructs more effectively (Nye et al. 2019).

In addition, analyzing firms' CSR and CSI separately allows scholars to explicitly capture the heterogeneity of firms' social responsibility and irresponsibility behaviors by looking at how firms respond to institutional pressures over time, which is a topic with substantial interest (Delmas and Toffel 2010; Doshi et al. 2013). This, in turn, allows for the examination of nuanced hypotheses such as how has firms' sustainability performance as it pertains to their supply chain evolved over time, what factors affect this evolution, as well as how the moderation effects change as time passes (Fitzmaurice, Laird, and Ware 2012). Moreover, our work also contributes to the literature by juxtaposing previous research in the assessment of social and environmental sustainability perception (e.g., Vincenzi et al. 2018), strategic management (e.g., Chen and Miller 2012), and supply chain sustainability (e.g., Chen and Ho 2019, Fernandes and Bornia 2019). In conclusion, the new measures conducted in this study are characterized by a host of advantages, such as marked increases in statistical power; broader psychometric assessment of theoretical constructs; longer longitudinal windows of study; the opportunity to test hypotheses not considered in the existing literature; and increased efficiency in developing new scales for concepts with latent characteristics.

1.6.2 Limitations and Suggestions for Future Research

As with all research, our work has limitations. First, the time span for this research consists of 16 years starting from 2003 to 2018, which can be extended further based on the data available. Second, the number of firms under investigation is limited because of the inconsistent ratings of some firms in the KLD database over the defined timespan. In other words, some firms have been rated irregularly between 2003 and 2018. As such, I have to exclude such firms from the analysis. Third, the measurement of firms' sustainability performance based on the log-logistic multidimensional IRT technique is rather a unique and novel approach, such that further research should examine and then extend this topic. Additionally, our approach can be replicated by specifically focusing on a certain country or a specific industry. Moreover, other IRT models can be applied in further research depending on the variables of a dataset. For example, develop new measures accounting for measurement nonequivalence across groups by how far the products are from final use (Antràs et al. 2012). Another direction of extension of this study can be to take the log-logistic multidimensional IRT scaled scores to some specific analytic frameworks, such as fitting a latent growth curve model directly to this measurement for longitudinal analysis.

APPENDIX

APPENDIX Comparison of Current Research with Selective Literature

Table 1.5 Academic Attempts to Measure Firms' Sustainability Performance via Using KLD
Rating Data

Citation	Sum items within categories	Use item response theory (IRT)
Graves and Waddock (1994)	Sum up all the strengths and concerns within categories on the basis of a weighting scheme and subtract the sum of concerns from the sum of strengths to form an overall numerical scale as a measure for research purposes.	No
Sharfman (1996)	Subtract each category's average concern score from its average strength score. Then, sum up the score for each category and divide by the number of categories to form an overall average scale as a measure.	No
Griffin and Mahon (1997)	Sum up all the strengths and concerns within categories and subtract the sum of concerns from the sum of strengths to form an overall numerical scale as. Then, combine this numerical scale with a firm's perceptual information to generate an aggregate measure for research purposes. .	No
Ruf, Muralidhar, and Paul (1998)	Weights were multiplied by the sum from the strengths minus the concerns for each category. Sum up all the categories to form an overall score as an aggregate measure for research purposes.	No
Johnson and Greening (1999)	Test a factor structure by focusing on five category and aggregate the strength and concerns scores into two dimensions via structural equation modeling.	No
Hillman and Keim (2001)	Strength scores minus concern scores to form a single indicator for each category. Then, construct them into two dimensions by simply summing up indicators.	No
Graafland, Eijffinger, and Smid (2004)	Aggregate an average score by summing up all the strengths and concerns within categories on the basis of a weighting scheme and dividing by the sum of weights involved..	No
Deckop, Merriman, and Gupta (2006)	Sum up all the strengths and concerns within categories and subtracting the sum of concerns from the sum of strengths to form an overall score as a measure for research purposes.	No

Table 1.5 (cont'd)

Mattingly and Berman (2006)	Sum separately the strengths and concerns within categories and leave them separate for uncovering latent factor structures by using the exploratory factor analysis (EFA) to form two dimensions as research measures. This study rejects the assumption that the strengths and concerns covary in opposing directions.	No
Waldman, Siegel, and Javidan (2006)	Strength scores minus concern scores to form a single indicator for each category. Then, conduct a factor analysis with principle components as the extraction procedure and varimax rotation to assess the factor structure and result in two dimensions as research measures of corporate social performance.	No
David, Bloom, and Hillman (2007)	Sum up the strengths and concerns within each category to form an indicator for each category. Then, construct them into two dimensions by simply summing up indicators.	No
Kempf and Osthoff (2007)	Convert concerns into strengths by taking binary complements (meaning that if a certain weakness is not present, i.e. rated as 0, it is considered a strength rated with 1; if the weakness is present, then its corresponding strength is rated 0). Then, sum up all 'strengths' to form an overall score.	No
Chatterji, Levine, and Toffel (2009)	Subtract the sum of the concerns from the sum of strengths to for a single net score as a measure	No
Statman and Glushkov (2009)	Rank companies without aggregating their ratings. Define a top-overall company as one that is in the top third of companies for at least two categories but not in the bottom third by any category based on the best-in-class adusted score. Similarly, a bottom-overall company is one that is in the bottom third of companies by two or more categories, but not in the top third by any others.	No
Chiu and Sharfman (2011)	Conduct an average strength score by summing the strength scores within categories and dividing by the number of categories.	No

Table 1.5 (cont'd)

Herzel, Nico losi, and Stărică (2012)	Average separately the strengths and concerns within each category and aggregate the average scores into three dimensions of sustainability by summing the average scores within each dimension and dividing by the number of average scores involved.	No
Delmas, Etzion, and Nairn-Birch (2013)	Sum separately the strengths and concerns to form two dimensions.	No
Nicolosi, Grassi, and Stanghellini (2014)	Subtract the average value over the concerns from the average value over the strengths to form an overall score. Then, based on the overall score, conduct a three-category response variable.	Adopt the polytomous item response models to conduct the latent dimension of the three-category response variable to measure a firm's social responsibility.
Carroll, Primo, and Richter (2016)	Sum up the strengths and concerns within categories on the basis of a measurement technique by generating measures of latent traits that are reflected by identical scores on an additive scale.	Adopt an IRT model focusing on the utility for adopting a particular sustainability-related policy.
Castillo, Mollenkopf, Bell, and Bozdogan (2018)	This study used three measures: total aggregate score, which is the difference between the total number of strengths and the total number of concerns in a given dimension for each year; total strengths, which is the total number of strengths; and total weaknesses, which is the total number of concerns.	No

REFERENCES

REFERENCES

- Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, 65(10), 4451-4469.
- Antràs, P., Chor, D., Fally, T., & Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review*, 102(3), 412-16.
- Baron, J. N., Mittman, B. S., & Newman, A. E. (1991). Targets of opportunity: Organizational and environmental determinants of gender integration within the California civil service, 1979-1985. *American Journal of Sociology*, 96(6), 1362-1401.
- Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychological methods*, 14(2), 101.
- Becchetti, L., & Ciciretti, R. (2009). Corporate social responsibility and stock market performance. *Applied financial economics*, 19(16), 1283-1293.
- Berg, F., Koelbel, J. F., & Rigobon, R. (2019). Aggregate confusion: The divergence of ESG ratings (pp. 1-42). Cambridge, MA, USA: MIT Sloan School of Management.
- Bird, R., Hall, A. D., Momentè, F., & Reggiani, F. (2007). What corporate social responsibility activities are valued by the market?. *Journal of business ethics*, 76(2), 189-206.
- Birnbaum, A. L. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29-51.
- Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 504-516.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate behavioral research*, 36(1), 111-150.
- Busse, C., Kach, A. P., & Bode, C. (2016). Sustainability and the false sense of legitimacy: How institutional distance augments risk in global supply chains. *Journal of Business Logistics*, 37(4), 312-328.
- Cai, L., & Houts, C. R. (2019). Diagnostic Classification Modeling with flexMIRT. In *Handbook of Diagnostic Classification Models* (pp. 573-579). Springer, Cham.

- Cai, L., & Monroe, S. (2013). IRT model fit evaluation from theory to practice: Progress and some unanswered questions. *Measurement: Interdisciplinary Research and Perspectives*, 11(3), 102-106.
- Carroll, R. J., Primo, D. M., & Richter, B. K. (2016). Using item response theory to improve measurement in strategic management research: An application to corporate social responsibility. *Strategic Management Journal*, 37(1), 66-85.
- Castillo, V. E., Mollenkopf, D. A., Bell, J. E., & Bozdogan, H. (2018). Supply chain integrity: A key to sustainable supply chain management. *Journal of Business Logistics*, 39(1), 38-56.
- Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility?. *Journal of Economics & Management Strategy*, 18(1), 125-169.
- Chen, C. M., & Ho, H. (2019). Who pays you to be green? How customers' environmental practices affect the sales benefits of suppliers' environmental practices. *Journal of Operations Management*, 65(4), 333-352.
- Chen, M.J., Miller, D. (2012). Competitive dynamics: Themes, trends, and a prospective research platform. *Academy of Management Annals* 6(1) 135-210.
- Chiu, S. C., & Sharfman, M. (2011). Legitimacy, visibility, and the antecedents of corporate social performance: An investigation of the instrumental perspective. *Journal of Management*, 37(6), 1558-1585.
- Chowdhury, M. M. H., & Quaddus, M. A. (2021). Supply chain sustainability practices and governance for mitigating sustainability risk and improving market performance: A dynamic capability perspective. *Journal of Cleaner Production*, 278, 123521.
- Cudeck, R., & Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annu. Rev. Psychol.*, 58, 615-637.
- Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: the role of item response theory in integrative data analysis. *Developmental psychology*, 44(2), 365.
- David, P., Bloom, M., & Hillman, A. J. (2007). Investor activism, managerial responsiveness, and corporate social performance. *Strategic Management Journal*, 28(1), 91-100.
- De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
- Deckop, J. R., Merriman, K. K., & Gupta, S. (2006). The effects of CEO pay structure on corporate social performance. *Journal of Management*, 32(3), 329-342.
- Delmas, M. A., Etzion, D., & Nairn-Birch, N. (2013). Triangulating environmental performance: What do corporate social responsibility ratings really capture?. *Academy of Management Perspectives*, 27(3), 255-267.

- Delmas, M. A., & Toffel, M. W. (2010). Institutional pressures and organizational characteristics: Implications for environmental strategy. Harvard Business School Technology & Operations Mgt. Unit Working Paper, (11-050).
- Distelhorst, G., Hainmueller, J., & Locke, R. M. (2017). Does lean improve labor standards? Management and social performance in the Nike supply chain. *Management Science*, 63(3), 707-728.
- Goodstein, J. D. (1994). Institutional pressures and strategic responsiveness: Employer involvement in work-family issues. *Academy of Management journal*, 37(2), 350-382.
- Doshi, A. R., Dowell, G. W., & Toffel, M. W. (2013). How firms respond to mandatory information disclosure. *Strategic Management Journal*, 34(10), 1209-1231.
- Dyck, A., Lins, K. V., Roth, L., & Wagner, H. F. (2019). Do institutional investors drive corporate social responsibility? International evidence. *Journal of Financial Economics*, 131(3), 693-714.
- Eccles, R. G., Lee, L. E., & Strohle, J. C. (2020). The social origins of ESG: An analysis of Innovest and KLD. *Organization & Environment*, 33(4), 575-596.
- Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah.
- Fahimnia, B., Sarkis, J., & Talluri, S. (2019). Design and management of sustainable and resilient supply chains. *IEEE Transactions on Engineering Management*, 66(1), 2-7.
- Fernandes, S. M., & Bornia, A. C. (2019). Reporting on supply chain sustainability: Measurement using item response theory. *Corporate Social Responsibility and Environmental Management*, 26(1), 106-116.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998). John Wiley & Sons.
- Flammer, C. (2015). Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. *Management Science*, 61(11), 2549-2568.
- Flora, D. B., Curran, P. J., Hussong, A. M., & Edwards, M. C. (2008). Incorporating measurement nonequivalence in a cross-study latent growth curve analysis. *Structural equation modeling: a multidisciplinary journal*, 15(4), 676-704.
- Foster, G. C., Min, H., & Zickar, M. J. (2017). Review of item response theory practices in organizational research: Lessons learned and paths forward. *Organizational Research Methods*, 20(3), 465-486.
- Godfrey, P. C., & Hill, C. W. (1995). The problem of unobservables in strategic management research. *Strategic management journal*, 16(7), 519-533.

- Graafland, J. J., Eijffinger, S. C., & SmidJohan, H. (2004). Benchmarking of corporate social responsibility: Methodological problems and robustness. *Journal of business ethics*, 53(1), 137-152.
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and psychological measurement*, 66(6), 930-944.
- Graves, S. B., & Waddock, S. A. (1994). Institutional owners and corporate social performance. *Academy of Management journal*, 37(4), 1034-1046.
- Griffin, J. J., & Mahon, J. F. (1997). The corporate social performance and corporate financial performance debate: Twenty-five years of incomparable research. *Business & society*, 36(1), 5-31.
- Hambleton, R. K., Swaminathan, H., Rogers, H. J., & Jaeger, R. M. (1991). *Fundamentals of Item Response Theory*. Sage Publications.
- Hardcopf, R., Shah, R., & Dhanorkar, S. (2021). The impact of a spill or pollution accident on firm environmental activity: An empirical investigation. *Production and Operations Management*, 30(8), 2467-2491.
- Hart, T. A., & Sharfman, M. (2015). Assessing the concurrent validity of the revised Kinder, Lydenberg, and Domini corporate social performance indicators. *Business & Society*, 54(5), 575-598.
- Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *The Journal of Finance*, 74(6), 2789-2837.
- Herzel, S., Nicolosi, M., & Stărică, C. (2012). The cost of sustainability in optimal portfolio decisions. *The European Journal of Finance*, 18(3-4), 333-349.
- Houts, C. R., & Cai, L. (2020). *flexMIRT user's manual version 3.6: flexible multilevel multidimensional item analysis and test scoring*. Chapel Hill, NC: Vector Psychometric Group.
- Ingram, P., & Simons, T. (1995). Institutional and resource dependence determinants of responsiveness to work-family issues. *Academy of Management Journal*, 38(5), 1466-1482.
- Ji, M., & Weil, D. (2015). The impact of franchising on labor standards compliance. *ILR Review*, 68(5), 977-1006.
- Johnson, R. A., & Greening, D. W. (1999). The effects of corporate governance and institutional ownership types on corporate social performance. *Academy of management journal*, 42(5), 564-576.

- Kagan, R. A., Gunningham, N., & Thornton, D. (2011). Fear, duty, and regulatory compliance: lessons from three research projects. *Explaining Compliance: Business Responses to Regulation*, 37-58.
- Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, 13(5), 908-922.
- Kirisci, L., Hsu, T. C., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied psychological measurement*, 25(2), 146-162.
- Knott, M., & Bartholomew, D. J. (1993). Constructing measures with maximum reliability. *Psychometrika*, 58(2), 331-338.
- Knott, M., & Bartholomew, D. J. (1999). *Latent variable models and factor analysis* (Vol. 7, No. 2nd). Edward Arnold.
- Krueger, P., Sautner, Z., & Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, 33(3), 1067-1111.
- Ladygina, I. (2021). Measurement of corporate carbon management with item response theory.
- Liang, H., & Renneboog, L. (2017). On the foundations of corporate social responsibility. *The Journal of Finance*, 72(2), 853-910.
- Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *the Journal of Finance*, 72(4), 1785-1824.
- Lord, F. (1952). A theory of test scores. *Psychometric monographs*.
- Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.
- Lucke, J. F. (2013). Positive trait item response models. In *New developments in quantitative psychology* (pp. 199-213). Springer, New York, NY.
- Lucke, J. F. (2015). Unipolar Item Response Models. *Handbook of item response theory modeling: Applications to typical performance assessment*, 272-284.
- Luo, X., & Bhattacharya, C. B. (2006). Corporate social responsibility, customer satisfaction, and market value. *Journal of marketing*, 70(4), 1-18.
- Mattingly, J. E., & Berman, S. L. (2006). Measurement of corporate social action: Discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society*, 45(1), 20-46.

- McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of personality assessment*, 100(1), 43-52.
- Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.
- Miller, J. W., & Kulpa, T. (2022). Econometrics and archival data: Reflections for purchasing and supply management (PSM) research. *Journal of Purchasing and Supply Management*, 100780.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. ETS Research Report Series, 1992(1), i-30.
- Nicolosi, M., Grassi, S., & Stanghellini, E. (2014). Item response models to measure corporate social responsibility. *Applied Financial Economics*, 24(22), 1449-1464.
- Negri, M., Cagno, E., Colicchia, C., & Sarkis, J. (2021). Integrating sustainability and resilience in the supply chain: A systematic literature review and a research agenda. *Business Strategy and the Environment*, 30(7), 2858-2886.
- Nye, C. D., & Drasgow, F. (2011). Assessing goodness of fit: Simple rules of thumb simply do not work. *Organizational Research Methods*, 14(3), 548-570.
- Nye, C. D., Joo, S. H., Zhang, B., & Stark, S. (2020). Advancing and evaluating IRT model data fit Indices in organizational research. *Organizational Research Methods*, 23(3), 457-486.
- Orlitzky, M., & Benjamin, J. D. (2001). Corporate social performance and firm risk: A meta-analytic review. *Business & Society*, 40(4), 369-396.
- Pagell, M. (2021). Replication without repeating ourselves: Addressing the replication crisis in operations and supply chain management research. *Journal of Operations Management*, 67(1), 105-115.
- Pierce, J. R., & Schott, P. K. (2016). The surprisingly swift decline of US manufacturing employment. *American Economic Review*, 106(7), 1632-62.
- PRI (2018). PRI Annual Report 2018. Available at: <https://www.unpri.org/annual-report-2018>. (accessed 10 April 2022).
- Pritikin, J. N., Brick, T. R., & Neale, M. C. (2018). Multivariate normal maximum likelihood with both ordinal and continuous variables, and data missing at random. *Behavior research methods*, 50(2), 490-500.
- Rabinovich, E., & Cheon, S. (2011). Expanding horizons and deepening understanding via the use of secondary data sources. *Journal of Business Logistics*, 32(4), 303-316.
- Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Routledge.

- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.
- Reckase, M. D. (2009). Multidimensional item response theory models. In *Multidimensional item response theory* (pp. 79-112). Springer, New York, NY.
- Reise, S. P., Du, H., Wong, E. F., Hubbard, A. S., & Haviland, M. G. (2021). Matching IRT models to patient-reported outcomes constructs: The graded response and log-logistic models for scaling depression. *psychometrika*, 86(3), 800-824.
- Reise, S. P., & Rodriguez, A. (2016). Item response theory and the measurement of psychiatric constructs: some empirical and conceptual issues and challenges. *Psychological Medicine*, 46(10), 2025-2039.
- Reise, S. P., Rodriguez, A., Spritzer, K. L., & Hays, R. D. (2018). Alternative approaches to addressing non-normal distributions in the application of IRT models to personality measures. *Journal of personality assessment*, 100(4), 363-374.
- Rowley, T., & Berman, S. (2000). A brand new brand of corporate social performance. *Business & society*, 39(4), 397-418.
- Ruf, B. M., Muralidhar, K., & Paul, K. (1998). The development of a systematic, aggregate measure of corporate social performance. *Journal of Management*, 24(1), 119-133.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.
- Sarkis, J., Gonzalez-Torre, P., Adenso-Diaz, B. (2010). Stakeholder pressure and the adoption of environmental practices: The mediating effect of training. *Journal of Operations Management* 28(2) 163-176.
- Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management science*, 59(5), 1045-1061.
- Sharfman, M. (1996). The construct validity of the Kinder, Lydenberg & Domini social performance ratings data. *Journal of Business Ethics*, 15(3), 287-296.
- Shaver, J. M. (2020). Causal identification through a cumulative body of research in the study of strategy and organizations. *Journal of Management*, 46(7), 1244-1256.
- Singer, J. D., Willett, J. B., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC.
- Stata Press. (2015). *Stata Item Response Theory Reference Manual Release 14*. Available at: <https://www.stata.com/manuals14/irt.pdf> (accessed 23 December 2021).

- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important?. *Journal of Applied Psychology*, 89(3), 497.
- Statman, M., & Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, 65(4), 33-46.
- Strike, V. M., Gao, J., & Bansal, P. (2006). Being good while being bad: Social responsibility and the international diversification of US firms. *Journal of International Business Studies*, 37(6), 850-862.
- Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393-408.
- Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, 18(1), 3-46.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models.
- United States Census Bureau. (2021). Business Dynamics Statistics: Annual Report: 2003-2018. Available at: <https://www.census.gov/data/tables/econ/awts/annual-reports.html> (accessed 19 March 2022).
- Vincenzi, S. L., Possan, E., de Andrade, D. F., Pituco, M. M., de Oliveira Santos, T., & Jasse, E. P. (2018). Assessment of environmental sustainability perception through item response theory: A case study in Brazil. *Journal of Cleaner Production*, 170, 1369-1386.
- Waddock, S. (2003). Myths and realities of social investing. *Organization & environment*, 16(3), 369-380.
- Wainer, H. (1993). Measurement problems. *Journal of educational measurement*, 30(1), 1-21.
- Waldman, D. A., Siegel, D. S., & Javidan, M. (2006). Components of CEO transformational leadership and corporate social responsibility. *Journal of management studies*, 43(8), 1703-1725.
- Wall, M. M., Park, J. Y., & Moustaki, I. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Applied Psychological Measurement*, 39(8), 583-597.
- Woods, C. M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika*, 71(2), 281-301.
- Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: current approaches and future directions. *Psychological methods*, 12(1), 58.

CHAPTER 2 - Longitudinal Examination of Firm-Level Supply Chain Sustainability

2.1 Introduction

Sustainability is growing increasingly important among the press, political activists, managers, analysts, investors, shareholders, and a company's other stakeholders, such as customers, employees, and local communities (Chen 2008; Kusi-Sarpong Gupta, and Sarkis 2019; McPeak, Devirian, and Seaman 2010; Negri et al. 2021). For example, a survey conducted by Nielsen in 2019 illustrated that 48% of consumers cared about sustainability issues, and this number increased to 83% among millennials (Langan and Menz 2022). And employees are more likely to trust the companies that develop socially responsible supply chain practices, as it is becoming increasingly evident that employees are emphasizing what they do, how they are treated, and whether their work is having a positive impact (Langan and Menz 2022; Solomon 2014). Therefore, a company must manage the supply chain practices aligned with the sustainability expectations of its stakeholders to achieve long-term success (Bai, Kusi-Sarpong, and Sarkis 2017), as failure to meet stakeholder sustainability expectations can result in various risks, such as consumers' negative perception of the firm, reputation and brand damage, labor disputes, market value reduce, or increased pressures from the community, government, and other social groups (Busse, Kach, and Bode 2016; Chowdhury and Quaddus 2021).

Investigating the development of socially responsible supply chain practices has been paid increasing attention and attached great importance to the sustainability and supply chain management field (Seuring and Müller 2008; Fahimnia, Sarkis, and Davarzani 2015). Yet, longitudinal examinations of firms' socially responsible supply chain practices have been noticeably scarce (Chowdhury and Quaddus 2021; Silvestre et al. 2020). This seems to run counter to the conceptual consensus that the development of socially responsible supply chain

practices goes through a complex and dynamic learning process, which cannot be achieved overnight but instead (Silvestre et al. 2020). Herein, this manuscript attempts to bridge this disconnect between theoretical and empirical research regarding the development of socially responsible supply chain practices by proposing and testing a longitudinal perspective that encourages researchers to leverage the past to predict the reactions of certain sustainability properties in the future. Literature suggests that a firm's socially responsible performance consists of not only "doing good through corporate social responsibility (CSR)" activities but also "doing bad through corporate social irresponsibility (CSI)" practices (Price and Sun 2017, p. 82). CSR initiatives have been generally recognized as highly desirable corporate behaviors that both benefit communities and help companies themselves perform better in business (Barnett 2007; McWilliams and Siegel 2001). In comparison, CSI activities, referred to as firms' irresponsible behaviors, have been regarded as harmful practices leading to substantive negative effects that may be harmful to various stakeholders (Armstrong and Green 2013). Researchers have traditionally focused on understanding firms' social responsibility by examining CSR activities, whereas CSI incidents have barely caught researchers' attention and thus have rarely been investigated so far among the existing studies (Murphy and Schlegelmich 2013; Price and Sun 2017). Only recently, has academic literature gotten started to broaden the understanding of firms' socially responsible practices through the inclusion of both CSR and CSI (e.g., Lenz, Wetzel, and Hammerschmidt 2017; Kang, Germann, and Grewal 2016; Price and Sun 2017). However, to date, the relationship between CSR and CSI has rarely been investigated (Groening and Kanuri 2013; Lenz, Wetzel, and Hammerschmidt 2017).

In this study, we adopt a between-firm orientation to examine how firm-level CSR and CSI in supply chain practices have changed over time, when rates of change are likely to be more

rapid, and what & how firm- and industry-level traits moderate such rates of change. This study focuses on firm-level practices, because emphasizing studying how subjects' performance evolves over time is particularly important in the supply chain discipline (Ketokivi and McIntosh 2017; Miller, Ganster, Griffis 2018). We expect that firm-level CSR supply chain practices will display different trajectories from CSI behaviors. Specifically, CSR activities are expected to grow more rapidly for a few years, and then the rates of growth slow down over time, whereas CSI activities are expected to increase at first but then go down rapidly. The beneficial role of CSR to firms has been confirmed such that firms have the motivation to adopt socially responsible practices rapidly, but rates of increase tend to slow down because firm's marginal costs of making improvements in CSR rise up as the absolute value in this measure goes up (Rosenthal, Quinn, and Harper 1997; Williams et al. 2005). In contrast, companies don't know what they were doing regarding CSI activities at first, as in the case of looking at how surprisingly quickly firms offshored in the early 2000s (Pierce and Schott 2016). The increasing CSI behaviors would then not only result in increased pressure regarding "handling relationships with the community, public, government, or other social groups" (Price and Sun 2017, p. 85), but also damage the companies' reputation and thus reduce the market value (Du, Bhattacharya, and Sen 2010; Fombrun, Gardberg, and Barnett 2000). The negative effect of CSI will eventually force the companies to reduce CSI behaviors as time goes by. To formulate and test our arguments, we draw on various theoretical lenses, such as the awareness-motivation-capability (AMC) framework (Chen 1996; Chen, Su, and Tsai 2007) and the attention-based perspective (Fiske and Taylor 1991; Ocasio 1997; Simon 1979), and test the theorizing with secondary archival data for thousands of public firms. More specifically, we use panel data from 2003 – 2018 for hundreds of publicly-traded manufacturers from KLD, which has been merged with

financial data from COMPUSTAT, market concentration from U.S. Census Bureau Economic Indicators (United States Census Bureau 2021), and Upstreamness measuring the average distance from final use from American economic review (Antràs et al. 2012).

This study makes theoretical and empirical contributions to the body of knowledge, regarding the longitudinal examinations of firms' supply chain sustainability, which has been limited by the numerous weaknesses of the prior approaches in the existing literature (Silvestre et al. 2020). First, this manuscript provides insights into the longitudinal examinations of firms' sustainability performance by splitting apart CSR and CSI in supply chain practices. Because investigating social responsibility by the inclusion of CSR and CSI can provide a broader, holistic viewpoint that enables researchers and practitioners to look at firms' regimes with new respect, such that outline better strategies involving socially responsible and irresponsible practices (Jones, Bowd, and Tench 2009; Lange and Washburn 2012; Sweetin, Knowles, Summey, and McQueen 2013). In particular, researchers have traditionally and primarily looked at the CSR perspective of sustainability and developed the logic and definition of social responsibility as an overall construct that was measured unidimensionally within the framework (Griffin and Mahon 1997). Such a unidimensional perspective is likely to limit our understanding of corporate social responsibility performance as it is usually comprised of responsible and irresponsible aspects of practices (e.g., Castillo, Mollenkopf, Bell, and Bozdogan 2018; Price and Sun 2017). Specifically, companies could involve CSR and CSI activities simultaneously in terms of doing both "good" and "bad" (Lenz, Wetzel, and Hammerschmidt 2017; Mattingly and Berman 2006; Strike, Gao, and Bansal 2006). For example, Nike has invested substantial resources in helping suppliers improve production processes yet has one or more suppliers engage in bad behavior (Distelhorst, Hainmueller, and Locke 2017). Second, our findings

highlight the fact that a firm goes through sequential learning loops regarding improving its socially responsible supply chain practices. Such loops could develop at different paces, which depends upon how decision-makers of the firm put their selective attention (Fiske and Taylor 1991; Ocasio 1997; Simon 1947). Our work responds to Matusik, Hohenbeck and Mitchell's (2021) call for organizational researchers to put more attention to investigating longitudinal phenomena, and particularly echoes Ketokivi and McIntosh's (2017) call for SCM scholars to examine how subjects' performance evolves over time. Moreover, our work theorizes what factors can affect firms' sustainability evolution by looking at how firm size would moderate rates of change in firm-level CSR and CSI supply chain practices over time. This echoes the call to develop and test "nuanced predictions" that probe for gaining a more precise understanding of underlying processes that bring about empirical relationships and characteristics (Edwards and Berry 2010). Collectively, this research directly adds to previous research in supply chain sustainability by developing and testing a nuanced theory via specifying piecewise latent growth models. In particular, our application of growth modeling techniques in understanding the longitudinal development of firm-level CSR and CSI in supply chain practices can provide important implications for empirical inquiry in both firms' sustainability domain and broader work focusing on understanding how firms respond to institutional pressure.

The rest of this study is organized into five sections. The next section covers the relevant literature. The second contains the theory and hypothesis development. The third details our research design and summarizes the relevant variables. The fourth describes the econometric methodology and presents the results of our analysis. The fifth explains theoretical contributions, presents managerial implications, notes limitations, and makes suggestions for future research.

2.2 Background Literature

2.2.1 Supply Chain Sustainability

Supply chain sustainability refers to corporates' efforts to manage the environmental and human impacts of their supply chain planning and decision-making (Ahi and Searcy 2013), which encourages good governance practices of operations throughout the lifecycles of goods and services. The goal of supply chain sustainability is to uphold long-term environmental and societal values for all stakeholders in and around the business operations, which requires organizations to take actions to minimize environmental and human harm throughout their supply chain (Bai, Kusi-Sarpong, and Sarkis 2017). The motivation for integrating sustainability practices into organizational supply chain operations is mainly derived from various social pressures, stricter government regulatory requirements, corporate image, increasing public awareness, and market pressures (Tseng, Lim, and Wong 2015; Esfahbodi, Zhang, and Watson 2016). A tremendous amount of time and effort has been and continues to be devoted to investigating the development of sustainable supply chain initiatives from different perspectives (e.g., Bai, Kusi-Sarpong, and Sarkis 2017; Fahimnia, Sarkis, and Davarzani 2015; Hofer, Cantor, and Dai 2012; Modi and Cantor 2020). In particular, literature on sustainability transition shows that supply chains are responsible for the reorientation of industries and regimes toward advancing sustainability and promoting broader sustainable development (Farla et al. 2012; Geels 2004). However, the longitudinal examinations of firm-level supply chain sustainability are still understudied. This occurs because companies tend to face complexity and context-specific challenges for sustainable business practices and particularly they are likely to take actions in their own interests, which leads to different trajectories of sustainable development (Geels 2014; Roy, Schoenherr, and Charan 2018). Specifically, such a supply chain sustainable

trajectory also depends on how efficiently companies learn and move toward advanced corporate sustainability, which can go through a “non-linear and multi-directional journey” by incorporating the fundamental role of time (Silvestre 2015).

2.2.2 Driving Factors of Firm-level Supply Chain Sustainability

The existing literature acknowledges that a firm’s sustainability transition starts with the firm’s awareness of changes in the business environment (Silvestre et al. 2020). Conceptualized as a firm’s strategic posture, some firms deliberately monitor and become aware of changes in their business conduct and regulatory environment like new environmentally friendly public policies (Rai and Tang 2010; Delmas and Montes-Sancho 2010; Kube et al. 2019). Indeed, Hofer, Cantor, and Dai (2012) theorized and found empirical support that market-leading firms observe and react to a rival firm’s environmental management practices. Modi and Cantor (2020) found that focal firms are likely to be under pressure from competitor firms to improve their environmental performance. Likewise, research finds that firms’ social sustainability implementation has also been gaining traction due to the awareness of changes in social responsibility towards society (Ahmadi, Kusi-Sarpong, and Rezaei 2017; Silvestre et al. 2020). Following this logic, we also contend that firms must remain aware of environmental and social changes in the business environment.

Next, research suggests the importance of firms’ motivation to alter their strategic posture toward the changing in the business environment. Dynamic changes regularly occur in many industries, including the introduction of new regulations. For example, in 2016, several prominent firms, including AT&T, Blue Cross/Blue Shield, Boeing, and Google, invested \$17 million, \$16 million, and \$15 million, respectively, in political lobbying (Martin, Josephson,

Vadakkepatt, and Johnson 2018). Thus, these, and many others, recognize that survival depends on the motivation to proactively influence the specific policies of new regulatory actions (Chen and Miller 2012; Oliver and Holzinger 2008). For instance, many firms are motivated to enhance their environmental and social performance because of various outside pressures that continue to change. In other words, many firms seek to improve their sustainability performance because of stakeholder pressure, changes in public policy, and competitive reasons (e.g., Donaldson and Preston 1995; Carter and Jennings 2002; Ellram and Murfield 2017; Hofer et al. 2012).

Additionally, studies emphasize the companies' capability to proactively respond to the changing in the business environment. Possessing superior resources or capabilities is critical to maintaining a competitive advantage. Financial and human resources provide firms with the capability to take actions that rivals cannot easily imitate or follow (Ndofor, Sirmon, and He 2011). This stream of research indicates that the availability of financial resources can directly influence a corporate's investment decisions to improve socially responsible supply chain practices, such as acquiring new physical capital assets that can have a positive environmental pay-off (Greve 2003; Hofer et al. 2012). For instance, Greve (2003) demonstrates that financial resources allow a company to have more flexibility to invest in new sustainable implications such as environmentally friendly technologies. In contrast, existing studies suggest that a firm with financial constraints (e.g., as reflected in high financial leverage) is more likely to opt to overlook the sustainable practices of its actions (Mishra and Modi 2013). According to the findings of existing research, we reason that a firm's capability can significantly influence its sustainability trajectory regarding the development of environmental and social practices (Parmigiani, Klassen, and Russo 2011).

To summarize, a firm can develop a proactive sustainable posture in response to the changing in the business environment when the firm has sufficient awareness, motivation, and capabilities (Chen and Miller 2012; Chi, Ravichandran, and Andreovski 2010). These three factors, in this study, allow us to reason how firms' CSR (i.e., doing good) and CSI (i.e., doing bad) evolved over time. Specifically, firms are made aware of sustainable development through repeated interactions with investors, customers, government regulators, and general awareness through articles in the trade press, along with word-of-mouth talks with industry associates. Likewise, firms are motivated to pursue better sustainable development because they are aware that improvement in sustainability performance is valued by various stakeholders and current/potential customers. Finally, larger firms have the resource capabilities to tend to have faster rates of change in improving their socially responsible supply chain practices. Collectively, these factors – awareness, motivation, and capabilities – provide firms with the ability to achieve superior sustainability performance.

2.2.3 Contributions to the Literature

Our work extends the literature in three ways. First, we rely on a longitudinal dataset to examine how firm-level CSR (i.e., doing good) and CSI (i.e., doing bad) in supply chain practices has evolved over time. By doing this, our work extends the literature to comprehensively understand the role of time in organizations' phenomena, which has rarely been examined in organizational research (Ancona et al. 2001; Miller, Ganster, and Griffis 2018). This responds to Matusik, Hollenbeck, and Mitchell's (2021) call for organizational researchers to take efforts to investigate longitudinal phenomena, and particularly echoes Ketokivi and McIntosh's (2017) call for SCM researchers to examine how subjects' performance evolves over time. Second, we develop a more holistic theory by analyzing firms' CSR and CSI activities separately. Our

findings highlight the fact that a firm goes through sequential learning loops regarding improving its socially responsible supply chain practices. Such loops could develop at different paces, which depends upon how decision-makers of the firm put their selective attention (Fiske and Taylor 1991; Ocasio 1997; Simon 1947). Third, our work theorizes what factors can affect firms' sustainability evolution by looking at how firm size would moderate rates of change in firm-level CSR and CSI supply chain practices over time. This echoes the call to develop and test “nuanced predictions” that probe for gaining a more precise understanding of underlying processes that bring about empirical relationships and characteristics (Edwards and Berry 2010). Collectively, this research directly adds to previous research in supply chain sustainability by developing and testing a nuanced theory via specifying piecewise latent growth models. As such, our work offers novel insights that extend knowledge regarding how firm-level socially responsible supply chain practices have evolved over time, when rates of change are likely to be more rapid, and what factors can affect this evolution over time.

2.3 Theory & Hypotheses Development

2.3.1 Longitudinal Development of Firm-level CSR in Supply Chain Practices

The theoretical lens of this study is the awareness-motivation-capability (AMC) framework (Chen 1996; Chen, Su, and Tsai 2007). We adopt this framework to theorize about several factors that motivate a firm to participate and improve its socially responsible supply chain practices. First, firms are starkly aware of the importance of CSR through doing good to ensure better publicity, and marketing (Hillman, Zardkoohi, and Bierman 1999; Tate, Ellram, and Kirchoff 2010). Specifically, the beneficial role of socially responsible supply chain practices has been justified by the stream of literature in various consumer areas. For example, more and more

consumers expect firms to develop social responsibility (e.g., Ipsos 2013; Kusi-Sarpong Gupta, and Sarkis 2019; Langan and Menz 2022). Such a consumer trend can be further revealed in the increased media attention sustainability received over time (Kang, Germann, and Grewal 2016). Specifically, as expectations of sustainable business operating have become a tendency worldwide (Du, Bhattacharya, and Sen 2010), consumers are willing to pay for goods and services with advanced corporate sustainability (De Pelsmacker, Driesen, and Rayp 2005). As such, socially responsible supply chain practices allow firms to create a positive consumer identification as a whole, given that such identification with firms has increasingly been sought by consumers (Bhattacharya and Sen 2004). Moreover, the positive perceptions of a firm can increase consumer loyalty that could promote the firm's financial performance, such as higher profits and incomes (Helgesen 2006); more successful cross-selling, and greater cost efficiency (Srinivasan, Anderson, and Ponnnavolu 2002), improved customer's lifetime value (Zhang, Dixit, and Friedmann 2010), and closer connections with consumers (Luo and Bhattacharya 2006).

Second, aligning CSR interests in supply chain practices allows firms to facilitate their relationships with suppliers given that firms can directly or indirectly incorporate their suppliers' offerings into their products and services (Castillo et al. 2018). In particular, CSR enables firms to closely link to various social groups surrounding them and thus creates a positive environment (Clarkson 1995). For instance, CSR can be an effective way to build close connections with local communities by incorporating social responsibility into their supply chain practices (Kapelus 2002). As a consequence, the positive social effect of CSR can enhance a firm's reputation and cost efficiency (Vanhamme and Grobben 2009). Therefore, CSR has been increasingly recognized by management research as a strategic action that allows firms to create both positive perceptions and promoted financial performance (McWilliams and Siegel 2011). Such practical

view reasons that CSR enables firms to obtain and maintain competitive advantage given socially responsible initiatives facilitate connections between firms and their stakeholders (Wang and Choi 2013). As such, firms have incentives to improve CSR because of the beneficial role of CSR (McWilliams and Siegel 2011; McWilliams, Siegel, and Wright 2006).

As time passes, however, we expect that the rate of change in improving firm-level CSR is likely to become less pronounced. Such a slower rate of change can be explained by the theory of “underlying progress curves” (Levy 1965). This theory suggests that the firm’s marginal costs of making performance improvements rise up with the absolute value in this measure going up (Rosenthal, Quinn, and Harper 1997; Williams et al. 2005). As such, firms tend to have low-cost opportunities regarding improving performance during the early stage (Chassin 2002). Yet, once such “low-hanging opportunities” have been picked, the marginal cost of the further improvement will increase as can be anticipated (Chatterji and Toffel 2010). Consequently, when making planning and decisions, managers are likely to turn their selective attention to other dimensions, such as turning to improve CSI rather than keep improving CSR. Taken together, we thus posit:

H1: Firm-level CSR will improve more rapidly at first, but the rate of change will become less pronounced as time elapses.

2.3.2 Longitudinal Development of Firm-level CSI in Supply Chain Practices

We then draw upon attention-based theoretical categories in reasoning for the longitudinal trajectory of firm-level CSI. Selective attention serves as a key principle of the attention-based theory (Fiske and Taylor 1991; Ocasio 1997; Simon 1947), suggesting that decision-makers are likely to selectively focus on handling certain issues while ignoring others. Such an attention-

based perspective indicates that decision-makers are limited in the number of issues (Ocasio 1997). Therefore, the decision-makers' selective attention is drawn to the most urgent issues in any particular situation, given that salience is found to be the most significant factor for grabbing selective attention (Fiske and Taylor 1991). As such, we expect firms will exhibit different adoption and implementation behaviors for CSI compared to CSR supply chain practices. The reason is that the beneficial role of socially responsible supply chain practices has been confirmed in different consumer areas (Price and Sun 2017) and thus firms are starkly aware of the importance of CSR through doing good to ensure better publicity, and marketing (Hillman, Zardkoohi, and Bierman 1999; Tate, Ellram, and Kirchoff 2010). As such, firms traditionally emphasize their CSR practices for advertising and marketing, thus CSR is more salience at first (Chen and Ho 2019), which means CSR is more likely to attract firms' attention during the early stage. Thus, firms would first put attention to working on doing good through CSR activities. During this period, CSI is likely to be ignored, given that individuals or organizations allocate their attention sequentially rather than holistically (Greve 2008; Ocasio 1997). In addition, companies don't know what they were doing regarding CSI activities at first, as in the case of looking at how surprisingly quickly firms offshored in the early 2000s (Pierce and Schott 2016). Therefore, we expect that firms' socially irresponsible supply chain incidents (i.e., CSI ratings) will increase rapidly at first.

However, firms would come under pressure as CSI incidents increase because those socially irresponsible behaviors can result in various negative outcomes, such as damaged firm image (Du, Bhattacharya, and Sen 2010) and increased pressures from stakeholders (Price and Sun 2017). In particular, when a firm's CSI activities are exposed, the positive consumer perspective created through its CSR can be interrupted. As a consequence, this firm is likely to

lose revenue, as consumers may boycott its goods or services once being aware of its social irresponsibility (Wagner, Bicen, and Hall 2008). Moreover, the negative effect of CSI behaviors can result in reduced brand value and market value (Du, Bhattacharya, and Sen 2010; Fombrun, Gardberg, and Barnett 2000). Specifically, the increasing CSI behaviors would also give rise to increased pressure from external actors such as the local community, regulators, and other stakeholders (Sweetin et al. 2013). However, the negative effect of CSI will eventually force the companies to reduce CSI behaviors as time goes. This is because organizations begin to develop new integrative knowledge to adjust and correct their actions after repeated negative outcomes resulting from prior behaviors (March, Sproull, and Tamuz 1991). Once developed, this new form of knowledge should coincidentally enable the organizations to improve their performance. Taken together, we thus posit:

H2: Firm-level CSI activities will increase rapidly at first, but then decrease significantly as time elapses.

2.3.3 Moderation Effect of Firm Size on Rates of Change regarding CSR

For firm-level CSR, we expect that larger firms compared to smaller peers will increase more rapidly during the early stage. First, the economic imperatives that firms face regarding being socially responsible can be one of the most important factors (Kagan, Gunningham, and Thornton 2011). The strongest economic incentive pushing firms toward improving socially responsible practices can be meeting customers' service expectations (Thornton, Kagan, and Gunningham 2009). For example, customers' demanding high levels of social responsibility can be a strong antecedent of doing business (Johnston 2010). The loyalty resulting from positive consumer perception of firm social responsibility yields more financial benefits for larger firms than their smaller peers (Genn 1993; Yan, Van Rooij, and Van der Heijden 2015). Specifically,

larger firms are more aware of their firm image than smaller peers, as firm reputation and brand are valuable as an integral part of firm assets for larger firms. Second, larger firms are likely to be under greater institutional pressure from outside than their smaller peers, given they are more visible (Baron, Mittman, and Newman 1991; Goodstein 1994; Ingram and Simons 1995). Third, the strength of social pressures from environmental and social activists, various media, and other social groups is likely to be higher for larger firms than for smaller firms (Gunningham, Kagan, and Thornton 2004). Additionally, larger firms are more likely to benefit from socially responsible practices given that CSR can generate “moral capital from stakeholders in the form of customer credibility, brand faith, employee affective commitment, community legitimacy, supplier trust, and shareholder reliability” (Price and Sun 2017, p. 86). Thus, larger firms will have more incentive to increase socially responsible practices than their smaller peers at first. In particular, larger firms are more likely to have the resource capabilities to take action rapidly, such as hiring specialists. As time passes, however, larger firms have less room to increase CSR practices because they eventually run out of opportunities, while smaller firms with the worst initial CSR have low marginal cost opportunities for increasing CSR. We thus expect that larger firms will display less pronounced rates of change than smaller peers regarding CSR, as time elapsed. Taken together, we posit:

H3: Larger firms compared to smaller ones will show more rapid rates of increase regarding CSR activities during the early stage, however, as time elapse, larger firms will display less pronounced rates of increase in CSR practices than smaller firms.

2.3.4 Moderation Effect of Firm Size on Rates of Change regarding CSI

Regarding the moderation effect of size on the rates of change for firm-level CSI, we expect that larger firms will expose higher increased rates of CSI activities than smaller firms at first,

however, as time elapses, larger firms will exhibit more rapid reduced rates of CSI than smaller firms. On the one hand, companies don't know what they were doing regarding CSI activities at first, as in the case of looking at how surprisingly quickly firms offshored in the early 2000s (Pierce and Schott 2016). On the other hand, larger firms are more visible than smaller peers, and thus socially irresponsible behaviors of larger firms are more likely to be exposed compared to those of smaller peers. Moreover, larger firms are more likely to be under greater scrutiny and institutional pressure than smaller peers (Baron, Mittman, and Newman 1991; Goodstein 1994; Ingram and Simons 1995). Consequently, external actors, such as activists, the media, and inspectors, are more likely to target larger firms because of their disproportionate effect on societal welfare (Genn 1993). Larger firms' CSI behaviors, thus are more likely to be exposed than smaller firms' (Kagan, Gunningham, and Thornton 2011; Ji and Weil 2015). Therefore, larger firms will have more pronounced increased CSI behaviors than smaller peers during the early stage. However, as time passes, we expect larger firms will be more aware of reducing CSI behaviors and thus display more rapidly decreased CSI than smaller firms. This is because once larger firms recognize that the increased CSI brings more firm idiosyncratic risks (Orlitzky and Benjamin 2001), they are more likely to put effort to reduce CSI behaviors. Specifically, larger firms have the resource capability to generate better knowledge that allows them to make improvements rapidly. In contrast, smaller firms have fewer incentives to reduce CSI activities as they are less visible. Taken together, we thus posit:

H4: Larger firms compared to smaller ones will show more rapid rates of increase regarding CSI activities during the early stage, however, as time elapse, larger firms will display more pronounced decreased CSI behaviors than smaller firms.

2.4 Research Setting and Data

2.4.1 Research Setting

We tested our study's hypotheses in the context of MSCI ESG KLD 2003-2018. KLD socially responsible ratings provide the longest time series of firms' environmental and social sustainability information (Chatterji, Levine, and Toffel 2009). KLD ratings for measuring firms' environmental, social, and governance performance were assigned based on the company's CSR reports and other relevant public information, which is released yearly (KLD 2018). Corporate behaviors are rated across seven dimensions: governance, community, diversity, employee relations, environment, human rights, and product quality. For each dimension, KLD ratings consist of paired items. Each such a paired item has both a strength and concern indicator, which is binary taking values 0 or 1. A score of 1 in a strength indicator indicates that the firm has a positive behavior in complying with the social responsibility standards, whereas a score of 1 in a concern indicator indicates the firm has a negative activity (i.e., social irresponsibility practice) that can be considered as a weakness to meet the standards of social responsibility.

Based on the findings in essay 1, we analyzed how manufacturing firm-level CSR and CSI evolved from 2003 to 2018, and how firm size moderated this evolution. We evaluated hypotheses about nonlinear change occurring for a longer overall time frame by adopting specifying piecewise latent trajectory models (Flora 2008; Li, Duncan, and Hops 2001). In so doing, we followed a similar approach used by other scholars in the existing literature (e.g., Flora 2008; Miller, Fugate, and Golicic 2017; Miller, Schwieterman, and Bolumole 2018; Miller, Bolumole, and Schwieterman 2020). In the sections that follow, we provide details of our methodology.

2.4.2 Data Sources

Between 1991 and 2000, KLD ratings focused on the largest 650 publicly-traded companies that belonged to the Domini 400 Social Index³ and/or to the S&P500 Index. In 2001, KLD ratings extended the coverage to consist of the top 1000 publicly-traded US companies by market capitalization. Since 2003, the coverage of KLD ratings expanded to include the top 3000 publicly-traded US companies by market capitalization, and this coverage went on to include the top 4,000 later on. Considering that, limited companies have been covered before 2003. We focus on the panel data from 2003 to 2018 for hundreds of publicly traded manufacturing, wholesale, and retailing firms included in KLD ratings. Once we obtained the KLD data, we merged it with financial data from COMPUSTAT, market concentration from US Census Bureau Economic Indicators, and Upstreamness measuring of the average distance from final use from American economic review (Antràs et al. 2012). To provide a better view of the datasets used in this manuscript, Table 2.1 provides all the datasets, using variables from them, and their sources.

Table 2.1 Datasets and their Sources

Data	Using Variables	Source
KLD data	ESG performance indicators	MSCI ESG KLD 2003-2018
Financial data	the current value of investment, cost of income, cost of investment, revenue, cost of goods sold, total assets, R&D expenditure, capital expenses	Standard & Poor's Compustat 2003-2018
Market concentration	HHI	US Census Bureau Economic Indicators 2003-2018
Upstreamness	How far is the product from final use (i.e., the average distance from final use)	American economic review (Antràs et al. 2012)

2.4.3 Sample

Firm-level CSR and CSI scores are the log-logistic IRT-scaled scores by adopting unipolar item response models (Lucke 2013, 2015) in essay 1. In this essay, we focus on manufacturing firms. This is because the Upstreamness of each wholesale and retailing firm is equal to one, which

makes it not interpretable for firms in these two industries. We set up a filter that requires a firm's observations no less than 8 repeats, which results in $N=689$. $N_t = 8,899$. Table 2.2 is sample statistics and correlation matrix for the manufacturing firms that we used for latent growth analysis.

Table 2.2 Sample Statistics and Correlation Matrix

Measure	1	2	3	4	5
1. CSR	1.00				
2. LnSize	0.46	1.00			
3. LnHHI	0.09	0.03	1.00		
4. Dur-Manufac	-0.13	0.01	-0.00	1.00	
5. Upstreamness	-0.09	0.03	0.07	0.04	1.00
Obs.	8,899	8,887	7,246	8,899	8,899
Mean	1.42	6.47	6.23	0.63	2.10
Standard deviation	1.60	1.64	0.80	0.48	0.87

Measure	1	2	3	4	5
1. CSI	1.00				
2. LnSize	0.49	1.00			
3. LnHHI	0.09	0.03	1.00		
4. Dur-Manufac	-0.14	0.01	-0.00	1.00	
5. Upstreamness	0.02	0.03	0.08	0.04	1.00
Obs.	8,899	8,887	7,246	8,899	8,899
Mean	1.38	6.47	6.23	0.63	2.10
Standard deviation	1.65	1.64	0.80	0.48	0.87

** Note. $N=689$. $N_t = 8,899$ manufacturing firms included in the analysis of CSR/CSI.*

2.4.4 Definition of Variables

We examine firm-level CSR and CSI supply chain practices separately. Our dependent variables are firm-level CSR and CSI, which are measured by CSR and CSI scores, which are the results of essay 1 by using multidimensional unipolar item response theory (UIRT) models (Lucke 2013, 2015).

One of the key independent variables measures is the passage of time in this study, which we specified two piecewise denoted as Occasion_1 and Occasion_2. Occasion_1, the first piecewise linear slope, captures the rates of change during the early stage. Occasion_2, the second piecewise linear slope, captures the rates of change after the first period. Table 2.3 shows metrics coding for measuring occasions of CSR and CSI. The moderating variable Size has been measured by the natural logarithm of the cost of goods sold from Compustat.

Table 2.3 Metrics Coding for Measurement Occasions of CSR/CSI

Year	Occasions	Intercept	CSR		CSI	
			1 st Period	2 nd Period	1 st Period	2 nd Period
2003	0	1	0	0	0	0
2004	1	1	1	0	1	0
2005	2	1	2	0	2	0
2006	3	1	3	0	3	0
2007	4	1	4	0	4	0
2008	5	1	5	0	5	0
2009	6	1	6	0	5	1
2010	7	1	7	0	5	2
2011	8	1	8	0	5	3
2012	9	1	8	1	5	4
2013	10	1	8	2	5	5
2014	11	1	8	3	5	6
2015	12	1	8	4	5	7
2016	13	1	8	5	5	8
2017	14	1	8	6	5	9
2018	15	1	8	7	5	10

** Note. Readers are referred to Flora (2008) for a more detailed treatment.*

This study has controlled for several control factors that may impact firms' (1) average CSR and CSI scores, (2) rate of change in CSR and CSI as time passes, and (3) extent of change in CSR and CSI for reducing concerns of "endogeneity" and obtaining more precise parameter estimations. Our work also includes a measure that captures market concentration adopted from US Census Bureau Economic Indicators (United States Census Bureau 2021). We further control

for firms' Upstreamness using a measurement of the average distance from final use from the American economic review (Antràs et al. 2012). We also control for whether firms' products are durable or nondurable given differences in products' durability are expected to affect firms' incentives to improve their CSR and CSI supply chain practices.

2.5 Methods and Results

2.5.1 Methods

We tested our hypotheses by specifying a piecewise linear latent growth model (e.g., Bollen and Curran, 2006; Flora 2008) and adopting a series of mixed-effects models to estimate our time-series of measures (Fitzmaurice, Laird, and Ware 2012; Ployhart, Holtz, and Bliese 2002). This is because mixed-effects models allow us to recover a large number of covariates more easily to test our hypotheses (Singer, Willett, and Willett 2003). We specified an identical model for firm-level CSR and CSI separately. Following the suggestions of Singer, Willett, and Willett (2003) and Fitzmaurice, Laird, and Ware (2012), our work first fitted a mixed-effects model to interconnect our two measures of CSR/CSI with "Occasion" whereby our model was designed to [1] map the parameter estimations well onto our theorizing, [2] make it easy to interpret the parameter estimations, and [3] efficiently recreate the observed covariance matrix. Adopting Cudeck's (1996) a system of notation, we specified the following model to test *H1* and *H2*:

$$CSR/CSI\ Scores_{it} = \beta_{0i} + \beta_{1i}Occasion_1_{it} + \beta_{2i}Occasion_2_{it} + e_{it} \quad [1]$$

$$\beta_{0i} = \beta_0 + b_{0i} \quad [2]$$

$$\beta_{1i} = \beta_1 + b_{1i} \quad [3]$$

$$\beta_{2i} = \beta_2 + b_{2i} \quad [4]$$

where *i* indexes each firm and *t* indexes each occasion of measurement. In Equations 1–4, we applied a two-piece linear-increment spline model, which has often been used in latent growth modeling literature to interpret how change happens as a "nonlinear function of time" (Newsom

2015). As for the treatment details, readers are referred to Flora (2008) for more information. Equation 1 includes time-varying covariates that occur in the “Level-1”, as in our case, such time-varying covariates have been represented as *Occasion_1* and *Occasion_2*. Our regression parameter estimations in Equation 1 have been indexed by *i* to indicate that the parameter estimations vary from firm to firm. Next, our inclusion of Equations 2–4 allows us to interpret time-varying covariates that occur in the “Level-2”. The design matrix of our model has been coded as that β_0 represents the average *CSR/CSI* scores across all firms on the 1st measurement occasion (i.e., 2003), β_1 indicates the average change in *CSR/CSI* scores with *Occasion_1* increasing one unit, and β_2 indicates the average change in *CSR/CSI* scores with *Occasion_2* increasing one unit. As such, β_1 captures the rate of change during the early stage, while β_2 captures the rate of change as time passes. For *CSR*, H_1 predicts that β_1 will be positive and β_2 will be no larger than β_1 , whereas H_1 , for *CSI*, predicts that β_1 will be positive while β_2 will be negative.

Next, we focus on testing $H3$ and $H4$ regarding how *Size* moderates firms’ rates of change in *Occasion_1* and *Occasion_2* by extending our previous model by incorporating the time-invariant covariates. We specified the full model as:

$$CSR/CSI\ Scores_{it} = \beta_{0i} + \beta_{1i}Occasion_1_{it} + \beta_{2i}Occasion_2_{it} + e_{it} \quad [5]$$

$$\beta_{0i} = \beta_0 + \gamma_1Size_i + \gamma_2LnHHI_i + \gamma_3Dur - Manufac_i + \gamma_4Upstreamness_i + b_{0i} \quad [6]$$

$$\beta_{1i} = \beta_1 + \varphi_1Size_i + b_{1i} \quad [7]$$

$$\beta_{2i} = \beta_2 + \varphi_2Size_i + b_{2i} \quad [8]$$

Equation 6 incorporates factors that can influence firms’ average *CSR/CSI* scores in our theory. Equation 7 incorporates our theorized predictors that can have moderating effects on firms’ rates of change in *Occasion_1*. Equation 8 incorporates our theorized predictors that can have moderating effects on firms’ rates of change in *Occasion_2*. Given that $H3$ predicts that

larger firms' *CSR* scores will improve more rapidly in *Occasion_1* but less pronounced in *Occasion_2*, we expect φ_1 will have a positive sign, and φ_1 will be larger than φ_2 . Given that *H4* predicts that larger firms' *CSI* scores will increase more than smaller peers in *Occasion_1* but reduced more rapidly in *Occasion_2*, we expect φ_1 will have a positive sign, while φ_2 will be negative, as *H2* predicts β_1 will be positive but β_2 will be negative for *CSI*.

2.5.2 Results

Our analysis started by using the mixed-effects model in Equations 1–4 via STATA to freely estimate the cut-off between *Occasion_1* and *Occasion_2*, as our purpose was to fit a nonlinear mixed-effects model (Cudeck and Harring 2007). As with any structural equation model analysis, model fitness detection serves as a critical part of evaluating model estimating. As in our case of evaluating the piecewise latent trajectory model, we tested the cut-off between *Occasion_1* and *Occasion_2* for *CSR/CSI* scores separately. For each calibration model, we computed several model-data fit indices: simulated loglikelihood of the fitted model, and the root mean squared error of approximation (RMSEA). By testing the fitness, for *CSR*, *Occasion_1* (the first linear slope represented latent change) is from 2003 to 2011; *Occasion_2* (the second linear slope represented latent change) is from 2011 to 2018; and the intercept, or status, factor represented the level of *CSR* scores is in 2003. For *CSI*, *Occasion_1* (the first linear slope represented latent change) is from 2003 to 2008; *Occasion_2* (the second linear slope represented latent change) is from 2008 to 2018; and the intercept, or status, factor representing the level of *CSI* scores is in 2003 (see Flora 2008 for details). The results for testing *H1* and *H2* have been reported in Table 2.4.

Table 2.4 Results for Testing H1 – H2 for Firm-level CSR/CSI

Parameter	Label	CSR		CSI		
		Coef.	SE.	Coef.	SE.	
<i>Fixed Effects</i>						
Intercept	β_0	0.83**	0.03	1.05**	0.06	
Occasion_1	β_1	0.07**	0.01	0.13**	0.01	
Occasion_2	β_2	0.01	0.01	-0.21**	0.01	
<i>Variance Components</i>						
Var. Intercept	$\sigma_{b_{0i}}^2$	0.51**	0.04	2.10**	0.02	
Var. Occasion_1	$\sigma_{b_{1i}}^2$	0.02**	0.00	0.05**	0.00	
Var. Occasion_2	$\sigma_{b_{2i}}^2$	0.08*	0.01	0.17**	0.01	
Corr. Intercept, Occasion_1	$\rho_{b_{0i},b_{1i}}$	0.03	0.06	0.13**	0.01	
Corr. Intercept, Occasion_2	$\rho_{b_{0i},b_{2i}}$	-0.18**	0.01	-0.39**	0.02	
Corr. Occasion_1, Occasion_2	$\rho_{b_{1i},b_{2i}}$	-0.54**	0.00	-0.94**	0.03	
Var. Residual	σ_e^2	0.65**	0.01	-0.36**	0.02	
<i>Model Fit</i>						
-2 Log Likelihood		25002		51396		
Wald		$\chi^2(2) = 189$		$\chi^2(2) = 152$		
Likelihood Ratio test vs. linear model		$\chi^2(6) = 8149$		$\chi^2(6) = 13191$		

* Notes: $\dagger = p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$ (two-tailed).

* All models were estimated using quasi-likelihood with mixed-effect models via STATA.

* For CSR, Occasion_1 is from 2003 to 2011, and Occasion_2 is from 2012 to 2018, while for CSI, Occasion_1 is from 2003 to 2008, and Occasion_2 is from 2009 to 2018. It was fixed in these analyses for simplicity.

* "Var." interprets the variance of a random effect, while "Corr." interprets the correlation between two random effects.

For CSR, β_1 is positive and statistically significant and β_1 is larger than β_2 , indicating that firm-level CSR practices will increase more rapidly during the early stage, but this rate of increase tapered over time. These results align with the prediction of H1. For CSI, as shown in Table 2.4, β_1 is positive and statistically significant and β_2 is negative and statistically significant, indicating that firm-level CSI behaviors will exhibit worse performance during the early stage, but improve more rapidly as time passes. As such, such results are consistent with H2. We have visualized our predictions by plotting out the model-implied trajectories for CSR and CSI in Figures 2.1 and 2.2, respectively. Figure 2.1 reveals that firm-level CSR was improving consistently, and the rate of increase was more rapid during the early stage, but this rate of increase tapered over time. Figures 2.2 displays that firms' CSI scoring was increasing

during the early stage, but got started to decrease as time passes. It indicates that firm-level *CSI* exhibited worse performance during the early stage, but improve rapidly as time elapses. Taken together, these results provide strong support for *H1* and *H2*.

Figure 2.1 Quasi-likelihood Estimated Means and Model-implied Means for Each Measurement Occasion for the Model with the Lowest -2 Log Likelihood for CSR

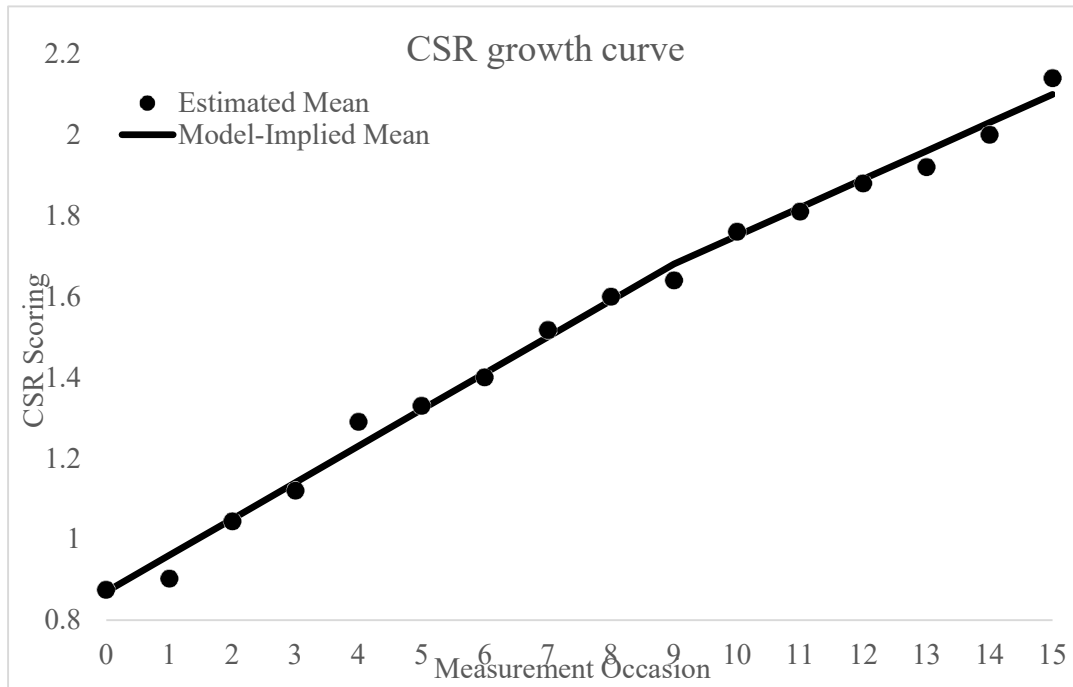
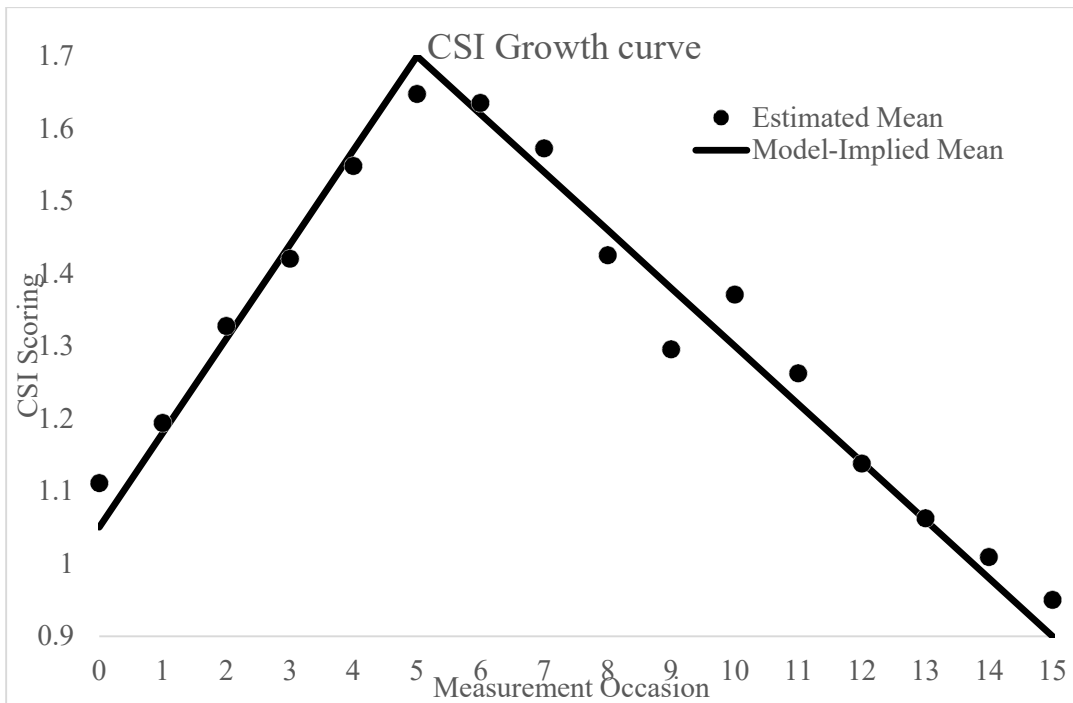


Figure 2.2 Quasi-likelihood Estimated Means and Model-implied Means for Each Measurement Occasion for the Model with the Lowest -2 Log Likelihood for CSI



In addition, we have conducted significance tests of between-subject random effects by examining the correlation between b_{0i} and b_{2i} (Ployhart, Holtz, and Bliese 2002). Positive values of b_{0i} in Equation 2 indicate a firm had better performance for CSR, however, it indicates a firm had a worse performance for CSI than average in 2003. Positive values of b_{1i} in Equation 4 indicate a firm had better performance than average during the first period for CSR, but it indicates a firm had a worse performance for CSI than average during the first period. Positive values of b_{2i} in Equation 4 indicate a firm had better performance than average during the second period for CSR, but it indicates a firm had a worse performance for CSI than average during the second period. Consistent with our expectations, $\rho_{b_{0i},b_{2i}}$ is negative and statistically significant for *CSR* ($\rho_{b_{0i},b_{2i}} = -0.18, z = 0.01$), but $\rho_{b_{0i},b_{1i}}$ is positive and insignificant ($\rho_{b_{0i},b_{1i}} = 0.03, z = 0.06$). Meanwhile, $\rho_{b_{0i},b_{2i}}$ is negative and statistically significant for *CSI* ($\rho_{b_{0i},b_{2i}} = -0.39, z = 0.02$), but $\rho_{b_{0i},b_{1i}}$ is positive and significant ($\rho_{b_{0i},b_{1i}} = 0.13, z = 0.01$). Given the coding of *CSR/CSI* measures and the specification of our model, such results indicate that firms with poor initial performance tend to have more pronounced improvement as time passes.

Regarding how size moderates firms' *CSR* and *CSI* evolution over time, we reported the results in Table 2.5. Looking first at *CSR*, φ_1 is statistically significant and positive, and φ_2 is statistically significant and negative. This is consistent with the prediction of H_3 . Such results indicate that larger firms compared with smaller peers make more rapid improvements during the first period, however, such improvements become less pronounced as time goes by. As for *CSI*, φ_1 is statistically significant and positive, and φ_2 is statistically significant and negative. This is consistent with the prediction of H_4 , indicating that larger firms would exhibit worse performance in CSI during the early stage, but make more rapid improvement as time passes. To

better understand how the effects of *Occasion_1* and *Occasion_2* on *CSR* and *CSI* have been moderated by *Size*, we have visualized our predictions in Figures 2.3 & 2.4 by plotting out the model-implied effects of *Occasion_1* and *Occasion_2* on *CSR* and *CSI* for small (10th percentile), medium (50th percentile), and large (90th percentile) manufacturing firms. Figure 2.3 reveals that larger firms' *CSR* activities were improving consistently, and the rate of improvement was more rapid during the early stage, while smaller firms with poor initial *CSR* scores have a more pronounced rate of increase later on. Figures 2.4 displays that larger firms' *CSI* scoring was increasing during the early stage, but got started to decrease rapidly as time passes. However, smaller firms' *CSI* behaviors have not changed much. This indicates that larger firms exhibited worse performance in *CSI* during the early stage given they are more exposed, but improve rapidly as time elapses, while smaller ones have no incentive to improve their *CSI* practices as they are less visible. Taken together, these results provide strong support for *H3* and *H4*.

Table 2.5 Results from Testing of Moderating Effects of *Size* on Firms' Rates of Change in Occasion_1 and Occasion_2 for CSR and CSI

Parameter	Label	CSR		CSI	
		Coef.	SE.	Coef.	SE.
<i>Fixed Effects</i>					
Intercept	β_0	1.03**	0.07	1.46**	0.07
Occasion_1	β_1	0.08**	0.00	0.12**	0.01
Occasion_2	β_2	-0.05**	0.01	-0.22**	0.01
Size	γ_1	0.05*	0.02	0.17**	0.00
Dur_Manufc	γ_2	-0.34**	0.08	-0.44**	0.08
LnHHI	γ_3	0.02	0.03	-0.03	0.04
Upstreamness	γ_4	-0.29**	0.09	0.08	0.10
Occasion_1× Size	φ_1	0.04**	0.00	0.08**	0.00
Occasion_2× Size	φ_2	-0.05**	0.01	-0.15**	0.01
<i>Model Fit</i>					
-2 Log Likelihood		88160		85730	
Wald		$\chi^2(8) = 1259$		$\chi^2(8) = 1964$	
Likelihood Ratio test vs. linear model		$\chi^2(1) = 3029$		$\chi^2(1) = 6055$	

$\dagger = p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$ (two-tailed).

All models were estimated using quasi-likelihood with mixed-effect models via STATA.

For CSR, Occasion_1 is from 2003 to 2011, and Occasion_2 is from 2012 to 2018, while for CSI, Occasion_1 is from 2003 to 2008, and Occasion_2 is from 2009 to 2018.

Figure 2.3 Plot of the Moderating Effects of *Size* for Each Measurement Occasion for the Model with the Lowest -2 Log Likelihood for *CSR*

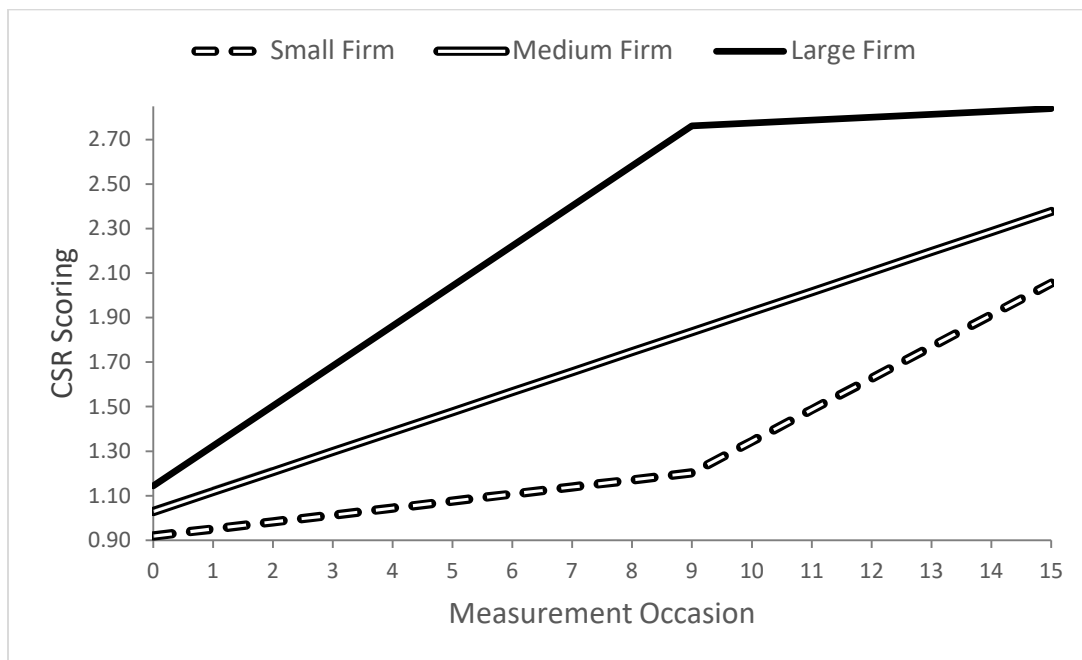
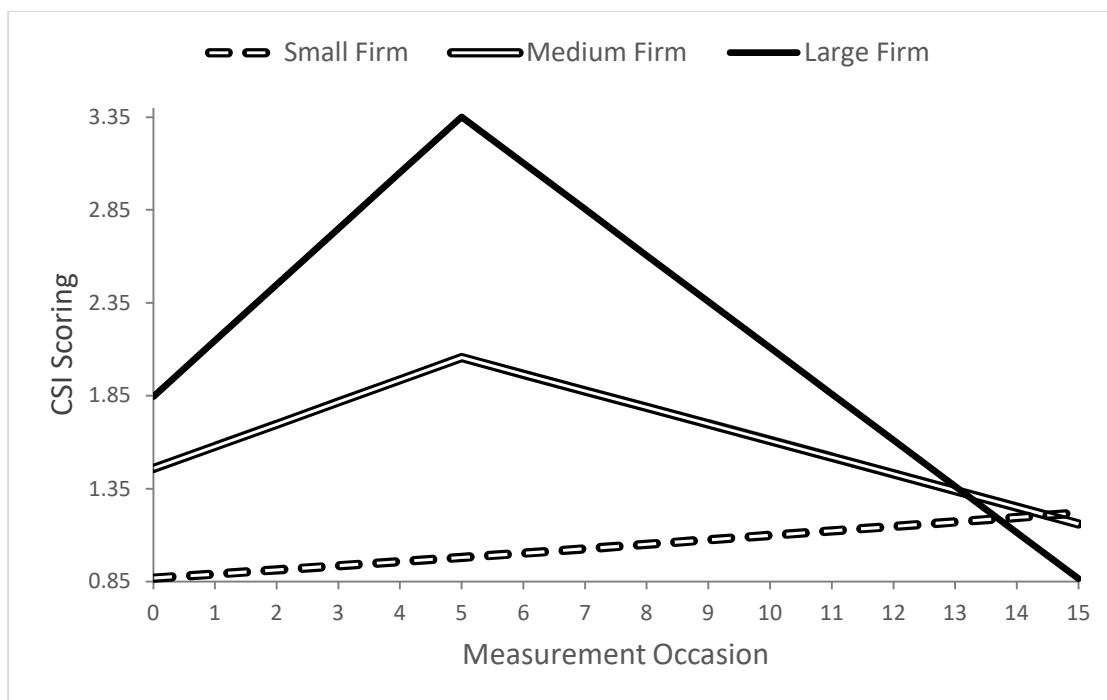


Figure 2.4 Plot of the Moderating Effects of *Size* for Each Measurement Occasion for the Model with the Lowest -2 Log Likelihood for *CSI*



2.6 Discussion

2.6.1 Theoretical and Empirical Contributions

Our work makes theoretical and empirical contributions to the body of knowledge, regarding the longitudinal examinations of firms' supply chain sustainability, which has been limited by the numerous weaknesses of the prior approaches in the existing literature (Silvestre et al. 2020). We particularly focus on examining how firms' sustainability performance evolved over time, and what factors affect this evolution by separating firms' CSR and CSI supply chain practices. This responds to Matusik, Hollenbeck, Mitchell's (2021) call for organizational researchers to put more attention to investigating longitudinal phenomena, and particularly echoes Ketokivi and McIntosh's (2017) call for SCM scholars to examine how subjects' performance evolves over time. The findings of this research and its expansion to supply chain sustainability open up multiple important avenues for further exploration of supply chain sustainability evolution. To be more specific, our findings contribute to the literature in several ways. First, this study elaborates on the theory of firm-level supply chain sustainability by examining details and highlighting empirical insights on how firms learn and collectively apply the knowledge to improve their socially responsible supply chain practices. More specifically, we find evidence that improving social responsibility is a fundamental process for a firm to be aware of the potential opportunity or threat in the business or regulatory environment, have the motivation to respond to opportunities or threats, and then develop the capability to effectively respond to changes in the environment by adapting them (Chen and Miller 2012). As a newly elaborated theory, our work calls for further large-scale empirical research that further examines the dynamics behind the evolution of organizations' longitudinal performance. Naturally, our findings can be well understood and leveraged to investigate organizations' longitudinal evaluation more broadly.

Second, this manuscript provides insights into the longitudinal examinations of firms' sustainability performance by splitting apart CSR and CSI in supply chain practices. Because investigating social responsibility by the inclusion of CSR and CSI can provide a broader, holistic viewpoint that enables researchers and practitioners to look at firms' regimes with new respect, such that outline better strategies involving socially responsible and irresponsible practices (Jones, Bowd, and Tench 2009; Lange and Washburn 2012; Sweetin, Knowles, Summey, and McQueen 2013). In particular, researchers have traditionally and primarily looked at the CSR perspective of sustainability and developed the logic and definition of social responsibility as an overall construct that was measured unidimensionally within the framework (Griffin and Mahon 1997). Such a unidimensional perspective is likely to limit our understanding of corporate social responsibility performance as it is usually comprised of responsible and irresponsible aspects of practices (e.g., Castillo, Mollenkopf, Bell, and Bozdogan 2018; Price and Sun 2017). Specifically, companies could involve CSR and CSI activities simultaneously in terms of doing both "good" and "bad" (Lenz, Wetzel, and Hammerschmidt 2017; Mattingly and Berman 2006; Strike, Gao, and Bansal 2006). For example, Nike has invested substantial resources in helping suppliers improve production processes yet has one or more suppliers engage in bad behavior (Distelhorst, Hainmueller, and Locke 2017). As such, the theorizing we proposed makes a theoretical contribution given its greater consilience (Thagard 1978) in comparison with the previous explanations (Chen, Su, and Tsai 2007) regarding how firms respond to social responsibility over time. Relying on core concepts (Meehl 1990) from a more general theory, our theorizing not only interprets why the rate of change in socially responsible supply chain practices tends to be pronounced following external pressures, but also accounts for why the rate of change is likely to be tapered.

Collectively, this research directly adds to previous research in supply chain sustainability by developing and testing a nuanced theory via specifying piecewise latent growth models. As such, we contribute to the literature by juxtaposing previous research in the assessment of social and environmental sustainability perception (e.g., Vincenzi et al. 2018), strategic management (e.g., Chen and Miller 2012), and supply chain sustainability (e.g., Chen and Ho 2019; Fernandes and Bornia 2019). By doing so, we developed new theoretical insights into several factors that motivate managers to improve a firm's sustainability performance. By integrating the AMC framework from the strategy literature, we explored how a firm considers industry-and firm-level conditions in its decision to improve performance. While scholars have considered other theoretical perspectives to study sustainability performance, previous research has not considered the AMC perspective to theorize on how and why a firm decides to improve its sustainability performance.

Additionally, our application of growth modeling techniques in understanding the longitudinal development of firm-level CSR and CSI in supply chain practices can provide important implications for empirical inquiry in both firms' sustainability domain and broader work focusing on understanding how firms respond to institutional pressure. In particular, we utilize a specifying piecewise latent growth technique to provide evidence that the passage of time plays an important role in longitudinal phenomena, which received limited attention in organizational research (Ancona et al. 2001; Miller, Ganster, and Griffis 2018). This technique enables researchers to explicitly examine a topic of substantial interest regarding the heterogeneity of firms' responses to institutional pressure (Delmas and Toffel 2010; Doshi, Dowell, and Toffel 2013). Moreover, such a technique allows each firm under investigation to have its own trajectory concerning how a dependent variable of interest changes as a function of

time (Singer, Willett, and Willett 2003). This, in turn, sheds light on the investigation of nuanced moderation effects such as “how organizational and environmental factors moderate rates of change on outcomes and whether these moderation effects continue to hold as time passes” (Miller, Fugate, and Golicic 2017, p. 1034).

This study has implications for managers to process their learning loops more efficiently regarding accumulating knowledge, applying the knowledge, and engaging in social responsibility practices to address better performance in supply chain sustainability. Specifically, decision-makers can be more attentive to balance their selective attention when implementing CSR and CSI practices, which may have different effects on their supply chain sustainability. Policy-makers can gain insights into the empirical relationship between CSR and CSI in firms’ socially responsible practices. More specifically, policy-makers can foster sustained performance improvement by getting insight into the mechanisms 1) how decision-makers change their selective attention over time, 2) whether/how initial performances affect firms’ rate of improvement regarding their longitudinal trajectories, and 3) what factors moderate the change of behaviors over time.

2.6.2 Limitations and Suggestions for Further Research

It is important to note that there are limitations to the present study. First, the time span for this study includes 16 years starting from 2003 to 2018. Such a range can be extended further based on the data available. Second, there is a limitation regarding the number of firms examined. This is because the ratings of certain firms in the KLD database are not consistent over the time span. In other words, some firms incorporated in the KLD participate a few times or participate irregularly over the defined timespan. Thus, we exclude these firms with limited observations over our defined timespan given that they are likely to add noise to our longitudinal examination

(Berg et al. 2020). Third, in this study, we focus on US manufacturing firms. Thus, we hesitate to generalize our findings to firms operating in other countries. Additionally, similar to most research that relies on statistical models to test theoretical predictions (Freedman 1991), this study does not directly observe if the processes theorized are operating in a way that brings about the posited relationships. However, the fact our theory accounts for a wide array of findings reduces concerns that an alternative explanation can better account for our set of empirical findings (Lipton 2003).

This research can be extended in multiple directions. Within the manufacturing industry, one avenue is to test the inter-dynamic relationship between firm-level CSR and CSI supply chain practices. A second direction would be to study how the changes in output, capital intensity, and change in R&D intensity would affect firms' supply chain sustainability performance over time. A third direction is to investigate how firms' sustainability performance interacts with the competition they face. Since economic data suggests that industry concentration and total sales for the sector, both nominal and potentially deflated (United States Census Bureau 2021), the challenges with sustainability as firms are facing competition, are whether they going to invest in this. Another avenue is to look at how the incentive structure as a moderator would affect firms' sustainable development over time because the literature suggests incentive structure matters a lot (Mukandwal et al. 2020; Jadhav, Orr, and Malik 2019; Silvestre et al. 2020). For example, look at whether such an interaction with incentives makes sustainability issues even more problematic or how industry-level imports affect firms' sustainability performance over time. Additionally, further research can be conducted by examining how either firm-level or industry-level productivity affects firms' investment in sustainability.

REFERENCES

REFERENCES

- Ahi, P., & Searcy, C. (2013). A comparative literature analysis of definitions for green and sustainable supply chain management. *Journal of cleaner production*, 52, 329-341.
- Ahmadi, H. B., Kusi-Sarpong, S., & Rezaei, J. (2017). Assessing the social sustainability of supply chains using Best Worst Method. *Resources, Conservation and Recycling*, 126, 99-106.
- Ancona, D. G., Goodman, P. S., Lawrence, B. S., & Tushman, M. L. (2001). Time: A new research lens. *Academy of management Review*, 26(4), 645-663.
- Antràs, P., Chor, D., Fally, T., & Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review*, 102(3), 412-16.
- Armstrong, J. S., & Green, K. C. (2013). Effects of corporate social responsibility and irresponsibility policies. *Journal of Business Research*, 66(10), 1922-1927.
- Bai, C., Kusi-Sarpong, S., & Sarkis, J. (2017). An implementation path for green information technology systems in the Ghanaian mining industry. *Journal of Cleaner Production*, 164, 1105-1123.
- Barnett, M. L. (2007). Stakeholder influence capacity and the variability of financial returns to corporate social responsibility. *Academy of management review*, 32(3), 794-816.
- Baron, J. N., Mittman, B. S., & Newman, A. E. (1991). Targets of opportunity: Organizational and environmental determinants of gender integration within the California civil service, 1979-1985. *American Journal of Sociology*, 96(6), 1362-1401.
- Becker-Olsen, K. L., Cudmore, B. A., & Hill, R. P. (2006). The impact of perceived corporate social responsibility on consumer behavior. *Journal of business research*, 59(1), 46-53.
- Bhattacharya, C. B., & Sen, S. (2004). Doing better at doing good: When, why, and how consumers respond to corporate social initiatives. *California management review*, 47(1), 9-24.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. 467). John Wiley & Sons.
- Busse, C., Kach, A. P., & Bode, C. (2016). Sustainability and the false sense of legitimacy: How institutional distance augments risk in global supply chains. *Journal of Business Logistics*, 37(4), 312-328.
- Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance. *Academy of management review*, 4(4), 497-505.

- Carter, C. R., & Jennings, M. M. (2002). Logistics social responsibility: an integrative framework. *Journal of business logistics*, 23(1), 145-180.
- Castillo, V. E., Mollenkopf, D. A., Bell, J. E., & Bozdogan, H. (2018). Supply chain integrity: A key to sustainable supply chain management. *Journal of Business Logistics*, 39(1), 38-56.
- Chatterji, A. K., & Toffel, M. W. (2010). How firms respond to being rated. *Strategic Management Journal*, 31(9), 917-945.
- Chen, M.J. (1996). Competitor analysis and interfirm rivalry: Toward a theoretical integration. *Academy of Management Review* 21(1) 100-134.
- Chen, Y. S. (2008). The driver of green innovation and green image–green core competence. *Journal of business ethics*, 81(3), 531-543.
- Chen, C. M., & Delmas, M. (2011). Measuring corporate social performance: An efficiency perspective. *Production and operations management*, 20(6), 789-804.
- Chen, C. M., & Ho, H. (2019). Who pays you to be green? How customers' environmental practices affect the sales benefits of suppliers' environmental practices. *Journal of Operations Management*, 65(4), 333-352.
- Chen, M. J., & Miller, D. (2012). Competitive dynamics: Themes, trends, and a prospective research platform. *Academy of management annals*, 6(1), 135-210.
- Chen, M. J., Su, K. H., & Tsai, W. (2007). Competitive tension: The awareness-motivation-capability perspective. *Academy of management Journal*, 50(1), 101-118.
- Chi, L., Ravichandran, T., & Andreovski, G. (2010). Information technology, network structure, and competitive action. *Information systems research*, 21(3), 543-570.
- Chowdhury, M. M. H., & Quaddus, M. A. (2021). Supply chain sustainability practices and governance for mitigating sustainability risk and improving market performance: A dynamic capability perspective. *Journal of Cleaner Production*, 278, 123521.
- Clarkson, M. E. (1995). A stakeholder framework for analyzing and evaluating corporate social performance. *Academy of management review*, 20(1), 92-117.
- Colbert, B. A. (2004). The complex resource-based view: Implications for theory and practice in strategic human resource management. *Academy of management review*, 29(3), 341-358.
- Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate behavioral research*, 31(3), 371-403.
- Cudeck, R., & Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annu. Rev. Psychol.*, 58, 615-637.

- Delmas, M. A., & Montes-Sancho, M. J. (2010). Voluntary agreements to improve environmental quality: Symbolic and substantive cooperation. *Strategic Management Journal*, 31(6), 575-601.
- Delmas, M. A., & Toffel, M. W. (2010). Institutional pressures and organizational characteristics: Implications for environmental strategy. Harvard Business School Technology & Operations Mgt. Unit Working Paper, (11-050).
- De Pelsmacker, P., Driesen, L., & Rayp, G. (2005). Do consumers care about ethics? Willingness to pay for fair-trade coffee. *Journal of consumer affairs*, 39(2), 363-385.
- Distelhorst, G., Hainmueller, J., & Locke, R. M. (2017). Does lean improve labor standards? Management and social performance in the Nike supply chain. *Management Science*, 63(3), 707-728.
- Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of management Review*, 20(1), 65-91.
- Doshi, A. R., Dowell, G. W., & Toffel, M. W. (2013). How firms respond to mandatory information disclosure. *Strategic Management Journal*, 34(10), 1209-1231.
- Du, S., Bhattacharya, C. B., & Sen, S. (2010). Maximizing business returns to corporate social responsibility (CSR): The role of CSR communication. *International journal of management reviews*, 12(1), 8-19.
- Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, 13(4), 668-689.
- Ellram, L. M., & Murfield, M. L. U. (2017). Environmental sustainability in freight transportation: A systematic literature review and agenda for future research. *Transportation Journal*, 56(3), 263-298.
- Esfahbodi, A., Zhang, Y., & Watson, G. (2016). Sustainable supply chain management in emerging economies: Trade-offs between environmental and cost performance. *International Journal of Production Economics*, 181, 350-366.
- Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics*, 162, 101-114.
- Farla, J., Markard, J., Raven, R., & Coenen, L. (2012). Sustainability transitions in the making: A closer look at actors, strategies and resources. *Technological forecasting and social change*, 79(6), 991-998.
- Fernandes, S. M., & Bornia, A. C. (2019). Reporting on supply chain sustainability: Measurement using item response theory. *Corporate Social Responsibility and Environmental Management*, 26(1), 106-116.

- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. McGraw-Hill Book Company.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998). John Wiley & Sons.
- Flora, D. B. (2008). Specifying piecewise latent trajectory models for longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(3), 513-533.
- Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological methodology*, 291-313.
- Fombrun, C. J., Gardberg, N. A., & Barnett, M. L. (2000). Opportunity platforms and safety nets: Corporate citizenship and reputational risk. *Business and society review*, 105(1).
- Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. *Research policy*, 33(6-7), 897-920.
- Geels, F. W. (2014). Reconceptualising the co-evolution of firms-in-industries and their environments: Developing an inter-disciplinary Triple Embeddedness Framework. *Research Policy*, 43(2), 261-277.
- Genn, H. (1993). Business responses to the regulation of health and safety in England. *Law & Policy*, 15(3), 219-233.
- Goodstein, J. D. (1994). Institutional pressures and strategic responsiveness: Employer involvement in work-family issues. *Academy of Management journal*, 37(2), 350-382.
- Greve, H. R. (2003). A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. *Academy of management journal*, 46(6), 685-702.
- Griffin, J. J., & Mahon, J. F. (1997). The corporate social performance and corporate financial performance debate: Twenty-five years of incomparable research. *Business & society*, 36(1), 5-31.
- Gunningham, N., Kagan, R. A., & Thornton, D. (2004). Social license and environmental protection: why businesses go beyond compliance. *Law & Social Inquiry*, 29(2), 307-341.
- Helgesen, Ø. (2006). Are loyal customers profitable? Customer satisfaction, customer (action) loyalty and customer profitability at the individual level. *Journal of Marketing Management*, 22(3-4), 245-266.
- Hillman, A. J., Zardkoohi, A., & Bierman, L. (1999). Corporate political strategies and firm performance: indications of firm-specific benefits from personal service in the US government. *Strategic Management Journal*, 20(1), 67-81.

- Hofer, C., Cantor, D. E., & Dai, J. (2012). The competitive determinants of a firm's environmental management activities: Evidence from US manufacturing industries. *Journal of Operations Management*, 30(1-2), 69-84.
- Ingram, P., & Simons, T. (1995). Institutional and resource dependence determinants of responsiveness to work-family issues. *Academy of Management Journal*, 38(5), 1466-1482.
- Jadhav, A., Orr, S., & Malik, M. (2019). The role of supply chain orientation in achieving supply chain sustainability. *International Journal of Production Economics*, 217, 112-125.
- Ji, M., & Weil, D. (2015). The impact of franchising on labor standards compliance. *ILR Review*, 68(5), 977-1006.
- Johnston, J.S. (2010). The Promise and Limits of Voluntary Management-Based Regulatory Reform: An Analysis of the EPA's Strategic Goals Program. In *Leveraging the Private Sector: Management-Based Strategies for Improving Environmental Performance*, edited by C. Coglianese and J. Nash, 183– 216. New York, NY: Routledge.
- Kagan, R. A., Gunningham, N., & Thornton, D. (2011). Fear, duty, and regulatory compliance: lessons from three research projects. *Explaining Compliance: Business Responses to Regulation*, 37-58.
- Kapelus, P. (2002). Mining, corporate social responsibility and the "community": The case of Rio Tinto, Richards Bay Minerals and the Mbonambi. *Journal of Business Ethics*, 39(3), 275-296.
- Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, 52, 1-14.
- Kube, R., von Graevenitz, K., Löschel, A., & Massier, P. (2019). Do voluntary environmental programs reduce emissions? EMAS in the German manufacturing sector. *Energy Economics*, 84, 104558.
- Kusi-Sarpong, S., Bai, C., Sarkis, J., & Wang, X. (2015). Green supply chain practices evaluation in the mining industry using a joint rough sets and fuzzy TOPSIS methodology. *Resources Policy*, 46, 86-100.
- Langan, R., & Menz, M. (2022). Does Your Company Need a Chief ESG Officer? *Harvard Business Review*. Available at <https://hbr.org/2022/02/does-your-company-need-a-chief-esg-officer> (accessed March 21 2022).
- Levy, F. K. (1965). Adaptation in the production process. *Management Science*, 11(6), B-136.
- Li, F., Duncan, T. E., & Hops, H. (2001). Examining developmental trajectories in adolescent alcohol use using piecewise growth mixture modeling analysis. *Journal of studies on alcohol*, 62(2), 199-210.

- Lipton, P. (2003). *Inference to the best explanation*. Routledge.
- Lucke, J. F. (2013). Positive trait item response models. In *New developments in quantitative psychology* (pp. 199-213). Springer, New York, NY.
- Lucke, J. F. (2015). Unipolar Item Response Models. *Handbook of item response theory modeling: Applications to typical performance assessment*, 272-284.
- Luo, X., & Bhattacharya, C. B. (2006). Corporate social responsibility, customer satisfaction, and market value. *Journal of marketing*, 70(4), 1-18.
- Martin, K. D., Josephson, B. W., Vadakkepatt, G. G., & Johnson, J. L. (2018). Political management, research and development, and advertising capital in the pharmaceutical industry: a good prognosis?. *Journal of Marketing*, 82(3), 87-107.
- Mattingly, J. E., & Berman, S. L. (2006). Measurement of corporate social action: Discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society*, 45(1), 20-46.
- Matusik, J. G., Hollenbeck, J. R., & Mitchell, R. L. (2021). Latent change score models for the study of development and dynamics in organizational research. *Organizational Research Methods*, 24(4), 772-801.
- McPeak, C., Devirian, J., & Seaman, S. (2010). Do environmentally friendly companies outperform the market?. *Journal of Global Business Issues*, 4(1), 61.
- McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of management review*, 26(1), 117-127.
- McWilliams, A., & Siegel, D. S. (2011). Creating and capturing value: Strategic corporate social responsibility, resource-based theory, and sustainable competitive advantage. *Journal of management*, 37(5), 1480-1495.
- McWilliams, A., Siegel, D. S., & Wright, P. M. (2006). Corporate social responsibility: Strategic implications. *Journal of management studies*, 43(1), 1-18.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological inquiry*, 1(2), 108-141.
- Mishra, S., & Modi, S. B. (2013). Positive and negative corporate social responsibility, financial leverage, and idiosyncratic risk. *Journal of business ethics*, 117(2), 431-448.
- Miller, J. W., Bolumole, Y., & Schwieterman, M. A. (2020). Electronic logging device compliance of small and medium size motor carriers prior to the December 18, 2017, mandate. *Journal of Business Logistics*, 41(1), 67-85.

- Miller, J. W., Ganster, D. C., & Griffis, S. E. (2018). Leveraging big data to develop supply chain management theory: The case of panel data. *Journal of Business Logistics*, 39(3), 182-202.
- Miller, J. W., Schwieterman, M. A., & Bolumole, Y. A. (2018). Effects of motor carriers' growth or contraction on safety: A multiyear panel analysis. *Journal of Business Logistics*, 39(2), 138-156.
- Modi, S. B., & Cantor, D. E. (2021). How coopetition influences environmental performance: Role of financial slack, leverage, and leanness. *Production and Operations Management*, 30(7), 2046-2068.
- Mukandwal, P. S., Cantor, D. E., Grimm, C. M., Elking, I., & Hofer, C. (2020). Do firms spend more on suppliers that have environmental expertise? An empirical study of US manufacturers' procurement spend. *Journal of Business Logistics*, 41(2), 129-148.
- Murphy, P. E., & Schlegelmilch, B. B. (2013). Corporate social responsibility and corporate social irresponsibility: Introduction to a special topic section. *Journal of Business Research*, 66(10), 1807-1813.
- Negri, M., Cagno, E., Colicchia, C., & Sarkis, J. (2021). Integrating sustainability and resilience in the supply chain: A systematic literature review and a research agenda. *Business Strategy and the Environment*, 30(7), 2858-2886.
- Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge.
- Ndofor, H. A., Sirmon, D. G., & He, X. (2011). Firm resources, competitive actions and performance: investigating a mediated model with evidence from the in-vitro diagnostics industry. *Strategic Management Journal*, 32(6), 640-657.
- Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic management journal*, 18(S1), 187-206.
- Oliver, C., & Holzinger, I. (2008). The effectiveness of strategic political management: A dynamic capabilities framework. *Academy of management review*, 33(2), 496-520.
- Orlitzky, M., & Benjamin, J. D. (2001). Corporate social performance and firm risk: A meta-analytic review. *Business & Society*, 40(4), 369-396.
- Parmigiani, A., Klassen, R. D., & Russo, M. V. (2011). Efficiency meets accountability: Performance implications of supply chain configuration, control, and capabilities. *Journal of operations management*, 29(3), 212-223.
- Pierce, J. R., & Schott, P. K. (2016). The surprisingly swift decline of US manufacturing employment. *American Economic Review*, 106(7), 1632-62.

- Ployhart, R. E., Holtz, B. C., & Bliese, P. D. (2002). Longitudinal data analysis: Applications of random coefficient modeling to leadership research. *The leadership quarterly*, 13(4), 455-486.
- Price, J. M., & Sun, W. (2017). Doing good and doing bad: The impact of corporate social responsibility and irresponsibility on firm performance. *Journal of Business Research*, 80, 82-97.
- Rai, A., & Tang, X. (2010). Leveraging IT capabilities and competitive process capabilities for the management of interorganizational relationship portfolios. *Information systems research*, 21(3), 516-542.
- Rosenthal, G. E., Quinn, L., & Harper, D. L. (1997). Declines in hospital mortality associated with a regional initiative to measure hospital performance. *American Journal of Medical Quality*, 12(2), 103-112.
- Roy, V., Schoenherr, T., & Charan, P. (2018). The thematic landscape of literature in sustainable supply chain management (SSCM): A review of the principal facets in SSCM development. *International Journal of Operations & Production Management*.
- Russo, M. V., & Fouts, P. A. (1997). A resource-based perspective on corporate environmental performance and profitability. *Academy of management Journal*, 40(3), 534-559.
- Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of cleaner production*, 16(15), 1699-1710.
- Solomon, M. 2014. “ 2015 is the Year of the Millennial Customer: 5 Key Traits These 80 Million Consumers Share.” Available at: <https://www.forbes.com/sites/micahsolomon/2014/12/29/5-traits-that-define-the-80-million-millennial-customers-coming-your-way/#78dd5e525e56>.
- Silvestre, B. S. (2015). Sustainable supply chain management in emerging economies: Environmental turbulence, institutional voids and sustainability trajectories. *International Journal of Production Economics*, 167, 156-169.
- Silvestre, B. S., Silva, M. E., Cormack, A., & Thome, A. M. T. (2020). Supply chain sustainability trajectories: learning through sustainability initiatives. *International Journal of Operations & Production Management*.
- Simon, H. A. (1979). Rational decision making in business organizations. *The American economic review*, 69(4), 493-513.
- Singer, J. D., Willett, J. B., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.
- Srinivasan, S. S., Anderson, R., & Ponnavaolu, K. (2002). Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of retailing*, 78(1), 41-50.

- Strike, V. M., Gao, J., & Bansal, P. (2006). Being good while being bad: Social responsibility and the international diversification of US firms. *Journal of International Business Studies*, 37(6), 850-862.
- Sweetin, V. H., Knowles, L. L., Summey, J. H., & McQueen, K. S. (2013). Willingness-to-punish the corporate brand for corporate social irresponsibility. *Journal of Business Research*, 66(10), 1822-1830.
- Tate, W. L., Ellram, L. M., & Kirchhoff, J. F. (2010). Corporate social responsibility reports: a thematic analysis related to supply chain management. *Journal of supply chain management*, 46(1), 19-44.
- Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The journal of philosophy*, 75(2), 76-92.
- Thornton, D., Kagan, R. A., & Gunningham, N. (2009). When social norms and pressures are not enough: Environmental performance in the trucking industry. *Law & Society Review*, 43(2), 405-436.
- Tseng, M., Lim, M., & Wong, W. P. (2015). Sustainable supply chain management: a closed-loop network hierarchical approach. *Industrial Management & Data Systems*.
- United States Census Bureau. (2021). Business Dynamics Statistics: Annual Report: 2003-2018. Available at: <https://www.census.gov/data/tables/econ/awts/annual-reports.html> (accessed 19 March 2022).
- Vanhamme, J., & Grobben, B. (2009). "Too good to be true!". The effectiveness of CSR history in countering negative publicity. *Journal of Business Ethics*, 85(2), 273-283.
- Vincenzi, S. L., Possan, E., de Andrade, D. F., Pituco, M. M., de Oliveira Santos, T., & Jasse, E. P. (2018). Assessment of environmental sustainability perception through item response theory: A case study in Brazil. *Journal of Cleaner Production*, 170, 1369-1386.
- Wang, H., & Choi, J. (2013). A new look at the corporate social–financial performance relationship: The moderating roles of temporal and interdomain consistency in corporate social performance. *Journal of Management*, 39(2), 416-441.
- Williams, S. C., Schmaltz, S. P., Morton, D. J., Koss, R. G., & Loeb, J. M. (2005). Quality of care in US hospitals as reflected by standardized measures, 2002–2004. *New England Journal of Medicine*, 353(3), 255-264.
- Wood, D. J. (1991). Toward improving corporate social performance. *Business Horizons*, 34(4), 66-74.
- Yan, H., Van Rooij, B., & Van der Heijden, J. (2015). Contextual Compliance: Situational and Subjective Cost-Benefit Decisions about Pesticides by Chinese Farmers. *Law & Policy*, 37(3), 240-263.

Zhang, J. Q., Dixit, A., & Friedmann, R. (2010). Customer loyalty and lifetime value: An empirical investigation of consumer packaged goods. *Journal of marketing theory and practice*, 18(2), 127-140.

CHAPTER 3 - Developing and Testing the Dynamic Inter-relationships between Corporate Social Responsibility and Irresponsibility Supply Chain Practices

3.1 Introduction

Understanding of firms' social responsibility has been increasingly emphasized in the sustainability and supply chain management field (Seuring and Müller 2008; Fahimnia, Sarkis, and Davarzani 2015). Firms' social responsibility consists of not only “doing good through corporate social responsibility (CSR)” activities but also “doing bad through corporate social irresponsibility (CSI)” practices (Price and Sun 2017, p. 82). CSR initiatives have been generally recognized as highly desirable corporate behaviors that both benefit communities and help companies themselves perform better in business (Barnett 2007; McWilliams and Siegel 2001). In comparison, CSI activities, referred to as firms' irresponsible behaviors, have been regarded as harmful practices leading to substantive negative effects that may be harmful to various stakeholders (Armstrong and Green 2013). Researchers have traditionally focused on understanding firms' social responsibility by examining CSR activities, whereas CSI incidents have barely caught researchers' attention and thus have rarely been investigated so far among the existing studies (Murphy and Schlegelmich 2013; Price and Sun 2017). Only recently, has academic literature gotten started to broaden the understanding of firms' socially responsible practices through the inclusion of both CSR and CSI (e.g, Lenz, Wetzel, and Hammerschmidt 2017; Kang, Germann, and Grewal 2016; Price and Sun 2017). Yet, to date, the relationship between CSR and CSI has rarely been investigated (Groening and Kanuri 2013; Lenz, Wetzel, and Hammerschmidt 2017).

In particular, the development of firms' social responsibility and irresponsibility practices has a dynamic nature, meaning that it cannot be achieved overnight but instead, goes through a

complex and dynamic process (Silvestre et al. 2020). Yet, little has been known about the dynamic relationships in the continually evolving firms' CSR and CSI activities as it pertains to the supply chains. Herein, this essay attempts to fill into gaps regarding firms' CSR and CSI by developing and testing a dynamic perspective that allows scholars to leverage the past to predict the response of certain firms' traits to events in the future. We suggest that as a complex, dynamic system, a firm's social responsibility and irresponsibility behaviors in the future are contingent on its past activities (Silvestre et al. 2020; Kozlowski and Klein 2000), given that Thelen (2005) refers to the term "dynamic" as the state of an entity "at any point in time depends on its previous states and is the starting point for future states" (p. 262).

To propose and test our argument, this manuscript adopts a within-firm orientation to examine how two core measures of firms' social responsibility and irresponsibility (i.e., CSR and CSI) converge over time, and what the dynamic inter-relationship between them looks like. To answer these questions, this manuscript develops a dynamic theory of socially responsible supply chain practices by detailing the underlying process (Sutton and Staw 1995; Whetten 1989) for why firms' CSR and CSI should be longitudinally interrelated. Apart from providing explanations for why CSR and CSI measures should be linked, our theorizing allows us to specify the magnitudes of the autoregressive effects of each measure. Regarding the logic of our dynamic theory, multiple sources have been adopted, such as the attention-based perspective (Fiske and Taylor 1991; Ocasio 1997), behavioral theory of the firm (Cyert and March 1992; Gavetti, Greve, Levinthal, and Ocasio 2012), and the theory of underlying progress curves (Levy 1965). Such synthesized theorizing echoes calls for theoretical pluralism to capture the inherent complexity of supply chain management phenomena (Fawcett and Waller 2011; Sanders and Wagner 2011) by developing middle-range theories (Merton and Merton 1968) to meet the needs

of specific supply chain management domain (Holmström, Ketokivi, and Hameri 2009). We implement a multivariate autoregressive moving average model we estimate in a covariance structural equation modeling (SEM) framework to test our theory (du Toit and Browne 2007) by utilizing six years of panel data obtained from the KLD sustainability ratings on firm-level CSR and CSI. To be specific, we used longitudinal data from 2013 – 2018 for hundreds of publicly-traded firms from KLD, which has been merged with financial data from COMPUSTAT, market concentration from U.S. Census Bureau Economic Indicators (United States Census Bureau 2021), and Upstreamness measuring the average distance from final use from American economic review (Antràs et al. 2012). Our results align well with our dynamic theorizing and provide evidence for a complex series of interrelationships between firm-level CSR and CSI measures.

This work makes contributions to the existing literature in several ways. First, this research extends literature that looks at firms' social responsibility, by investigating the dynamic longitudinal interrelationships of firms' CSR and CSI practices, which has not been examined in the existing literature. Since firms' social responsibility development is a complex, dynamic system (Silvestre et al. 2020), studying how firms' social responsibility and irresponsibility practices are associated with each other through a repeated-measures design sheds light on the understanding of the complex dynamics underlying supply chain sustainability domain. Second, this manuscript provides insights into firms' social responsibility by splitting apart CSR and CSI in supply chain practices. Because investigating social responsibility by the inclusion of CSR and CSI can provide a broader, holistic viewpoint that enables researchers and practitioners to look at firms' regimes with new respect, such that outline better strategies involving socially responsible and irresponsible practices (Jones, Bowd, and Tench 2009; Lange and Washburn 2012; Sweetin,

Knowles, Summey, and McQueen 2013). In particular, researchers have traditionally and primarily looked at the CSR perspective of social responsibility and developed the logic and definition of social responsibility as an overall construct that was measured unidimensionally within the framework (Griffin and Mahon 1997). Such a unidimensional perspective is likely to limit our understanding of corporate social responsibility performance as it is usually comprised of responsible and irresponsible aspects of practices (e.g., Castillo, Mollenkopf, Bell, and Bozdogan 2018; Price and Sun 2017). Specifically, companies could involve CSR and CSI activities simultaneously in terms of doing both “good” and “bad” (Lenz, Wetzel, and Hammerschmidt 2017; Mattingly and Berman 2006; Strike, Gao, and Bansal 2006). For example, Nike has invested substantial resources in helping suppliers improve production processes yet has one or more suppliers engage in bad behavior (Distelhorst, Hainmueller, and Locke 2017). In addition, our findings indicate that the insurance mechanism explains better than the penance mechanism regarding the relationship between CSR and CSI, such that “CSR should not be used to atone for past CSI but rather as insurance against future CSI” (Kang, Germann, and Grewal 2016, p. 63). This is consistent with the literature that is grounded in the insurance mechanism, suggesting that CSR creates a positive perception of the company as a whole and thus allows it to set up close connections with various stakeholders, such close connections can safeguard the company against firm idiosyncratic risk (e.g., Flammer 2013; Godfrey, Merrill, and Hansen 2009; Schnietz and Epstein 2005). Third, our work offers one of the few longitudinal investigations of firm-level social responsibility, which allows researchers to examine complex systems consisting of change, development, and dynamics (Matusik, Hollenbeck, and Mitchell 2021). The scarcity of longitudinal examinations can be a problem, as cross-sectional studies face methodological issues and have inherent limitations on their capability regarding answering the types of

questions (McArdle and Nesselroade 2014). As a consequence, this study sheds light on more comprehensively addressing a research problem that is still unsolved in prior investigations to date, as “[a]dvancements in methodology have historically precipitated advancements in organizational research” (Matusik, Hollenbeck, and Mitchell 2021, p. 772). In particular, this manuscript aids to illustrate the unique capabilities of longitudinal structural equation modeling frameworks to reveal theoretical interest and practical relationships amongst constructs (Little 2013; Matusik, Hollenbeck, Matta, Oh 2019).

The rest of this study is organized into five sections. The next section covers the relevant literature. The second contains the theory and hypothesis development. The third details our research design and summarize the relevant variables. The fourth describes the econometric methodology and presents the results of our analysis. The fifth explains theoretical contributions, presents managerial implications, notes limitations, and makes suggestions for future research.

3.2 Background Literature

A substantial body of research looking at the value and impact of social responsibility investment has increasingly emerged in multiple disciplines including marketing (Doh et al. 2010), returns on the stock market (Becchetti and Ciciretti 2009; Statman and Glushkov 2009), organizational behavior (Chatterji, Levine, and Toffel 2009; Chatterji and Toffel 2010), stakeholder management (Coombs and Gilley 2005; Choi and Wang 2009; Deckop, Merriman, and Gupta 2006), organizational psychology (Delmas and Blass 2010), operations (Chen and Delmas 2011), and information systems (Escrig-Olmedo, Muñoz-Torres, and Fernandez-Izquierdo 2010). For example, Becchetti and Ciciretti (2009) looked at how a firm’s social responsibility investment affects its stock market performance. Cho, Lee, and Pfeiffer (2013) tested how firms’ social responsibility investment affects information asymmetry. Other scholars have examined the

factors that affect firms' socially responsible performance, such as firms' managers (Johnson and Greening 1999), and pressures from outside (Post, Rahman, and Rubow 2011; Walls, Berrone, and Phan 2012).

A brief review of the existing studies reveals that researchers have traditionally and primarily devoted themselves to understanding firms' social responsibility by examining CSR practices (Griffin and Mahon 1997; Price and Sun 2017). However, firms' social responsibility consists of not only "doing good" through CSR activities but also "doing bad" through CSI practices (Price and Sun 2017, p. 82). Yet, very little attention to date has been paid to CSI, such that the dearth of examining CSI is clearly seen in the literature (Murphy and Schlegelmich 2013; Price and Sun 2017). Researchers have recently sought to broaden the understanding of firms' social responsibility by including both CSR and CSI (e.g, Lenz, Wetzel, and Hammerschmidt 2017; Kang, Germann, and Grewal 2016; Price and Sun 2017).

Building on the knowledge generated by the existing literature, our work seeks to look at the question of firms' social responsibility development from a new perspective. Rather than developing and testing theory regarding the impact of socially responsible investment or the factors that affect socially responsible performance, this study adopts a repeated-measure design to examine the longitudinal dynamic relationship between two core measures of firms' social responsibility—CSR and CSI. Such a design enables us to extend the literature in several ways. First of all, we provide insights into firms' social responsibility by splitting apart CSR and CSI in supply chain practices. This is because investigating social responsibility by including CSR and CSI simultaneously can provide a broader, holistic viewpoint that enables researchers and practitioners to accurately assess firms' regimes in addressing better strategies for socially responsible performance (Jones, Bowd, and Tench 2009; Lange and Washburn 2012; Sweetin,

Knowles, Summey, and McQueen 2013). In particular, researchers have traditionally and primarily examined social responsibility from the CSR perspective and developed the social responsibility construct as a unidimensional measure within the framework (Griffin and Mahon 1997). Such a unidimensional perspective results in an obvious limitation in understanding firms' social responsibility performance given that is comprised of both the responsible and irresponsible aspects of practices (e.g., Castillo, Mollenkopf, Bell, and Bozdogan 2018; Price and Sun 2017). This is because corporates can involve CSR and CSI activities simultaneously by doing both “good” and “bad” (Kang, Germann, and Grewal 2016; Lenz, Wetzels, and Hamerichmidt 2017).

Second, this study adopts multiple theoretical perspectives to build the mechanistic-based explanations for why the two measures should be related (Hedström and Ylikoski 2010). This echoes calls by several studies (Sutton and Staw 1995; Whetten 1989) to uncover the reasons for the correlation of phenomena, given that scientific theories attempt to the depth and breadth of the explanatory power (Lipton 2003; Thagard 2007). Specifically, such explanatory depth provides insights into further examination (Ylikoski and Kuorikoski 2010), which is a desirable feature of scientific theories (Wacker 1998). Third, our findings indicate that a longitudinal structural equation modeling framework can be valuable for examining supply chain sustainable development. As in the case of this study, such a framework allows us to find out empirical evidence regarding how CSR and CSI converge and whether the longitudinal relationship between CSR and CSI stays invariant over time. As explained by previous longitudinal structural equation modeling literature, this identification can be of great theoretical interest and practical significance as it indicates the stability of the underlying process that brings about the observed relationships (du Toit and Browne 2007; Langeheine and Van de Pol 1990; Little 2013). In

addition, our work identifies a parsimonious, theoretically-grounded modeling framework that adequately presents our multivariate time series. Identifying parsimonious modeling frameworks that can approximate the covariance structure of complex data precisely should be valued because [1] it can simplify understanding of an underlying process that can bring about observed relationships, and [2] it could reveal several intriguing aspects of the data which went undetected before conducting statistic models (Cudeck and Henly 1991, 2003; du Toit and Browne 2007). Considering the significant implications of social responsibility measures, developing a parsimonious modeling framework that adequately approximates the observed relationship between CSR and CSI contributes to the body of knowledge both theoretically and empirically.

3.3 Theory and Hypotheses Development

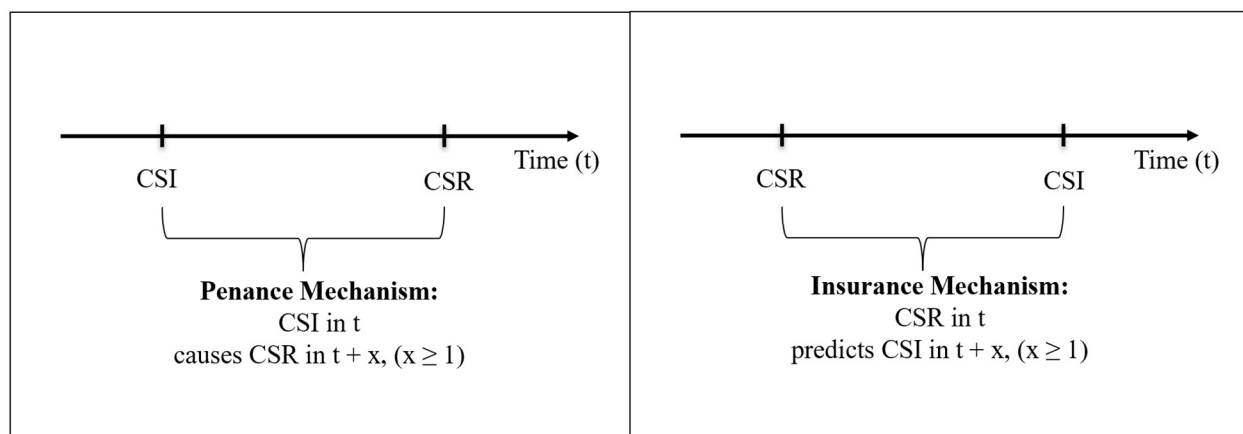
Our dynamic theory of firm-level social responsibility and irresponsibility is developed around two IRT-scaled scores measures that have been conducted in essay 1: CSR and CSI. CSR reflects positive environmental and social practices (i.e., doing good in products and services, diversity, employee relations, working environment, human rights, management systems, etc), whereas CSI indicates negative environmental and social behaviors (i.e., doing bad in products and services, diversity, employee relations, working environment, human rights, management systems, etc).

3.3.1 Cross-lagged Effects between CSR and CSI

A review of the prior literature reveals that there are two independent mechanisms concerning the relationship between CSR and CSI (Kang, Germann, and Grewal 2016). One is the penance mechanism, meaning that firms put efforts into CSR activities to atone for their previous CSI behaviors (e.g., Heal 2005; Kotchen and Moon 2012). The other is the insurance mechanism, meaning that CSR creates a positive perception of the company as a whole and thus allows it to

set up close connections with various stakeholders, such close connections can safeguard the company against the potentially negative reactions to future CSI behaviors (e.g., Flammer 2013; Godfrey, Merrill, and Hansen 2009). As such, the reversed causal relations of CSR and CSI have been proposed by penance and insurance mechanisms. Specifically, the penance mechanism indicates that companies put efforts into CSR in time $t + x$ ($x \geq 1$) to atone for CSI occurring in time t , while the insurance mechanism indicates that companies put efforts into CSR in time t , to safeguard against CSI in time $t + x$ ($x \geq 1$). We visualize the two mechanisms that propose the relationship between CSR and CSI in Figure 3.1.

Figure 3.1 Mechanisms that Propose the Relationship between CSR and CSI



Extending from Heal's (2005) proposition that socially responsible activities are used to pull down potential costs caused by socially irresponsible initiatives, Kotchen and Moon (2012) suggest that companies put effort into current CSR practices to atone for past CSI behaviors. In particular, Kotchen and Moon (2012) indicate that CSR can be regarded as a "Coasian solution" to allow companies to achieve economic efficiency by reducing externalized costs resulting from CSI activities. That is to say, a company can engage in CSR to offset its unpaid bills caused by CSI, as in the case of an oil spill, the firm causing the spill is typically likely to pay for part of

the expenses resulting from the spill, given that the total costs of such an incident cannot be estimated precisely. Likewise, if employees have been treated poorly by their companies, it is impossible to “gauge the negative ripple effect of this poor treatment on the individual workers, their families, and the communities in which they live” (Kang, Germann, and Grewal 2016, p. 63). Empirical evidence, however, indicates that companies can be penalized when they quibble or excuse themselves from their responsibility to take responsibility for their CSI behaviors, which can result in negative identification as being perceived to be in a position that tries to free themselves of responsibility for what happened (Kang, Germann, and Grewal 2016). As in the case of the Deepwater Horizon oil spill in 2010, BP was criticized for its initial muted response to the incident and the identification became negative, which led to BP’s sales dropping by 40%. Therefore, Kotchen and Moon (2012) suggest that companies are motivated to put effort into CSR by viewing social responsibility activities as a form of “penance”, allowing them to offset externalized costs caused by prior social irresponsibility behaviors. Based on such a penance mechanism, we expect the current CSI will positively predict future CSR, as in our case CSI at time t causes CSR at time $t+1$. We thus posit:

H1: CSI at time t will cause CSR at time $t+1$.

In contrast to the arguments offered in the penance mechanism, the insurance mechanism views CSR as a form of safeguard that can insure against the potentially negative reactions to future CSI behaviors (e.g., Flammer 2013; Godfrey, Merrill, and Hansen 2009; Minor and Morgan 2011). The penance and insurance mechanisms “differ in that proponents of the insurance mechanism posit that CSR should not be used to atone for past CSI but rather as insurance against future CSI” (Kang, Germann, and Grewal 2016, p. 63). As such, in comparison to the penance mechanism, in which CSI in time t is likely to cause CSR in time $t + x$ ($x \geq 1$), the

insurance mechanism proposes that CSR in time t can positively predict CSI in time $t + x$ ($x \geq 1$). The insurance mechanism has been supported by the literature suggesting that the positive perception of a firm can be leveraged as an intangible asset to protect against the potential firm idiosyncratic risk and reduce negative reactions to various events in times of crisis (e.g., Jones, Jones, and Little 2000; Godfrey, Merrill, and Hansen 2009; Schnietz and Epstein 2005). For example, Klein and Dawar (2004) find empirical evidence for the insurance mechanism viewing CSR as a safeguard for firms to reduce the negative effect of CSI. In particular, CSR facilitates consumers' positive identification with a firm, which can attenuate consumers' negative perception of the firm's harmful events. Additionally, Flammer (2013) suggests that firms with a higher level of CSR are less likely to suffer from adverse stock market reactions in times of crisis compared with firms with a lower level of CSR. Moreover, literature has confirmed the beneficial role of CSR in creating value for a company to recoup loss resulting from future CSI (e.g. Minor 2015; Minor and Morgan 2011). Specifically, Kang, Germann, and Grewal (2016) highlight that "CSR presumably helps build a reservoir of goodwill among the firm's stakeholders, which endows the firm with idiosyncrasy credits that act as safeguards (i.e., as insurance against CSI)" (p. 63). Based on this insurance mechanism, we expect current CSR will positively predict future CSI. That is to say, firms that have high CSR scores in the present tend to have high CSI scores in the future. Thus we posit:

H1 Alt: CSR at time t can be caused by CSI at time $t+1$.

3.3.2 Autoregressive Effects of CSR and CSI

We now focus on the anticipated autoregressive effects of CSR and CSI, in which our theorizing looks at how (i) CSR at time t relates to CSR at time $t+1$, and (ii) CSI at time t relates to CSI at time $t+1$. In the context of a longitudinal structural equation model for time series data, an

autoregressive relationship captures a proportional change tendency of a process as this process unfolds (Miller, Golicic, and Fugate 2017; Zyphur et al. 2020). It is critical to understand that autoregressive effects reflect the link between past and future, meaning that the current state of a system can be looked at as a function of its past. This occurs because the current state of a process depends on its past state rather than arising spontaneously. For example, under a first-order autoregressive process that remains stable with the passage of time, an autoregressive parameter less than one implies that subjects that start at higher initial scores tend to show larger reductions relative to subjects that start at lower initial scores. Such an autoregressive relationship aligns with a process whereby change has steadily occurred and the covariance of a repeated measure within itself tends to go down over time (Little 2013). By comparison, an autoregressive parameter greater than one implies that subjects that start at higher initial scores show larger increases relative to subjects that start at lower initial scores. Such an autoregressive relationship aligns with a process whereby small differences between subjects can be amplified with the passage of time and the covariance of a repeated measure within itself tends to increase over time (Little 2013; Miller, Golicic, Fugate 2017).

Extant theory indicates that the autoregressive parameters regarding our CSR and CSI measures are supposed to be less than one. That is to say, initial differences in both CSR and CSI are expected to decline with the passage of time, rather than being amplified. There are two explanations for such a prediction. First, the behavioral theory of the firm suggests that managers tend to engage in improving performance on a given measure if the performance on the measure has been poor (Cyert and March 1992; Gavetti, Greve, Levinthal, and Ocasio 2012; Greve 2003). Considering that the metrics in our work are used to assess the level of commitment of its members to environmental and social concerns (Chatterji, Levine, and Toffel 2009) and disclose

information to various stakeholders including customers, investors, and local communities (Hart and Sharfman 2015), firms with poor performance (e.g., lower scores in CSR or higher score in CSI) at the initial measurement occasion are likely to have more incentive to improve their performance. Second, the organizational learning literature, specifically studies related to learning curves (Lapr   and Tsikriktsis 2006; Levy 1965), suggests that “the rate of improvement in a given domain is proportional to current performance in that domain” (Miller, Golicic, Fugate 2017, p. 98). Aligning with this notion, Chatterji and Toffel (2010) looking at environmental management, and Williams et al. (2005) focusing on quality improvement in healthcare provide empirical support for the explanation that firms performing poorly on the given metrics, on average, make more improvements compared with firms with good initial performance on these given metrics. As such, we expect:

H2: The autoregressive parameter regarding CSR/CSI at time t to CSR/CSI at time $t+1$ will be less than one.

3.3.3 Moving Average Effects of CSR and CSI

We are also concerned about how an exogenous shock that affects a firm’s behaviors on each of our CSR and CSI measures would influence subsequent behaviors on the respective measure. That is to say, we are concerning the effect of the error terms for CSR and CSI measures at time t on the performance of each given measure at time $t+1$. In time-series analyses terms, we are looking at moving average effects (McArdle and Nesselroade 2014), which modify the autoregressive effects via making observations a direct function of past impulses on the future (Box, Jenkins, Reinsel, and Ljung 2015; Hamaker, Dolan, and Molenaar 2002). In another word, moving average effects reflect the short-run persistence of exogenous shocks, while autoregressive effects imply long-run dynamics (Zyphur et al. 2020). In our case, the moving

average parameters are expected to be negative. For example, a firm that has a negative residual for its behaviors in a given domain at time t indicates that its adaptation to changes of an exogenous shock occurs rapidly at first but then fades or slows over time. This negative residual for CSR and CSI may arise for different reasons. The negative residual for CSR, for instance, can occur because the firm's marginal costs of making improvements in CSR rise up as the absolute value in this measure goes up (Rosenthal, Quinn, and Harper 1997; Williams et al. 2005). As firms have increased marginal costs of improving CSR practices over time, we expect that a negative residual for CSR at time t will negatively predict CSR at time $t+1$. However, the negative residual for CSI, for instance, can occur because the firm is more likely to react to its socially irresponsible behaviors by changing processes and making investments to improve its poor performance on the given measure (Cyert and March 1992; Gavetti et al. 2012; Greve 2003). In our case, higher CSI indicates worse performance. Improving performance requires reducing CSI. As such, we expect that a negative residual for CSI at time t will negatively predict CSI at time $t+1$. Thus, we predict:

H3: The moving average parameter regarding CSR/CSI at time t to CSR/CSI at time $t+1$ will be negative.

3.4 Research Setting and Data

3.4.1 Research Setting

We tested our study's hypotheses in the context of MSCI ESG KLD 2013-2018. KLD socially responsible ratings provide the longest time series of firms' environmental and social sustainability information (Chatterji, Levine, and Toffel 2009). KLD ratings for measuring firms' environmental, social, and governance performance were assigned based on the company's CSR

reports and other relevant public information, which is released yearly (KLD 2018). Corporate behaviors are rated across seven dimensions: governance, community, diversity, employee relations, environment, human rights, and product quality. For each dimension, KLD ratings consist of paired items. Each such a paired item has both a strength and concern indicator, which is binary taking values 0 or 1. A score of 1 in a strength indicator indicates that the firm has a positive behavior in complying with the social responsibility standards, whereas a score of 1 in a concern indicator indicates the firm has a negative activity (i.e., social irresponsibility practice) that can be considered as a weakness to meet the standards of social responsibility.

3.4.2 Data Sources

We focus on the panel data from 2013 to 2018 for hundreds of publicly traded manufacturing, wholesale, and retailing firms included in KLD data. Once we got the KLD data, we merged it with financial data from COMPUSTAT, market concentration from US Census Bureau Economic Indicators, and Upstreamness measuring of the average distance from final use from American economic review (Antràs et al. 2012). To provide a better view of the datasets used in this manuscript, Table 3.1 provides all the datasets, using variables from them, and their sources.

Table 3.1 Datasets and their Sources

Data	Using Variables	Source
KLD data	ESG performance indicators	MSCI ESG KLD 2013-2018
Financial data	the current value of investment, cost of income, cost of investment, revenue, cost of goods sold, total assets, R&D expenditure, capital expenses	Standard & Poor's Compustat 2013-2018
Market concentration	HHI	US Census Bureau Economic Indicators 2013-2018
Upstreamness	How far is the product from final use (i.e., the average distance from final use)	American economic review (Antràs et al. 2012)

3.4.3 Measure Description

We examine firms' CSR and CSI separately. Firm-level CSR/ CSI is measured by CSR/ CSI scores, which are the results of essay 1 by using multidimensional item response theory (IRT) models. At the same time, it is important to note that a high score on CSR indicates good social responsibility, whereas a high score on CSI indicates bad social responsibility (i.e., social irresponsibility).

We first screen the data to identify potential outliers. In order to do this, we developed a series of boxplots for CSR and CSI scores across each given measurement occasion under investigation. When observation was beyond the $1.5 \times$ interquartile range above/below the third/first quartile for a given measurement occasion, it would be marked as an outlier (Moore, McCabe, and Craig 2009). As a consequence, 8 firms have been removed for having no less than three observations being outliers, which results in the final sample size of $N = 607$ public companies. This process allows us to not only reduce concerns about abnormal firm biasing results but also make our data more Gaussian. As such, the final sample can be more consistent with the assumptions underlying maximum Wishart likelihood estimation. Table 3.2 displays the full correlation matrix, means, and standard deviations of the measures.

Table 3.2 Correlations, Means, and Standard Deviations for all Measures on all Occasions

	CSR1	CSR2	CSR3	CSR4	CSR5	CSR6	CSI1	CSI2	CSI3	CSI4	CSI5	CSI6
CSR1	1											
CSR2	0.870	1										
CSR3	0.810	0.849	1									
CSR4	0.708	0.730	0.854	1								
CSR5	0.659	0.681	0.785	0.865	1							
CSR6	0.600	0.614	0.547	0.514	0.461	1						
CSI1	0.782	0.727	0.611	0.524	0.502	0.443	1					
CSI2	0.696	0.806	0.641	0.547	0.514	0.461	0.892	1				
CSI3	0.684	0.758	0.819	0.698	0.646	0.564	0.745	0.827	1			
CSI4	0.633	0.689	0.748	0.845	0.742	0.657	0.659	0.714	0.830	1		
CSI5	0.613	0.653	0.692	0.746	0.831	0.689	0.636	0.679	0.766	0.879	1	
CSI6	0.602	0.626	0.654	0.687	0.719	0.822	0.611	0.643	0.706	0.789	0.868	1
Mean	0.220	0.162	0.154	0.235	0.371	0.642	0.093	0.064	0.010	0.028	0.065	0.182
Stdev.	0.755	0.763	0.770	0.827	0.871	0.853	0.658	0.653	0.598	0.522	0.616	0.581

3.5 Methods and Results

3.5.1 Methods

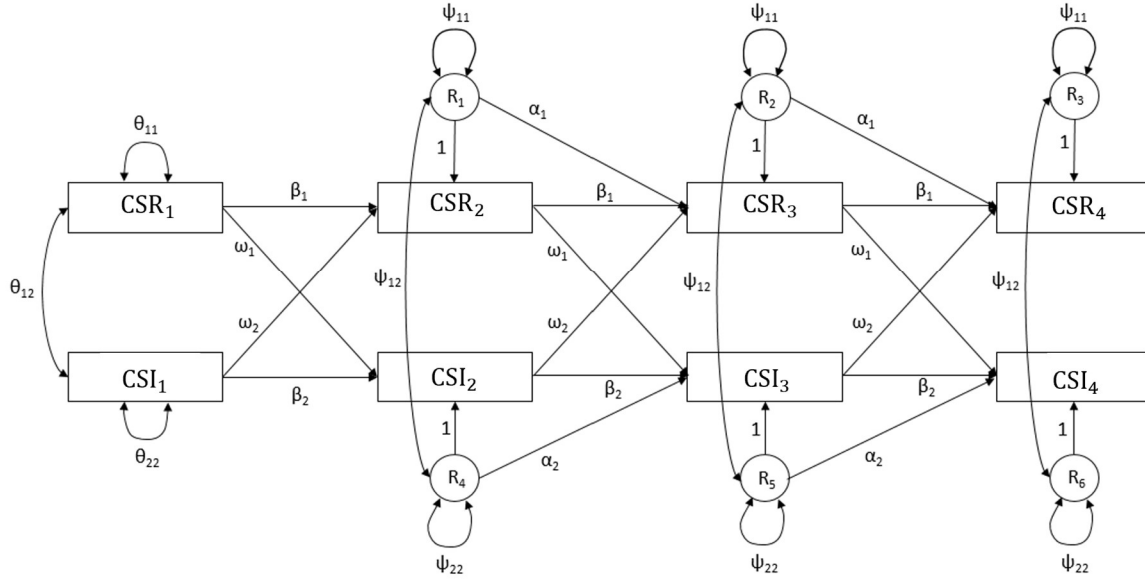
We use the covariance structure of multivariate time series of CSR and CSI to test our dynamic theorizing. “A statistical model for such a data structure should adequately recover the observed covariance matrix using a limited number of parameters that provide interesting information regarding a process that may have generated the data” (Miller, Golicic, Fugate 2017, p. 101). After reviewing the literature on the covariance structure of multivariate time series (Little 2013; McArdle and Nesselroade 2014), this study decided to adopt a vector (multivariate) autoregressive moving average model (i.e., the VARMA model) (du Toit and Browne 2007) regarding testing the dynamic process that we theorized. The VARMA model should be an ideal framework for testing our hypotheses, given that such a modeling framework consists of

parameter estimations for cross-lagged effects, autoregressive effects, and moving average effects that are essential for testing the dynamic process that we theorized (du Toit and Browne 2007). We specify our estimating through a VARMA(1,1) model, which is a specific model that incorporates one period of autoregressive and cross-lagged effect as well as one moving average period. We opted for a VARMA(1,1) model, given that such a specification reflects the simplest VARMA process (du Toit and Browne 2007) that can be easily falsified, which is a desirable characteristic for a preferred modeling framework (MacCallum, Roznowski, Mar, and Reith 1994; Meehl 1990). The VARMA(1,1) model, in our case, reflects a very parsimonious covariance structure of the multivariate time series⁴, since our model depends on 12 free parameters to approximate the 400 unique elements of the observed covariance matrix. A simplified version of the VARMA(1,1) model has been shown in Figure 3.2 with our two core measures, *CSR* and *CSI scores*.

⁴ Our model includes 2 autoregressive parameters, 4 coupling parameters, 2 variances for each measure at the first measurement occasion, 2 covariances between the measures at the first occasion, 2 residual variances for each measure regarding the endogenous measurement occasions, as well as 2 covariances between the residual variances at the same measurement occasion. Aligning with common practice, the parameters in our model have been constrained to be constant across measurement occasions. The coupling parameter that links *Strength Scores* at t to *Weakness Scores* at $t+1$, for instance, has been set to equality across measurement occasions.

Figure 3.2 Simplified Version of the VARMA(1,1) Model

Note. This model includes our core measures and four measurement occasions. Greek notation has been used to label our parameters, where the same labels have been constrained to be equal.



Given that we focus on firms' performance from 2013 to 2018, our specification started in 2013, which is our initial measurement occasion. For the second measurement occasion ($t = 2$), the linear equations can be specified as follows

$$CSR_2 = \beta_1 CSR_1 + \omega_1 CSI_1 + u_{SS_2} \quad [1]$$

$$CSI_2 = \beta_2 CSI_1 + \omega_2 CSR_1 + u_{WS_2} \quad [2]$$

From the third to sixth measurement occasions, the linear equations can be specified as follows

$$CSR_t = \beta_1 CSR_{t-1} + \omega_1 CSI_{t-1} + \alpha_1 u_{SS_{t-1}} + u_{SS_t} \quad t \geq 3 \quad [3]$$

$$CSI_t = \beta_2 CSI_{t-1} + \omega_2 CSR_{t-1} + \alpha_2 u_{WS_{t-1}} + u_{WS_t} \quad t \geq 3 \quad [4]$$

We'd like to note that including a one-period lagged value of the dependent variable in each linear equation provides support for us to account for significant cross-lagged relationships

captured by parameters ω_1 and ω_2 , which indicates Granger causality (Zyphur et al. 2020). This is because lagged values of one time series (e.g., *CSR*) “contain information regarding future values of another time series” (e.g., *CSI*), “not captured by that time series’ past information” (Granger 1988; Wooldridge 2009, 649–650). Likewise, literature (e.g., Maxwell and Cole 2007; Little 2013) suggests that researchers can “have greater confidence that significant cross-lagged estimates represent causal effects when these relationships are statistically significant holding constant prior values of dependent variables” (Miller, Golicic, Fugate 2017, p. 102). Therefore, our modeling design can be regarded as a strong test of our hypotheses.

3.5.2 Results

Our modeling framework is estimated by utilizing Mplus Version 8.2 with the robust maximum likelihood (MLR) estimator, which allows us to correct for deviations from multivariate normality. Our results are reported in Table 3.3 including the regression weights and moving average parameters. Our VARMA(1,1) model fits the data well, with $\chi^2 = 215.59$ with $DF = 66$, correction factor = 1.70, sample-corrected RMSEA = 0.061 with 90% RMSEA CI [0.052, 0.070], and SRMR = 0.037.

Table 3.3 Results from Fitting the VARMA(1,1) Model to the Multivariate Time Series
regarding the Measures of CSR and CSI

	CSR	CSI
Regression Weights		
CSR _t → CSR _{t+1}	0.919*** (0.016)	
CSI _t → CSR _{t+1}	-0.001 (0.020)	
CSI _t → CSI _{t+1}		0.796*** (0.022)
CSR _t → CSI _{t+1}		0.072*** (0.013)
Moving Average		
Resid. CSR _t → CSR _{t+1}	-0.238*** (0.026)	
Resid. CSI _t → CSI _{t+1}		-0.179*** (0.029)
Model Fit		
R ² Occasion 2	0.69	0.79
R ² Occasion 3	0.70	0.76
R ² Occasion 4	0.74	0.69
R ² Occasion 5	0.74	0.74
R ² Occasion 6	0.73	0.71

Notes: • = $p < 0.10$; ** = $p < 0.05$; *** = $p < 0.01$ (one-tailed tests).

Z-values are reported below parameter estimates in parentheses.

Model fit using MLR estimation. Model fit: $\chi^2 = 215.59$ with $DF = 66$, correction factor = 1.70, sample-corrected RMSEA = 0.061 with 90% RMSEA CI [0.052, 0.070], SRMR = 0.037.

Our work starts by examining parameters for the two mechanisms (i.e., penance and insurance mechanism) regarding the inter-relationship between CSR and CSI. As shown in Table 3.3, CSR at time t is significantly and positively related to CSI at time $t+1$, indicating that the insurance mechanism is supported, whereas CSI at time t is insignificantly related to CSR at time $t+1$, suggesting that the penance mechanism is not supported. Therefore, our findings support *H1Alt* rather than *H1*. Turning to the autoregressive effects, the autoregressive parameter relating CSR at time t to CSI at time $t+1$ is less than 1.0, with a point estimate of 0.919 and a 95% confidence interval of [0.887, 0.950]. Meanwhile, the autoregressive parameter relating CSI at time t to CSI at time $t+1$ is less than 1.0, with a point estimate of 0.796 and a 95% confidence

interval of [0.753, 0.839]. As such, *H2* is supported. This indicates that firms with worse social responsibility performance, on average, make greater improvements in their performance. Additionally, investigating these autoregressive parameter estimates, our observation was that the magnitude of the autoregressive parameter for *CSI* was significantly less than the autoregressive parameter for *CSR*. We then focused on testing whether these parameters were indeed different. To do this, we re-estimated the VARMA(1,1) model and imposed parameter constraints. Aligning with what we observed, the $\Delta\chi^2 = 12.57$ with DF=1 ($p < 0.01$) whereby the autoregressive parameters were constrained to be equal, providing evidence that the magnitude of the autoregressive parameter for *CSR* was indeed larger. We interpret such a finding in the managerial implications. Furthermore, consistent with *H3*, the moving average parameter relating *CSR/CSI* at time t to *CSR/CSI* at time $t+1$ is statistically significant and negative, indicating that firms' adaptation to changes of an exogenous shock occurs rapidly at first but then fades or slows over time.

A model of multivariate time series implies complex dynamics, which makes it difficult to be interpreted, when solely relying on observing parameter estimations (McArdle and Nesselrode 2014). To effectively convey the dynamics that our model implies, we visualized the relationships by plotting the predicted values for firms starting with combinations of high (75th percentile) or low (25th percentile) values of *CSR* and *CSI* (see Figures 3.3 & 3.4). When developing these plots, we included the autoregressive parameters and cross-lagged parameters that link the previous values of each measure to the future values of the next measure. Specifically, the estimated intercepts were incorporated for completeness, when we were generating our plots. Moreover, we specified that the predicted values at time t for each measure were included to predict the next value of the given measure and the other measure, when

developing these plots. Several significant findings can be uncovered in Figures 3.3 & 3.4. First, Figure 3.3 implies that CSR at time t is unrelated to CSI at time $t+1$. Even though we have included CSI when generating the plot for CSR . Relative to CSI , the initial value of CSR has a more profound effect on its trajectory, which aligns with the fact that CSR has a larger point estimate for its autoregressive parameter than CSI as shown in Table 3.3. Second, Figure 3.4 indicates a firm's predicted trajectory on CSI has been heavily contingent on its initial performance on CSR . The strong interdependence between previous CSI and present CSR can be revealed in Table 3.3 where the parameter estimate linking CSR at t to CSI at $t+1$ is positive and significant. This provides more evidence to support the insurance mechanism rather than the penance mechanism, meaning that “CSR should not be used to atone for past CSI but rather as insurance against future CSI ” (Kang, Germann, and Grewal 2016, p. 63).

Figure 3.3 Plot of CSR Implied by the VARMA Model Contingent on CSI at the First Measurement Occasion

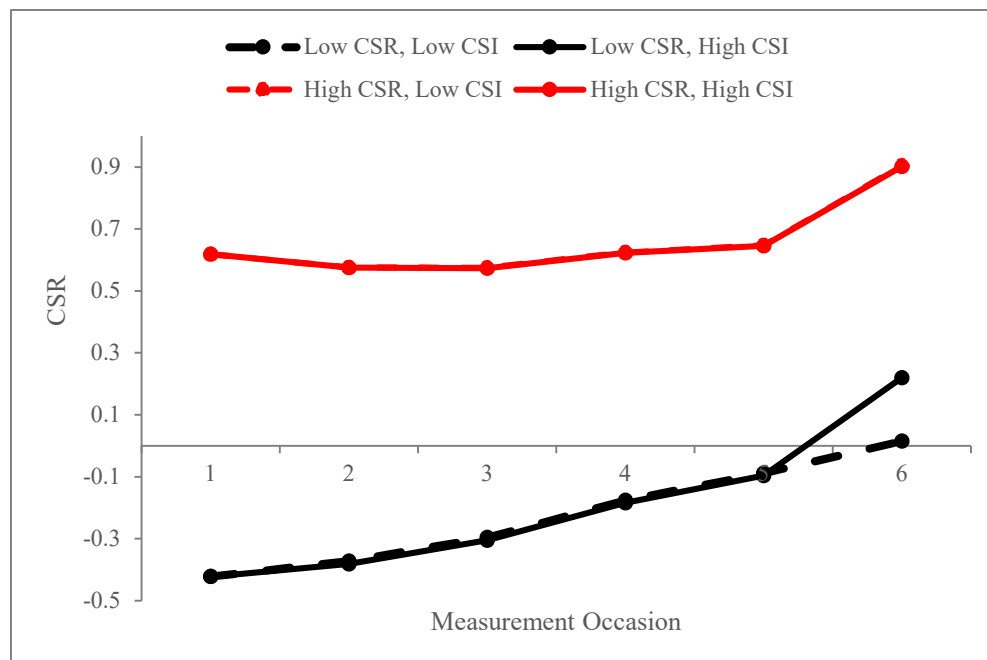
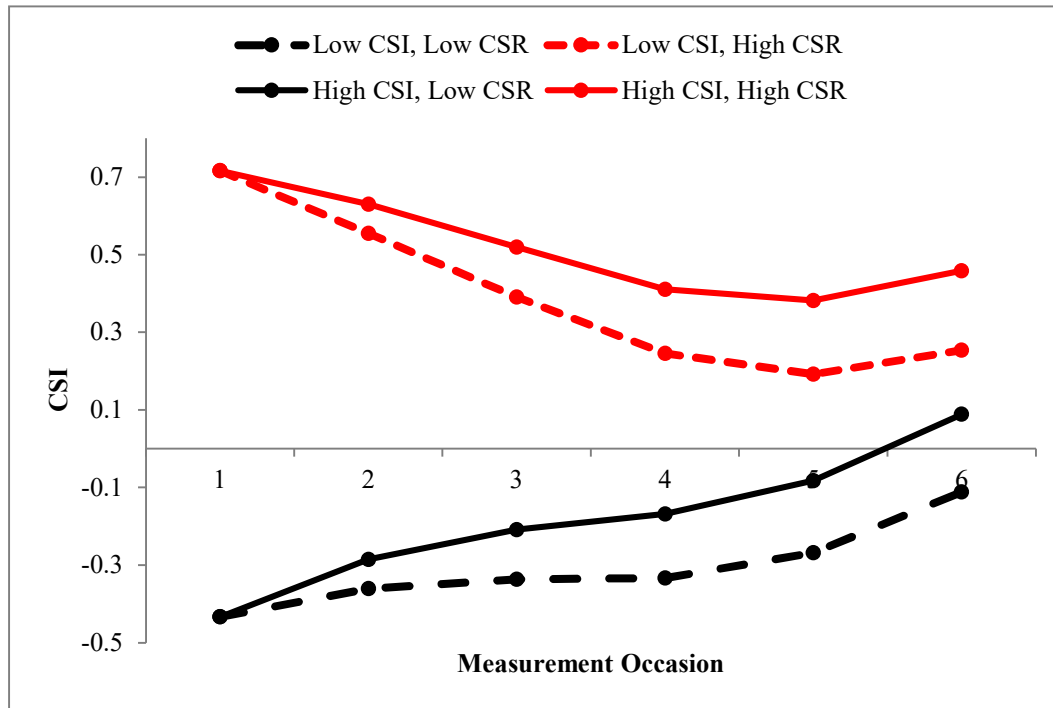


Figure 3.4 Plot of CSI Implied by the VARMA Model Contingent on CSR at the First Measurement Occasion



3.5.3 Cross-Validating VARMA(1,1) Model

The capability of generalizing to a broader population regarding a statistical model could make it more desirable, meaning that results hold up for cross-validation estimates by fitting this model to a different sample (Browne 2000; Cudeck and Browne 1983). For cross-validating, we fit the same VARMA(1,1) model to a new data on CSR and CSI from a second independent sample of $N = 656$ publicly traded manufacturing, wholesale, and retailing firms included in KLD data from 2003 to 2008. To conduct such cross-validating, we constrained our structural parameter estimations to the acquired values from our original sample. We opted to free the parameter estimations regarding variances, residual variances, and covariances, according to the insights of the psychometrics literature that parameter estimations concerning variances and covariances can be different within random samples in the same population because of sampling error

(MacCallum et al. 1994). As such, cross-validating our VARMA(1,1) model by fitting to a raw data from 2003 to 2008 turned out to be acceptable ($\chi^2 = 171.58$, $DF = 66$, correction factor = 1.67, sample-corrected RMSEA = 0.049, 90% sample-corrected RMSEA CI = [0.040, 0.059], SRMR = 0.023). In addition, following MacCallum, Browne, and Cai's (2006) procedures, we conducted a hypothesis test for a small difference of RMSEA of 0.01 regarding the constrained models vis-à-vis models without constraints. As a consequence, the null hypothesis could not be rejected, meaning that "there was not a small difference in RMSEA" (Miller, Golicic, and Fugate, p. 105). As such, results from cross-validation present solid evidence to support generalizing the findings from our first sample to the broader population of publicly traded manufacturing, wholesale, and retailing firms included in KLD data.

3.6 Discussion

3.6.1 Theoretical and Empirical Contributions

This work contributes to the body of knowledge, both theoretically and empirically. First, this research extends literature that looks at firms' social responsibility, by investigating the dynamic longitudinal interrelationships of firms' CSR and CSI practices, which has not been examined in the existing literature. Since firms' social responsibility development is a complex, dynamic system (Silvestre et al. 2020), studying how firms' social responsibility and irresponsibility practices are associated by adopting a repeated-measures design allows us to more fully understand the complex dynamics implied by corporate social responsibility. In particular, we rely on multiple theoretical perspectives to develop mechanistic-based explanations for why the two measures should be related (Hedström and Ylikoski 2010). Our work echoes the call by several studies (Sutton and Staw 1995; Whetten 1989) to uncover the reasons for the correlation of phenomena, given that scientific theories attempt to the depth and breadth of the explanatory

power (Lipton 2003; Thagard 2007). Specifically, such explanatory depth provides insights into further examination (Ylikoski and Kuorikoski 2010), which is a desirable feature of scientific theories (Wacker 1998).

Second, this manuscript provides insights into firms' social responsibility by splitting apart CSR and CSI in supply chain practices. This is because investigating social responsibility by including CSR and CSI simultaneously can provide a broader, holistic viewpoint that enables researchers and practitioners to accurately assess firms' regimes in addressing better strategies for socially responsible performance (Jones, Bowd, and Tench 2009; Lange and Washburn 2012; Sweetin, Knowles, Summey, and McQueen 2013). In particular, researchers have traditionally and primarily examined social responsibility from the CSR perspective and developed the social responsibility construct as a unidimensional measure within the framework (Griffin and Mahon 1997). Such a unidimensional perspective results in an obvious limitation in understanding firms' social responsibility performance given that is comprised of both the responsible and irresponsible aspects of practices (e.g., Castillo, Mollenkopf, Bell, and Bozdogan 2018; Price and Sun 2017). This is because corporates can involve CSR and CSI activities simultaneously by doing both "good" and "bad" (Kang, Germann, and Grewal 2016; Lenz, Wetzels, and Hamerichmidt 2017). For example, Nike has invested substantial resources in helping suppliers improve production processes yet has one or more suppliers engage in bad behavior (Distelhorst, Hainmueller, and Locke 2017).

Third, our work offers one of the few longitudinal investigations of firm-level social responsibility, which allows researchers to examine complex systems consisting of change, development, and dynamics (Matusik, Hollenbeck, and Mitchell 2021). The scarcity of longitudinal examinations can be a problem, as cross-sectional studies face methodological

issues and have inherent limitations on their capability regarding answering the types of questions (McArdle and Nesselroade 2014). As a consequence, this study sheds light on more comprehensively addressing a research problem that is still unsolved in prior investigations to date, as “[a]dvancements in methodology have historically precipitated advancements in organizational research” (Matusik, Hollenbeck, and Mitchell 2021, p. 772). In particular, this manuscript aids to illustrate the unique capabilities of longitudinal structural equation modeling frameworks to reveal theoretical interests and practical relationships amongst constructs (Little 2013; Matusik, Hollenbeck, Matta, Oh 2019). For example, these models allow us to [1] predict the development of processes through the longitudinal relationships between measures (Grimm, Castro-Schio, and Davoudzadeh 2013; Langeheine and Van de Pol 1990), [2] examine the magnitude of the autoregressive relationships across measures, and [3] provide empirical evidence for how constructs coevolve over time (Matusik, Hollenbeck, and Mitchell 2021). Our work, thus, echoes the calls for the longitudinal examination of organizational phenomena (e.g., Mitchell and James 2001; Miller, Golicic, and Fugate 2017; Vantiborgh, Hofmans, and Judge 2018).

In addition, our findings indicate that a longitudinal structural equation modeling framework can be valuable for examining supply chain sustainable development. As in the case of this study, such a framework allows us to find out empirical evidence regarding how CSR and CSI converge and whether the longitudinal relationship between CSR and CSI stays invariant over time. As explained by previous longitudinal structural equation modeling literature, this identification can be of great theoretical interest and practical significance as it indicates the stability of the underlying process that brings about the observed relationships (du Toit and Browne 2007; Langeheine and Van de Pol 1990; Little 2013). In addition, our work identifies a

parsimonious, theoretically-grounded modeling framework that adequately presents our multivariate time series. Identifying parsimonious modeling frameworks that can approximate the covariance structure of complex data precisely should be valued because [1] it can simplify understanding of an underlying process that can bring about observed relationships, and [2] it could reveal several intriguing aspects of the data which went undetected before conducting statistic models (Cudeck and Henly 1991, 2003; du Toit and Browne 2007). Considering the significant implications of social responsibility measures, developing a parsimonious modeling framework that adequately approximates the observed relationship between CSR and CSI contributes to the body of knowledge both theoretically and empirically.

3.6.2 Managerial Implications

This study has several implications for managers at multiple stakeholders such as buyers, suppliers, inspectors, and regulators. Our first practical implication is to shed light on the question that how CSR and CSI converge, as firms increasingly engage in both CSR and CSI (Kang, Germann, and Grewal 2016). Specifically, our results show support for the insurance mechanism rather than the penance mechanism, meaning that “CSR should not be used to atone for past CSI but rather as insurance against future CSI” (Kang, Germann, and Grewal 2016, p. 63). In particular, we speculate that increasing expectations of business social responsibility behavior (Langan and Menz 2022) can significantly contribute to the increasing correlation between CSR and CSI. For example, more and more consumers expect firms to develop social responsibility (e.g., Ipsos 2013; Kusi-Sarpong Gupta, and Sarkis 2019; Langan and Menz 2022). Such a consumer trend can be further revealed in the increased media attention CSR received over time (Kang, Germann, and Grewal 2016). Thus, we suggest that managers are likely to view CSR as a strategy to protect against future CSI. This insurance mechanism has been supported by

the literature suggesting that the positive perception of a firm can be leveraged as an intangible asset to protect against the potential firm idiosyncratic risk and reduce negative reactions to various events in times of crisis (e.g., Jones, Jones, and Little 2000; Godfrey, Merrill, and Hansen 2009; Schnietz and Epstein 2005). For example, Klein and Dawar (2004) find empirical evidence for the insurance mechanism viewing CSR as a safeguard for firms to reduce the negative effect of CSI. In particular, CSR facilitates consumers' positive identification with a firm, which can attenuate consumers' negative perception of the firm's harmful events. Additionally, Flammer (2013) suggests that firms with a higher level of CSR are less likely to suffer from adverse stock market reactions in times of crisis compared with firms with a lower level of CSR. Moreover, literature has confirmed the beneficial role of CSR in creating value for a company to recoup loss resulting from future CSI (e.g. Minor 2015; Minor and Morgan 2011). Specifically, Kang, Germann, and Grewal (2016) highlight that "CSR presumably helps build a reservoir of goodwill among the firm's stakeholders, which endows the firm with idiosyncrasy credits that act as safeguards (i.e., as insurance against CSI)" (p. 63).

Our second practical implication is that the strength of the autoregressive relationships between CSR and CSI can be quantified. Such information is critical, as it can be used to make predictions about a firm's future social responsibility and irresponsibility grounded on its current performance on that measure. The capability to predict the levels of social responsibility can be important implications for buyers in that choosing a socially responsible and irresponsible supplier heavily affects their supply chain sustainability. For suppliers, predicting and then improving socially responsible and irresponsible practices allow them to reduce future firm idiosyncratic risks and/or increase future market value, as benefits from improving corporate social performance have been confirmed by literature (e.g., McWilliams and Siegel 2011; Wang

and Choi 2013). Our results indicate that future *CSR* is more strongly associated with prior value than *CSI*, meaning that the socially responsible practices are much more stable than the socially irresponsible practices. In other words, firms' socially responsible practices are more of cumulative investments, whereas firms' socially irresponsible behaviors involve more reactions according to their internal and external business environment. Such a finding, in conjunction with the fact that our model explains less variability in *CSI*, is most likely rooting in managers putting less selective attention on *CSI* than *CSR*, as managers are starkly aware of the importance of doing good through *CSR* to ensure better publicity, and marketing (Hillman, Zardkoohi, and Bierman 1999; Tate, Ellram, and Kirchoff 2010). Therefore, although our model is designed to work well through examining the interrelationships between *CSR* and *CSI*, it is important to note that more caution needs to be proceeded with predicting a firm's future *CSI*.

The third practical implication is that our findings provide empirical evidence regarding how different measures of firm-level social responsibility can act as leading indicators of other measures. That being said, our work provides evidence of which social responsibility measure plays the leading role of "Granger cause" other measures (Wooldridge 2009). As in our case, *CSR* serves as a strong leading indicator of future *CSI*, suggesting that a company's current *CSR* can be used to predict its future *CSI*. As such, *CSR* is the "Granger cause" of *CSI*. Such information can be valuable for multiple stakeholders, such as buyers, suppliers, inspectors, and regulators. However, it is important to inform stakeholders that the interrelationships of core measures are likely to be asymmetric. For instance, *CSI* at time $t+1$ Granger causes *CSR* at time t , whereas *CSR* at time $t+1$ cannot Granger cause *CSI* at time t .

Our fourth practical implication is that our findings identify the dynamic stability of both the autoregressive and cross-lagged relationships for *CSR* and *CSI* across time. This invariance

can increase confidence that the underlying process that brings about the relationships between social responsibility and irresponsibility remains stable. Such a stable nature has considerable practical importance, as this indicates that the observed relationships tend to continue into the future by following a similar path (Little 2013). The amount of information regarding the stability of our observed relationships can be practically important for various stakeholders. In addition, this study provides practical implications regarding the amount of information from the KLD database buyers supposedly include when making supplier selection decisions. The challenge for buyers is how to incorporate enough information to probe for an adequate understanding of supplier social responsibility and irresponsibility performance without overloading information. Our findings suggest that a single lag should be enough to obtain a sufficient approximation of the correlations amongst the social responsibility measures for one-year measurement intervals. As such, our work suggests that it would be better for buyers to focus more on a supplier's most recent performance of social responsibility vis-à-vis that farther in the past.

3.6.3 Limitations and Suggestions for Future Research

It is important to note that there are some limitations to the present study. First, the time span for this study includes 6 years starting from 2013 to 2018. Such a range can be extended further based on the data available. Second, there is a limitation regarding the number of firms examined. This is because the ratings of certain firms in the KLD database are not consistent over the time span. In other words, some firms incorporated in the KLD participate a few times or participate irregularly over the defined timespan. Thus, we exclude these firms with limited observations over our defined timespan given that they are likely to add noise to our longitudinal examination (Berg et al. 2020). Third, although the KLD social indicators have been broadly used in the

existing social responsibility literature, firms' social responsibility performance can be measured by somewhat subjective rather than objective items. As such, scholars can replicate our analysis based on other data rather than the KLD index to measure firms' social responsibility performance. Yet, to date, we are not aware of a source other than the KLD database that can provide such data. Moreover, in this study, we focus on US manufacturing, wholesale, and retailing firms. Thus, we hesitate to generalize our findings to firms operating in other countries. Fourth, similar to most research that relies on statistical models to test theoretical predictions (Freedman 1991), this study does not directly observe if the processes theorized are operating in a way that brings about the posited relationships. However, the fact our theory accounts for a wide array of findings reduces concerns that an alternative explanation can better account for our set of empirical findings (Lipton 2003). In addition, this study focuses on one possible model to investigate the correlations of multivariate time series. A VARMA(1,1) model has been selected to approximate the covariance structure, because such an operating model performed well in testing our proposed hypotheses. Nevertheless, it is worth investigating whether there are other statistical frameworks that work as well or better than our VARMA(1,1) model. Moreover, this investigation may facilitate revealing other interesting aspects of multivariate time series data. For instance, future research can evaluate the efficacy of possible covariance pattern modeling frameworks (Browne 1977; Jöreskog 1978) to interpret the longitudinal relationships observed in this study. However, identifying any possible operating model that performs as well as the VARMA(1,1) model we adopted cannot challenge our research, given that there is no "true" model to be found (Cudeck and Henly 2003). Identifying other operating models, in fact, enriches our knowledge of the dynamics of firm-level social responsibility and irresponsibility practices.

This research can be extended in multiple directions. One avenue for future research to extend our work could be to address the limitation that we cannot observe the underlying processes that gave rise to the relationships we found. According to Steel (2004), researchers, in this case, are suggested to conduct some qualitative investigations (e.g., case studies) considering the beneficial role of such approaches to reveal the underlying processes that give rise to relationships between phenomena. Such investigations could particularly add value to understanding the processes that give rise to the cross-lagged relationships between social responsibility and irresponsibility measures, because scholars disagree on mechanisms regarding the relationship between CSR and CSI (Kang, Germann, and Grewal 2016). Although our results show support for the insurance mechanism that reviews CSR as insurance against future CSI rather than the penance mechanism that views CSR as a form of “penance” to offset firms’ past CSI, future research could examine more about the mechanisms between CSR and CSI to provide empirical evidence. Another way for future research to extend our work is to examine how the changes in output, capital intensity, and change in R&D intensity would affect firms’ CSR and CSI behaviors over time. A third direction is to investigate how firms’ CSR and CSI behaviors interact with the competition they face. Since economic data suggests that industry concentration, total sales for the sector, both nominal and potentially deflated (United States Census Bureau 2021), the challenges with social responsibility and irresponsibility as firms are facing competition, are whether they going to invest in socially responsible practices. Another avenue is to investigate how the incentive structure as a moderator affects firms’ CSR and CSI behaviors over time because the literature suggests incentive structure matters a lot (Mukandwal et al. 2020; Jadhav, Orr, and Malik 2019; Silvestre et al. 2020). For example, look at whether such an interaction with incentives makes social responsibility behaviors even more problematic

or how industry-level imports affect firms' social responsibility and irresponsibility over time. Additionally, further research can be conducted by examining how either firm-level or industry-level productivity affects firms' investment in social responsibility practices.

REFERENCES

REFERENCES

- Antràs, P., Chor, D., Fally, T., & Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review*, 102(3), 412-16.
- Armstrong, J. S., & Green, K. C. (2013). Effects of corporate social responsibility and irresponsibility policies. *Journal of Business Research*, 66(10), 1922-1927.
- Barnett, M. L. (2007). Stakeholder influence capacity and the variability of financial returns to corporate social responsibility. *Academy of management review*, 32(3), 794-816.
- Becchetti, L., & Ciciretti, R. (2009). Corporate social responsibility and stock market performance. *Applied financial economics*, 19(16), 1283-1293.
- Browne, M. W. (1977). The analysis of patterned correlation matrices by generalized least squares. *British Journal of Mathematical and Statistical Psychology*, 30(1), 113-124.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance. *Academy of management review*, 4(4), 497-505.
- Castillo, V. E., Mollenkopf, D. A., Bell, J. E., & Bozdogan, H. (2018). Supply chain integrity: A key to sustainable supply chain management. *Journal of Business Logistics*, 39(1), 38-56.
- Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility?. *Journal of Economics & Management Strategy*, 18(1), 125-169.
- Chatterji, A. K., & Toffel, M. W. (2010). How firms respond to being rated. *Strategic Management Journal*, 31(9), 917-945.
- Chen, C. M., & Delmas, M. (2011). Measuring corporate social performance: An efficiency perspective. *Production and operations management*, 20(6), 789-804.
- Cho, S. Y., Lee, C., & Pfeiffer Jr, R. J. (2013). Corporate social responsibility performance and information asymmetry. *Journal of Accounting and Public Policy*, 32(1), 71-83.
- Choi, J., & Wang, H. (2009). Stakeholder relations and the persistence of corporate financial performance. *Strategic management journal*, 30(8), 895-907.
- Coombs, J. E., & Gilley, K. M. (2005). Stakeholder management as a predictor of CEO compensation: Main effects and interactions with financial performance. *Strategic Management Journal*, 26(9), 827-840.

- Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: a clarification. *Psychological bulletin*, 109(3), 512.
- Cudeck, R., & Henly, S. J. (2003). A realistic perspective on pattern representation in growth data: comment on Bauer and Curran (2003).
- Cyert, R. M., & March, J. G. (1992). *A behavioral theory of the firm* (Second ed.). Malden, MA: Blackwell Publishers.
- Deckop, J. R., Merriman, K. K., & Gupta, S. (2006). The effects of CEO pay structure on corporate social performance. *Journal of Management*, 32(3), 329-342.
- Delmas, M., & Blass, V. D. (2010). Measuring corporate environmental performance: the trade-offs of sustainability ratings. *Business Strategy and the Environment*, 19(4), 245-260.
- Distelhorst, G., Hainmueller, J., & Locke, R. M. (2017). Does lean improve labor standards? Management and social performance in the Nike supply chain. *Management Science*, 63(3), 707-728.
- Doh, J. P., Howton, S. D., Howton, S. W., & Siegel, D. S. (2010). Does the market respond to an endorsement of social responsibility? The role of institutions, information, and legitimacy. *Journal of Management*, 36(6), 1461-1485.
- du Toit, S. H., & Browne, M. W. (2007). Structural equation modeling of multivariate time series. *Multivariate Behavioral Research*, 42(1), 67-101.
- Escrig-Olmedo, E., Muñoz-Torres, M. J., & Fernandez-Izquierdo, M. A. (2010). Socially responsible investing: sustainability indices, ESG rating and information provider agencies. *International journal of sustainable economy*, 2(4), 442-461.
- Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics*, 162, 101-114.
- Fawcett, S. E., & Waller, M. A. (2011). Moving the needle: making a contribution when the easy questions have been answered. *Journal of Business Logistics*, 32(4), 291-295.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. McGraw-Hill Book Company.
- Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, 56(3), 758-781.
- Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological methodology*, 291-313.
- Gavetti, G., Greve, H. R., Levinthal, D. A., & Ocasio, W. (2012). The behavioral theory of the firm: Assessment and prospects. *Academy of Management Annals*, 6(1), 1-40.

- Godfrey, P. C., Merrill, C. B., & Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic management journal*, 30(4), 425-445.
- Granger, C. W. (1988). Some recent development in a concept of causality. *Journal of econometrics*, 39(1-2), 199-211.
- Greve, H. R. (2003). A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. *Academy of management journal*, 46(6), 685-702.
- Griffin, J. J., & Mahon, J. F. (1997). The corporate social performance and corporate financial performance debate: Twenty-five years of incomparable research. *Business & society*, 36(1), 5-31.
- Grimm, K. J., Castro-Schilo, L., & Davoudzadeh, P. (2013). Modeling intraindividual change in nonlinear growth models with latent change scores. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, 26(3), 153.
- Groening, C., & Kanuri, V. K. (2013). Investor reaction to positive and negative corporate social events. *Journal of Business Research*, 66(10), 1852-1860.
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. (2002). On the nature of SEM estimates of ARMA parameters. *Structural Equation Modeling*, 9(3), 347-368.
- Hart, T. A., & Sharfman, M. (2015). Assessing the concurrent validity of the revised Kinder, Lydenberg, and Domini corporate social performance indicators. *Business & Society*, 54(5), 575-598.
- Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual review of sociology*, 36, 49-67.
- Heal, G. (2005). Corporate social responsibility: An economic and financial framework. *The Geneva papers on risk and insurance-Issues and practice*, 30(3), 387-409.
- Hillman, A. J., Zardkoohi, A., & Bierman, L. (1999). Corporate political strategies and firm performance: indications of firm-specific benefits from personal service in the US government. *Strategic Management Journal*, 20(1), 67-81.
- Holmström, J., Ketokivi, M., & Hameri, A. P. (2009). Bridging practice and theory: A design science approach. *Decision Sciences*, 40(1), 65-87.
- Ipsos (2013). Eight out of Ten Australians Rate Corporate Social Responsibility as Important. Available at: <http://ipsos.com.au/wp-content/uploads/2013/06/Corporate-social-responsibility-media-release-FINAL.pdf> (accessed June 13 2022).
- Jadhav, A., Orr, S., & Malik, M. (2019). The role of supply chain orientation in achieving supply chain sustainability. *International Journal of Production Economics*, 217, 112-125.

- Johnson, R. A., & Greening, D. W. (1999). The effects of corporate governance and institutional ownership types on corporate social performance. *Academy of management journal*, 42(5), 564-576.
- Jones, B., Bowd, R., & Tench, R. (2009). Corporate irresponsibility and corporate social responsibility: competing realities. *Social Responsibility Journal*.
- Jones, G. H., Jones, B. H., & Little, P. (2000). Reputation as reservoir: Buffering against loss in times of economic crisis. *Corporate Reputation Review*, 3(1), 21-29.
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4), 443-477.
- Kang, C., Germann, F., & Grewal, R. (2016). Washing away your sins? Corporate social responsibility, corporate social irresponsibility, and firm performance. *Journal of Marketing*, 80(2), 59-79.
- Klein, J., & Dawar, N. (2004). Corporate social responsibility and consumers' attributions and brand evaluations in a product-harm crisis. *International Journal of research in Marketing*, 21(3), 203-217.
- Kotchen, M., & Moon, J. J. (2012). Corporate social responsibility for irresponsibility. *The BE Journal of Economic Analysis & Policy*, 12(1).
- Kozlowski, S. W., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes.
- Kusi-Sarpong, S., Gupta, H., & Sarkis, J. (2019). A supply chain sustainability innovation framework and evaluation methodology. *International Journal of Production Research*, 57(7), 1990-2008.
- Langan, R., & Menz, M. (2022). Does Your Company Need a Chief ESG Officer? *Harvard Business Review*. Available at: <https://hbr.org/2022/02/does-your-company-need-a-chief-esg-officer> (accessed March 21 2022).
- Langeheine, R., & Van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods & Research*, 18(4), 416-441.
- Lange, D., & Washburn, N. T. (2012). Understanding attributions of corporate social irresponsibility. *Academy of management review*, 37(2), 300-326.
- Lapr , M. A., & Tsikriktsis, N. (2006). Organizational learning curves for customer dissatisfaction: Heterogeneity across airlines. *Management science*, 52(3), 352-366.
- Lenz, I., Wetzel, H. A., & Hammerschmidt, M. (2017). Can doing good lead to doing poorly? Firm value implications of CSR in the face of CSI. *Journal of the Academy of Marketing Science*, 45(5), 677-697.

- Levy, F. K. (1965). Adaptation in the production process. *Management Science*, 11(6), B-136.
- Lipton, P. (2003). *Inference to the best explanation*. Routledge.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press.
- Mattingly, J. E., & Berman, S. L. (2006). Measurement of corporate social action: Discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society*, 45(1), 20-46.
- Matusik, J. G., Hollenbeck, J. R., Matta, F. K., & Oh, J. K. (2019). Dynamic systems theory and dual change score models: Seeing teams through the lens of developmental psychology. *Academy of Management Journal*, 62(6), 1760-1788.
- Matusik, J. G., Hollenbeck, J. R., & Mitchell, R. L. (2021). Latent change score models for the study of development and dynamics in organizational research. *Organizational Research Methods*, 24(4), 772-801.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological methods*, 12(1), 23.
- McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. American Psychological Association.
- MacCallum, R. C., Roznowski, M., Mar, C. M., & Reith, J. V. (1994). Alternative strategies for cross-validation of covariance structure models. *Multivariate behavioral research*, 29(1), 1-32.
- McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of management review*, 26(1), 117-127.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological inquiry*, 1(2), 108-141.
- Merton, R. K., & Merton, R. C. (1968). *Social theory and social structure*. Simon and Schuster.
- Miller, J. W., Golicic, S. L., & Fugate, B. S. (2017). Developing and testing a dynamic theory of motor carrier safety. *Journal of Business Logistics*, 38(2), 96-114.
- Minor, D. (2015). *The value of corporate citizenship: protection*. Harvard Business School Strategy Unit Working Paper, (16-021).
- Minor, D., & Morgan, J. (2011). CSR as reputation insurance: Primum non nocere. *California Management Review*, 53(3), 40-59.
- Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review*, 26(4), 530-547.

- Moore, D.S., McCabe, G.P., & Craig, B.A. (2009). *Introduction to the Practice of Statistics* (Sixth ed.). New York, NY: W. H. Freeman and Company.
- Mukandwal, P. S., Cantor, D. E., Grimm, C. M., Elking, I., & Hofer, C. (2020). Do firms spend more on suppliers that have environmental expertise? An empirical study of US manufacturers' procurement spend. *Journal of Business Logistics*, 41(2), 129-148.
- Murphy, P. E., & Schlegelmilch, B. B. (2013). Corporate social responsibility and corporate social irresponsibility: Introduction to a special topic section. *Journal of Business Research*, 66(10), 1807-1813.
- Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic management journal*, 18(S1), 187-206.
- Post, C., Rahman, N., & Rubow, E. (2011). Green governance: Boards of directors' composition and environmental corporate social responsibility. *Business & society*, 50(1), 189-223.
- Price, J. M., & Sun, W. (2017). Doing good and doing bad: The impact of corporate social responsibility and irresponsibility on firm performance. *Journal of Business Research*, 80, 82-97.
- Rosenthal, G. E., Quinn, L., & Harper, D. L. (1997). Declines in hospital mortality associated with a regional initiative to measure hospital performance. *American Journal of Medical Quality*, 12(2), 103-112.
- Sanders, N. R., & Wagner, S. M. (2011). Multidisciplinary and multimethod research for addressing contemporary supply chain challenges. *Journal of Business Logistics*, 32(4), 317-323.
- Schnietz, K. E., & Epstein, M. J. (2005). Exploring the financial value of a reputation for corporate social responsibility during a crisis. *Corporate reputation review*, 7(4), 327-345.
- Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of cleaner production*, 16(15), 1699-1710.
- Silvestre, B. S., Silva, M. E., Cormack, A., & Thome, A. M. T. (2020). Supply chain sustainability trajectories: learning through sustainability initiatives. *International Journal of Operations & Production Management*.
- Statman, M., & Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, 65(4), 33-46.
- Steel, D. (2004). Social mechanisms and causal inference. *Philosophy of the social sciences*, 34(1), 55-78.
- Strike, V. M., Gao, J., & Bansal, P. (2006). Being good while being bad: Social responsibility and the international diversification of US firms. *Journal of International Business Studies*, 37(6), 850-862.

- Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative science quarterly*, 371-384.
- Sweetin, V. H., Knowles, L. L., Summey, J. H., & McQueen, K. S. (2013). Willingness-to-punish the corporate brand for corporate social irresponsibility. *Journal of Business Research*, 66(10), 1822-1830.
- Tate, W. L., Ellram, L. M., & Kirchoff, J. F. (2010). Corporate social responsibility reports: a thematic analysis related to supply chain management. *Journal of supply chain management*, 46(1), 19-44.
- Thagard, P. (2007). Coherence, truth, and the development of scientific knowledge. *Philosophy of science*, 74(1), 28-47.
- Thelen, E. (2005). Dynamic systems theory and the complexity of change. *Psychoanalytic dialogues*, 15(2), 255-283.
- United States Census Bureau. (2021). Business Dynamics Statistics: Annual Report: 2003-2018. Available at: <https://www.census.gov/data/tables/econ/awts/annual-reports.html> (accessed 19 March 2022).
- Vantilborgh, T., Hofmans, J., & Judge, T. A. (2018). The time has come to study dynamics at work. *Journal of Organizational Behavior*, 39(9), 1045-1049.
- Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of operations management*, 16(4), 361-385.
- Walls, J. L., Berrone, P., & Phan, P. H. (2012). Corporate governance and environmental performance: Is there really a link?. *Strategic management journal*, 33(8), 885-913.
- Whetten, D. A. (1989). What constitutes a theoretical contribution?. *Academy of management review*, 14(4), 490-495.
- Williams, S. C., Schmaltz, S. P., Morton, D. J., Koss, R. G., & Loeb, J. M. (2005). Quality of care in US hospitals as reflected by standardized measures, 2002–2004. *New England Journal of Medicine*, 353(3), 255-264.
- Wood, D. J. (1991). Toward improving corporate social performance. *Business Horizons*, 34(4), 66-74.
- Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach* (4th edition). Mason, OH: South-Western Cengage Learning.
- Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical studies*, 148(2), 201-219.

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., ... & Diener, E. (2020). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, 23(4), 651-687.