COEXISTENCE OF DISPOSITION AND REVERSE DISPOSITION EFFECTS ACROSS MULTIPLE REFERENCE POINTS ON BORSA ISTANBUL

BEKİR SERHAT ÇEVİKEL

BOĞAZİÇİ UNIVERSITY

2023

COEXISTENCE OF DISPOSITION AND REVERSE DISPOSITION EFFECTS

ACROSS MULTIPLE REFERENCE POINTS ON BORSA ISTANBUL

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Management

by

Bekir Serhat Çevikel

Boğaziçi University

2023

DECLARATION OF ORIGINALITY

I, Bekir Serhat Çevikel, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature...

Date ....

ABSTRACT

Coexistence of Disposition and Reverse Disposition Effects Across Multiple Reference Points on Borsa Istanbul

Disposition effect is the propensity to sell winners earlier and hold losers longer while reverse disposition effect is the opposite. Using transaction data from Borsa Istanbul, coexistence of disposition effect and reverse disposition effect is tested with respect to three reference points: Cost, maximum and minimum prices. The overhang levels are discretized using a method introduced here as MUHBOD, varying the cutting points across the levels of other discretized variables. Extended Cox model is the main method, supported by Kaplan-Meier survival curves and an enhanced version of Odean's measure. Coefficients and measures derived from these models are tested for univariate monotonicity using Kendall's Tau and for multivariate monotonicity using Bayesian regression models with monotonic effects.

The results suggest that investors exhibit disposition effect with respect to cost reference and reverse disposition effect with respect to maximum and minimum references, simultaneously. Furthermore, these effects also interact with each other: As prices are perceived to be closer to the minimum and/or maximum references, the strength of disposition effect diminishes. The findings can be explained by the affect account theory. Within the anticipated feelings dimension, investors feel elation in gains and disappointment in losses with respect to the cost reference. Within the anticipatory feelings dimension, investors feel the fear of losing more when prices are closer down to the minimum references, and they feel the hope of earning more when prices are closer up to the maximum reference.

ÖZET

Borsa İstanbul'da Çoklu Referans Noktaları Boyunca Eğilim ve Ters Eğilim Etkilerinin Birlikte Varolması

Eğilim etkisi, kazanan pozisyonları daha erken satma ve kaybeden pozisyonları daha uzun süre elde tutma eğilimidir, ters eğilim etkisi ise tam tersidir. Borsa İstanbul'dan alınan işlem verileri kullanılarak, eğilim etkisi ile ters eğilim etkisinin birlikte varlığı üç referans noktasına göre test edilmektedir: Maliyet, maksimum ve minimum fiyatlar. Kâr seviyeleri, burada MUHBOD olarak tanıtılan ve diğer kesikli değişkenlerin seviyeleri boyunca kesme noktalarını değiştiren bir yöntem kullanılarak kesikleştirilmektedir. Genişletilmiş Cox modeli ana yöntem olup, Kaplan-Meier hayatta kalma eğrileri ve Odean ölçütünün geliştirilmiş bir versiyonu ile de desteklenmektedir. Bu modellerden elde edilen katsayı ve ölçütlerin tek değişkenli tekdüzeliği Kendall's Tau ile, çok değişkenli tekdüzeliği ise tekdüze etkilere sahip Bayes regresyon modelleri ile test edilmektedir.

Sonuçlar, yatırımcıların eş zamanlı olarak maliyet referansına göre eğilim etkisi ve maksimum ve minimum referanslara göre ters eğilim etkisi sergilediğini göstermektedir. Ayrıca bu etkiler birbirleriyle de etkileşim halindedir: Fiyatların minimum ve/veya maksimum referanslara yakın olduğu algılandıkça, eğilim etkisinin gücü azalmaktadır. Bulgular etki hesabı teorisi ile açıklanabilmektedir. Beklenen duygular boyutunda yatırımcılar, maliyet referansına göre kazançlarda sevinç, kayıplarda ise hayal kırıklığı yaşamaktadırlar. Beklenti duyguları boyutunda yatırımcılar, fiyatlar minimum referanslara yaklaştıkça daha fazla kaybetme korkusu, maksimum referansa yaklaştıkça ise daha fazla kazanma umudu duymaktadırlar.

CURRICULUM VITAE

NAME: Bekir Serhat Çevikel

DEGREES AWARDED

PhD in Management, 2023, Boğaziçi University

MBA, 2009, İstanbul Bilgi University

BA in Management, 1999, Boğaziçi University

AREAS OF SPECIAL INTEREST

Behavioral finance, big data analytics, machine learning, management simulation, algorithmic trading, valuation

PROFESSIONAL EXPERIENCE

Instructor, Department of Management, Boğaziçi University, 2018 - present

Specialist, Department of Management, Boğaziçi University, 2016 - 2018

Research Assistant, Department of Management, Boğaziçi University, 2010 - 2015

Advisor to Board, Türk Telekom, 2005 - 2009

Vice President, Is Investment Securities, 2005

Senior Associate, Is Investment Securities, 2002 - 2005

Associate, Is Investment Securities, 1999 - 2002

Associate, Dış Investment Securities (later Fortis Investment Securities), 1998 - 1999

AWARDS AND HONORS

Graduate Education Exam of Türkiye, 28th standing in Weighted Score, 2001

Graduate Education Exam of Türkiye, 14th standing in Weighted Score, 1999

University Entrance Exam of Türkiye, 32nd standing in Turkish-Math Score, 1995

PUBLICATIONS

Journal Articles

Ercan, M. & Çevikel, S. (2011). Survey on IPO underpricing in the Istanbul Stock Exchange. *İktisat İşletme ve Finans*, *26*(305), 85–117. doi:10.3848/iif.2011.305.2999

Book Chapters

Coskun, A., Cevikel, S., Özçelik, Z. & Akgiray, V. (2019). The role of the state in MENA capital markets. In I. Diwan, A. Malik & I. Atiyas (Eds.), *Crony Capitalism in the Middle East: Business and Politics from Liberalization to the Arab Spring* (pp. 344–362). doi:10.1093/oso/9780198799870.003.0014

ACKNOWLEDGEMENTS

In the early 2000s, long before I began to pursue my doctorate degree and without being aware of its lengthy history of research in the literature, I first had the idea of the disposition effect, which I investigate in detail in this thesis. I initiated the preliminary work for the thesis in 2012. Since then, I have surveyed or read hundreds of resources and I have also immersed myself in the reproducible and programmed side of data analysis including the meticulous editing of the document. Even though I was the sole person participating in each step of the entire workflow, many other people inspired me to pursue academic research, gave me with opportunities to overcome obstacles, or offered assistance and support when it came to administrative or procedural phases. I would like to express my sincere gratitude to them.

First and foremost, my thesis advisor Prof. Rıza Metin Ercan has always been much more than an academic figure for me during the last quarter century of my life. We grew closer as a family, and I always felt his support and assistance in many facets of my life. I was his first student at Boğaziçi University back in 1997. I started my professional life in finance with him in 1998. Later we worked together in Türk Telekom between 2005-2009 and afterwards he was the one to motivate and invite me to start my doctorate degree. I was honored to act as his assistant and advisor, and we collaborated in many projects.

Data act as the lifeblood of academic research and without data, a scholar is left with nothing more than theoretical arguments and a review of the literature. Due to the nature of the investor level data required for this thesis, I confronted difficulties to access that data. At a point when I was stuck with no data, Prof. Ahmet Vedat Akgiray unlocked the most critical door to make this big dataset available to me. And that was the first big dataset I had personal access to. After that step, apart from being a study in the finance area, this thesis also became a research of big data analytics. And this kind of a research drove me to acquire many new skills related to coding, system administration, database management, DevOps, and similar areas. Thus, the door that Prof. Akgiray unlocked had a significant impact on my orientation to pursue a career in big data analytics with an emphasis on finance.

Dr. Recep Bildik provided me that lifeblood, the dataset which I long struggled to find. So along with Prof. Akgiray, the assistance of Dr. Bildik allowed this thesis to advance and gave me access to new opportunities in big data analytics.

Assoc. Prof. Attila Odabaşı was the first one to spark the idea that in order to write a good thesis, it is a valuable opportunity to invest in coding skills and especially focus on R for statistical tasks. So I first heard about R programming language from Prof. Odabaşı. And R provided the backbone of many tasks, including data wrangling, modeling, visualization and reporting in this thesis.

My supervisory committee members Assoc. Prof. Ali Coşkun and Assoc. Prof. Gözde Erhan Ünal, my thesis committee members Assist. Prof. Kadri Özgüneş and Assist. Prof. Emrah Ahi, my academic advisors Prof. Mine Uğurlu and Assoc. Prof. Neslihan Yılmaz, former chair of Management Department Assist.
Prof. Mustafa Hayri Tongarlak, former assistant director of the Institute for Graduate Studies in Social Sciences Assoc. Prof. Tolga Umut Kuzubaş, former vice chair of Management Department Assist. Prof. Cenk Cevat Karahan, former executive

secretary of the Faculty of Economics and Administrative Sciences Hatıra Şenkon, secretary of Management Department Barış Yağmur and secretary of the Institute for Graduate Studies in Social Sciences Ersen Göl provided me valuable assistance, support and help during the progression of my thesis.

I used many software components, libraries and extensions, which I also mention in the thesis. From the operating system and statistical computing environment to the typesetting and publishing components, all software tools I used are free and open source software. The free and open software movement has risen upon the shoulders of countless altruist hackers, in the original and laudatory meaning of hacker, and developers during the last fifty years. The pioneers of the movement and/or originators or contributors to the software that I used in this thesis include but are not limited to Ken Thompson, Dennis Ritchie, Brian Kernighan, Donald Knuth, Richard Stallman, Linus Torvalds, John Chambers, Ross Ihaka, Robert Gentleman, Matt Dowle, Arun Srinivasan, Hadley Wickham, Paul-Christian Bürkner, Terry M. Therneau, Leslie Lamport, Yihui Xie, John MacFarlane and Chester Ismay. Without the tools they created, developed and made available to public use, this thesis would be impossible.

And last, my wife İkbal Çevikel has always provided the support and motivation for me to carry out this long journey.

To Mustafa Kemal Atatürk, the savior of the Turkish Nation, the founder of the Republic of Türkiye and the inspiration for the enlightenment and scientific progress towards modern civilization in our country…

To all our martyrs and veterans who fought and sacrificed their lives for the independence and indivisible integrity of the Republic of Türkiye and the Turkish Nation…

To my parents and grandparents…

To my wife for her constant support and tolerance during my long journey of doctorate degree…

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

API	:	Application programming interface
BIST	:	Borsa Istanbul
brms	:	Bayesian regression models using ’Stan’
CAGR	:	Compound annual growth rate
COVID-19	:	Coronavirus Disease 2019
GNU	:	GNU’s Not Unix (recursive acronym)
IQ	:	Intelligence quotient
ISE	:	Istanbul Stock Exchange
KM	:	Kaplan-Meier
LDB	:	Large discount brokerage
MCMC	:	Markov chain Monte Carlo
MUHBOD	:	Multivariate hierarchical balanced ordinal discretization
PGR	:	Proportion of gains realized
PLR	:	Proportion of losses realized
SP/A	:	Security, potential and aspiration
SQL	:	Structured query language
US	:	United States

CHAPTER 1

INTRODUCTION

The purpose of this thesis is to test whether investors trading on Borsa Istanbul (BIST) display a monotonous and consistent *disposition* or *reverse disposition* effect with respect to multiple reference points simultaneously and whether these reference points are updated.

A novel approach is introduced to create ordinal variables from overhang (unrealized capital gains/losses) measures controlling for other variables such as length of holding period and market return since the respective reference points are formed or updated. These holding period and market return variables are also discretized ordinally in a hierarchical fashion. With this approach, distribution of overhang across ordinal categories remains balanced for different market return and holding period levels. The term *multivariate hierarchical balanced ordinal discretization* (MUHBOD) is coined in this thesis for this new method. The approach echoes the investors' evaluation and perception of returns, as a result of which the investors may be labeling and comparing those returns as high, mediocre, low, etc. also controlling for categorical perception of the market returns (e.g. high, low, etc.) and investment horizons (e.g. short, medium, long, etc.), instead of making complex financial calculations (e.g. abnormal returns and/or annualization) to incorporate these additional measures since unsophisticated investors may not be able to conduct such complex analyses.

Ordinal overhang measures for multiple reference points can be tested together as main effects or multiple way interactions. An advantage of using ordinals as covariates in modeling is the flexibility of the underlying relationship between the selling status of positions as dependent variable and the covariates. Inspired by Shumway and Wu (2005), extended Cox model, Cox proportional hazards model with time-varying covariates, is selected as the major method. The semi-parametric nature of the method and partial maximum likelihood calculation allow for more flexibility. Following Odean (1998), the Odean's measure, proportion of realized positions at different overhang levels, is used as a secondary model for confirmation of results from the Cox model.

Another innovation of this thesis is that the multivariate monotonicity of coefficients or ratios which represent propensity of investors to sell across ordinals of overhang measures with respect to reference points are tested using a Bayesian framework incorporating monotonic effects. So while the numeric coefficient representing the propensity to sell is the dependent variable, multiple ordinal variables for overhang measures taken from interaction terms of the Cox model are introduced into this second level model as explanatory variables. These ordinal variables are allowed to have varying degrees of effects on the dependent variable between their sequential levels. While computationally expensive, the Bayesian framework allows for a flexible and multivariate relationship among variables where the monotonicity of the effect of sequential levels of an ordinal is tested.

A total of 13 extended Cox models are run in this thesis. The models with interaction terms up to 998 covariates are tested further with Bayesian monotonic effects model.

The main contributions of this thesis to the literature are as follows:

- Testing the shape of the propensity to sell versus overhang levels with respect to multiple reference points simultaneously

- Discretization of numeric overhang levels controlling for holding periods and market returns. Hence, those ordinal variables become context-sensitive.
- Testing the multivariate monotonicity of coefficients from extended Cox models at a second level of modeling by sampling values from the coefficients' estimates and standard errors and using those values within a Bayesian approach.

The thesis is organized as follows: Chapter 2 (p. 4) summarizes literature on behavioral finance, disposition effect and categorical view of data. Chapter 3 (p. 50) develops the hypotheses tested in this thesis. Chapter 4 (p. 57) lists various software tools, all open source in nature, utilized in this thesis. Chapter 5 (p. 61) details the nature of data, complete data provenance from the raw source to the dataset included in the model, all wrangling steps to transform the data, descriptive statistics of data and the novel approach for discretization introduced here. Chapter 6 (p. 115) explains all methods to measure the propensity to sell in the first level and to test the monotonicity of propensity measures in the second level. Chapter 7 (p. 138) presents the results of the models tested. Chapter 8 (p. 192) interprets and discusses the findings in the light of psychological concepts from the study field of economics. Chapter 9 (p. 202) recaps the hypotheses, methods, results and the discussion and concludes.

CHAPTER 2

THEORETICAL BACKGROUND AND LITERATURE REVIEW

The quest for the existence of disposition effect in capital and financial markets is a part of the behavioral finance literature. And the main rationale for behavioral finance is that investors and markets demonstrate some degree of inefficiencies, in contrast to what efficient market hypothesis suggests.

So the theoretical background will start with a quick history and review of behavioral finance and then commence to discussion about the disposition effect.

2.1 A review of behavioral finance

Evolution of finance discipline covers three schools of thought (Haugen, 2001; Ramiah, Xu, & Moosa, 2015):

- Old finance: Focuses on analysis of financial statements
- Modern finance: Focuses on asset pricing and valuation based on rationality and efficient markets
- New finance: Focuses on inefficient markets and adopts behavioral models

Standard finance or modern asset pricing theory is based on several important landmarks in finance (Antony, 2019; Statman, 1999; Zahera & Bansal, 2018):

- Expected utility theory is the basis for risk return trade-off.
- Markowitz portfolio theory is the basis for creating an optimal portfolio on the risk-return plane utilizing risk reduction through diversification.
- Capital asset pricing model introduces a new notion of risk, systematic or non-diversifiable risk and remodels the risk-return trade-off using systematic risk.

- Black-Scholes-Merton model introduces a method for valuing options and other derivative instrument.

Ramiah et al. (2015) list the major sources of those landmarks and pillars as follows:

- Neoclassical finance or standard finance started when Markowitz (1952) introduced modern portfolio theory.
- Next important step in the field was the introduction of the irrelevance of capital structure by Modigliani and Miller (1958, 1963).
- Capital asset pricing model was developed by Sharpe (1964) and Lintner (1965).
- Fama (1965, 1970) introduced efficient market hypothesis.
- Black and Scholes (1973) pioneered option pricing theory.

These theories require that markets are efficient and investors are rational. Rationality has two meanings (Barberis & Thaler, 2003):

- Agents update their beliefs correctly upon arrival of new information.
- Given their beliefs, agents make choices that are normatively acceptable.

Efficient market hypothesis was introduced by Fama (1965, 1970) and the hypothesis simply claims that prices are right in the sense that they reflect available information and are set by rational agents. Efficient market hypothesis is defined in three forms based on type of available information in the market: Weak, semi-strong, strong. Defenders of efficient market hypothesis argue that when an attractive opportunity arises, rational investors buy the security at its bargain price instantly (Barberis & Thaler, 2003; Kapoor & Prosad, 2017).

However, some market anomalies that call market efficiency into question have been identified in time (Zahera & Bansal, 2018), most important of them being the existence of bubbles and crashes in the markets.

Most famous historic incidence of market bubble and subsequent crash is that of the Tulip bubble in 1630s in the Netherlands (Kapoor & Prosad, 2017). Possession of tulip flower became a status symbol in upper circles. The price of the flower skyrocketed until investors realized that the price was excessively high for a commodity with such a low utility value and the market crashed.

Some financial phenomena can be better understood with models incorporating agents that are not fully rational (Barberis & Thaler, 2003).

Behavioral finance is about the psyche of investors and its role in financial decisions (Kapoor & Prosad, 2017). Behavioral finance accepts that market participants are subject to errors arising from heuristics and biases (Ramiah et al., 2015). According to Barberis and Thaler (2003), behavioral finance field has two major building blocks:

- Limits to arbitrage: Difficulty of rational traders to exploit the dislocations caused by less rational traders
- Psychology: Deviations from full rationality

Behavioral finance accepts the fact that mispricings in the market are noticed and exploited quickly. However, actions to correct the mispricings are risky and costly, so they remain unattractive and mispricings can persist (Barberis & Thaler, 2003). Behavioral finance literature mentions several factors for limited market efficiency (Aggarwal, 2014):

- Transaction costs: Costs of negotiation, search, enforcement and contracting for a business
- Limits to arbitrage: Limits arising from costs of leverage and short-selling and also from noisy information

- Behavioral biases: Cognitive and psychological biases, heuristics and emotional affects that deviate from rationality
- Other factors: Informational asymmetries and adverse selection making it harder to identify good or bad investments, herding behavior, positive feedback loops and bubbles

Equity return anomalies have also been found empirically as evidences of market inefficiencies and they are as follows (Aggarwal, 2014; Barberis & Thaler, 2003):

- Size effect: Smaller firms outperform larger firms
- Book-to-market ratio effect: Lower book-to-market ratio firms outperform higher book-to-market ratio firms
- Short-run momentum: Tendency of the recent performance to continue
- Long-run reversal: Tendency of performance to revert in the long term
- Other: High volatility of returns compared to discounted future streams and post-earnings announcement drift

Behavioral finance substitutes normal investors for the rational ones, behavioral portfolio theory for mean-variance portfolio theory and behavioral asset pricing model for capital asset pricing model (Statman, 2014).

Investors may look for three kinds of benefits (Statman, 2014):

- Utilitarian benefits answer the question of “What does it do for me and my pocket?”.
- Expressive benefits answer the question of “What does it say about me to others and me?”.
- Emotional benefits answer the question of “How does it make me feel?”.

Normal investors, unlike rational ones, are prone to cognitive errors and misleading emotions. Yet they are not alike, instead they have varying wants of utilitarian, expressive and emotional benefits, and they stand along the range of normal-ignorant to normal-knowledgeable investors who have learned imperfectly to subdue their cognitive errors and misleading emotions (Statman, 2014).

Statman (2014) also defines hard-to-beat markets, distinct from rational ones. While prices always equal to intrinsic values in rational markets, they may deviate in hard-to-beat markets. However, it is hard to exploit those deviations and earn above market returns.

The assumptions of behavioral finance are different from those of traditional finance models in following respects (Ramiah et al., 2015):

- Investors may incorporate non-statistical and psychological factors while making investment decisions.
- Investors may perceive trends, while no such patterns exist in reality.
- Information is imperfect with trader heterogeneity.
- Investors may have different investment opportunities depending on taste, while herding may induce a common taste.
- The market may be in disequilibrium and available arbitrage opportunities may be subject to market sentiment.

This line of literature emerged in late 1970’s and questioned the rational investor and market efficiency (Zahera & Bansal, 2018). The first paper in this regard came from Kahneman and Tversky (1979) introducing *prospect theory*. According to prospect theory, contrary to expected utility theory, investors exhibit risk aversion in choices with sure gains, while they exhibit risk seeking in choices with sure losses.

Thaler (1980) posits that consumers' actions are inconsistent with economic theory, and economic theory will make systematic errors in these situations.

Shiller (1981) remarks that in the long term stock price volatility is too high to be attributed to new information about future real dividends, and casts doubt on rationality of investors and efficiency of markets.

Debondt and Thaler (1985) are among the first to test contrarian investment strategies, overreaction and long-term price reversal. They find that over the last half-century, loser portfolios outperform the market by 19.6% while winner ones underperform the market by 5% in 36 month windows.

Jegadeesh and Titman (1993) test momentum investment strategies, short-term price momentum and underreaction hypothesis. According to their results, between 1965 and 1989, winner portfolios selected on past six months' return and held for six months outperform the market by 12.01% on CAGR basis.

Thaler (1999) discusses five areas that contradicts the modern finance theory:

- Volume: Investors trade more frequently than the amount of new information arrivals into the market may justify.
- Volatility: The prices in the market fluctuate more than arrival of new information can justify.
- Dividends: The fact that stock prices react to dividend decisions contradicts the dividend irrelevance theory.
- The equity premium puzzle: The return premium of equities over T-bills is higher than that can be justified by risk alone.
- Predictability: Contrary to what efficient market hypothesis posits, stock prices are at least partly predictable.

Zahera and Bansal (2018) list some behavioral biases frequently discussed in literature as follows:

- Overconfidence: Investors pay too much attention to their own capabilities and are highly optimistic about their trading outcomes.
- Disposition effect: Investors tend to realize their gains early, but they tend to delay the realization of their losses.
- Herding effect: Investors tend to follow the decisions of other investors.
- Mental accounting: Investors create portfolios based on mental categories and assign separate investment policies to each.
- Confirmation bias: People rely on their own preconceived impressions and avoid other information.
- Hindsight bias: Investors can form cause and effect relationship between two unrelated events.
- House money effect: Investors with huge profits tend to take more risks.
- Endowment effect: Investors pay too much emphasis on their current holdings and refrain from changing their positions.
- Framing: Investors avoid or take risks depending on whether an information is presented within a positive or negative frame.
- Home bias: As a result of belongingness feeling, investors tend to invest in domestic companies regardless of the return potential.
- Self-attribution: Investors attribute success to their own actions, while they attribute failure to actions of others.
- Conservatism: Investors stick to their own beliefs and disregard useful information by others.

- Regret aversion: When people regret their decisions, they update their future decisions accordingly.
- Recency: Investors might put more emphasis on recent information, while they neglect useful information that took place a while ago.
- Anchoring: Investors might place their judgments on the initial information they receive, and later decisions are also based on that same information.
- Representativeness: Investors tend to assess an event to be similar to other events, and so consider that the assessed event is more likely to happen.

2.2 Disposition effect

The timeline of influential studies in disposition effect is presented below[1]. The development of the idea of disposition effect went through some major milestones:

- Schlarbaum, Lewellen, and Lease (1978) first observe that investors tend to sell the securities in the gain region, and hold the ones in the loss region with hopes of recovery, a phenomenon they regard as psychological rather than economic.
- Kahneman and Tversky (1979) introduce the notion that the perceptions of outcomes are measured as gains and losses with respect to a reference point rather than the wealth level, and a kinked value function is defined: People are risk averse in gain region, and are risk seeking in loss region.
- Thaler (1985) posits that investors segregate buy and sell decisions of separate stocks in separate *mental accounts*.

[1] The backbone of this timeline is formed by selecting articles that are highly cited, taking into account also the years since the publication dates of the studies. The citation counts are gathered from OpenCitations, an independent non-profit organization for open bibliographic and citation data, Crossref, another non-profit organization with the goal of making scholarly communications better, and Google Scholar, a service by Google that provides a simple way to broadly search for scholarly literature (“Crossref”, n.d.; “Google Scholar”, n.d.; “Open Citation”, n.d.). Highest of the three citation counts is taken for each publication.

- Shefrin and Statman (1985) define the disposition effect as the disposition to sell winners too early and hold losers too long within a framework incorporating prospect theory, mental accounting, regret aversion, self-control and tax considerations.
- Odean (1998) is the first to test disposition effect using an investor level large dataset and brings the first widely accepted measure to define the disposition effect, *Odean's measure*.
- Feng and Seasholes (2005) incorporate survival methods in testing disposition effect using a much larger dataset and a larger set of variables.

Schlarbaum et al. (1978) are among the pioneers for studies of individual investment experiences and focus on comparing individual round-trip[2] investment returns with those of broad market portfolios or collection of mutual funds. The study predates and foreshadows prospect theory and disposition effect literature. 75,123 round trip investments between January 1964 and December 1970 on the US capital markets are divided into 0-30, 31-182, 183-365 and longer than 365 days categories so that return differences among them can be observed. The study also highlights the selection issue with respect to investment positions that are not closed by the end of the sampling period. This issue is solved by subsequent studies utilizing survival analysis methodologies that allow for censoring such positions, so that their data prior to censoring can be included in the result set.

They find out that positions with longer round-trip durations have a lower rate of continuously compounded and annualized returns (5.7% for the longest category versus 40% for the shortest category, both after transaction costs). This finding is

[2] Complete cycle from purchase to sale in long positions, or from sale to purchase in short positions.

correctly attributed to psychological factors and a tendency or disposition[3] of investors to sell securities in the gain domain earlier with quick profit realization, and hold securities in the loss domain longer in order to recover the losses. However, the authors refute this explanation on the grounds that proportion of positive rates of return in the shorter durations does not exceed the same proportion in longer durations.

Kahneman and Tversky (1979) basically start a criticism towards expected utility theory's assumption that outcome utilities are probability weighted. They introduce the certainty effect by hypothesizing that more certain outcomes are overweighted as compared to less probable ones. For positive outcomes, in cases where winning probabilities are substantial (e.g. 0.45 and 0.90) people choose the prospect with higher probability (even if gain is smaller). In cases where probabilities are very low (e.g. 0.01 and 0.002), people choose the prospect with higher gain (even if the probability is smaller).

However, when the signs of outcomes are changed to negative, the preference of prospects are reversed. This reversal of preference around 0 gains are labeled as the reflection effect. It implies that people are risk averse with positive outcomes while risk loving with negative outcomes. Incorporating the certainty effect to the interpretation of preferences, people exhibit risk averse behavior preferring a smaller sure gain over a larger but less probable gain, and they exhibit risk loving behavior preferring a larger but less probably loss over a smaller sure loss. These interpretations form the basis of prospect theory. According to their theory, agents perceive outcomes as gains or losses instead of final wealth. And gains or losses are calculated relative to a reference point. Reflecting the risk behavior defined above, the

[3] So disposition term is coined to the phenomenon in later studies.

value function is concave above the reference point (risk aversity in gains) and convex below the reference point (risk loving in losses). Furthermore, the function is steeper in the loss region, since the disutility of losing is greater than the utility of winning the same amount.

Kahneman and Tversky (1982) elaborate on and introduce new concepts related to prospect theory. Providing intuitive examples, they propose that when a situation is presented or framed in a different setting preserving the outcomes, the decision may change. With framing, the reference points that events are evaluated for gains or losses may be altered to induce difference decisions. Following the reference point argument, they also theorize the regret aversion issue that is previously mentioned in Thaler (1980).

Kahneman and Tversky (1982) are the first to mention that gains and losses can be assigned to separate mental accounts. Thaler (1985) provides a more detailed explanation for mental accounting of gains and losses based on value function as per the prospect theory. He applies the idea to household budgeting process as an example. In mental accounting, there are two ways consumers may value the outcomes of multiple events:

- When the outcomes are valued jointly, *integration* happens.
- When the outcomes are valued separately, *segregation* happens.

He proposes four possible combinations, in each of which either integration or segregation occurs on the basis that consumers maximize their values according to the value function as per the prospect theory:

- Multiple gains are segregated.
- Multiple losses are integrated.

- In the case of a larger gain and a smaller loss (so with a net gain), outcomes are integrated.
- In the case of a smaller gain and a larger loss (so with a net loss), outcomes are segregated (silver lining principle).

Shefrin and Statman (1985) form the starting point of disposition effect as a new area of study. Shefrin and Statman (1985) build upon the initial empirical findings by Schlarbaum et al. (1978) and normative optimal gain/loss realization strategy by Constantinides (1984), and they develop a theory of disposition effect. Schlarbaum et al. (1978) find that stock positions with longer round-trip durations have a lower rate of return. Constantinides (1984) mentions a tax regime where long term capital gains and losses are taxed at a rate half of that used for taxing short term capital gains and losses. He proposes that under such a tax regime, investors should optimally defer realizing their gains, while they realize their losses in the short term. Shefrin and Statman (1985) extend an earlier model introduced in Shefrin and Statman (1984). In that model, they suggest investors are willing to pay more tax to receive cash dividends instead of capital gains. Hence, these two options are not perfect substitutes. This willingness is explained on behavioral bases:

- According to self-control theory (Thaler, 1980; Thaler & Shefrin, 1981), an individual can be considered a long-term planner and short-term doer, schizophrenically at the same time. The goals of the two roles may conflict such as a planner role with the goal of long-term savings and a doer role with the goal of short term consumption. The conflict between the dual roles of this model of the individual consumer resembles the principal-agent conflict. This conflict is resolved either by giving a discretion to the doer role with modified incentives and preferences or by limiting the choice set of the doer with a rule. In the context of Shefrin and Statman (1984), a rule for prohibiting capital spending for the doer fulfills the long term savings goal of the planner. And doer has more discretion in the case of cash dividend payout.

- Shefrin and Statman (1984) use prospect theoretic explanation of individual behavior which is risk averse in the gain domain while it is risk loving in the loss domain (Kahneman & Tversky, 1979). They also use a preliminary version of the mental accounting system first defined by Kahneman and Tversky (1982) and later formalized in (Thaler, 1985) for the cases for segregation and integration of gains and losses. And they propose that the capital gains and dividend payout from stock investments are segregated to maximize the prospect theoretic value. Considering dividend payout always yields a gain, in the case of large positive capital gains, segregation results in maximizing the value from both gains. In the case of large capital losses, the segregation of dividends from capital losses results in a consolation, coined as silver linings.
- Regret aversion argument is built upon Kahneman and Tversky (1982) and Thaler (1980). Thaler (1980) relates regret to concepts of guilt and responsibility in actions and first observes the preference of inaction to action in many contexts for managing regret. Kahneman and Tversky (1982) mention that regret is more severe for contexts in which the actors are more close to achieve a gain but miss the opportunity, although the final set of outcomes are the same across the contexts. So the feeling of regret is determined by the reference point by which the outcomes are compared. Regret can be formally defined as the ex-post feeling when an alternative action or inaction ex-ante would lead to a better outcome. They mention the

observation that regret from inaction is less severe than regret from actions and regret from conventional or routine action is less severe than regret from innovative action. In the context of dividends, Shefrin and Statman (1984) propose that regret from a consumption decision using proceeds from dividends is less than that using proceeds from sales of stocks when the stock price increases thereafter. Hence, dividends are preferred to mitigate regret.

Shefrin and Statman (1985) apply these building blocks to the argument of disposition effect, tendency of investors to realize profits too early while they defer the realization of losses too long:

- Prospect theory explains the risk-loving behavior in the loss region and risk-averse behavior in the gain region. And the theory also predicts the existence of disposition effect when a reference point for gains/losses persist as opposed to resetting the reference point by closing the positions.
- Mental accounting provides an explanation for evaluating the losses and gains separately, segregation, or combining the outcomes in a portfolio, integration. A stock purchase triggers the opening of a new mental account for the investor with the cost of the stock acting as the natural reference point. So proceeds reinvested in new stock positions are not considered as a part of the same continuous account by subsequent mental accounts.
- Regret aversion explains the difficulty in realizing a position. In the context of disposition effect, realizing a position at loss admits a wrong decision while sticking to the position at loss creates a relief that the paper loss will disappear at the reversal of the stock prices.

- Self-control explains the motives for realizing a position. Closing a position at gain induces the feeling of pride based on the correct decision as a part of the self-control of the doer.
- They also consider the potential gain from the tax optimizing strategy of Constantinides (1984). They conclude that heavy tax selling decisions in December does not conform with long term rational tax planning, but it can be explained by behavioral reasons due to the perception of December as a natural deadline.

Shefrin and Statman (1985) interpret the findings of Schlarbaum et al. (1978) using the measure of round-trip duration, the time length between the purchase and sale of a certain stock, in the context of disposition effect. They conclude that actual findings on the share of gain and loss in realized positions does not confirm the shares as suggested by tax induced trades, and interpret that disposition effect induced trades offset tax motivated trades. To support the thesis, they also utilize aggregated data on mutual funds trades. And they interpret the fact that share of positions closed on gain is higher than the share of positions closed on loss as an evidence of the existence of disposition effect. A shortcoming of this study, also mentioned by the authors, is that transaction level data are not available for funds or individual investors.

After Shefrin and Statman (1985), there are only two studies published in the field of disposition effect until 1998. Ferris, Haugen, and Makhija (1988) predict year-end aggregate stock trading volume in line with disposition effect model and not tax-loss-selling model according to which traders realize their losses for tax optimization purposes. Weigand (1993) mentions disposition effect as a possible explanation for the correlation between stock prices and trading volume. There are no further studies on disposition effect conducted until 1998. Possible reasons for this ignorance might be as follows:

- The lack of availability of investor and transaction level data
- The lack of a plausible measure and model to quantify disposition effect
- The insufficiency of the existing hardware and software to deal with large size of investor and transaction level data even if that data existed

Odean (1998) surged the interest in the study of the disposition effect. Two main innovations he makes over and above previous two studies on disposition effect are utilization of investor and transaction level data and the development of a new measure to quantify the disposition effect

The dataset is provided by a large discount brokerage house and filtered for randomly selected 10,000 customer accounts. Records cover the 1987-1993 period. Multiple buy or sell transactions of an investor's position in a stock in a single day are aggregated separately. He defines proportion of gains realized (PGR) as the realized share of positions that are at a gain, such that they are trading above the investor's purchase price. Proportion of losses realized (PLR) is defined accordingly for positions trading at a loss. The PGR/PLR ratio, subsequently coined as Odean's measure in many citing studies, shows disposition effect above 1, so the propensity of investors to close positions at a gain is higher than the propensity to close position at a loss. He finds that except for December, PGR/PLR ratio is above 1. In December, realization of losses increase due to tax selling by investors. The study mentions that the price path after the purchase may also affect the reference point used by the investor. However, due to the difficulty in analysis to consider the true reference point, only the purchase price is accepted as the reference point.

An updated version of the dataset in Odean (1998) from the large discount brokerage house covers the 1991-1996 period, consists of 78,000 accounts and was first used in Barber and Odean (2000). The dataset is known as the large discount brokerage dataset or LDB dataset. The LDB dataset is a staple in the disposition effect literature. Among studies using the LDB dataset are Akay (2008), An (2015), Balkanska (2018), Barber and Odean (2013), Ben-David and Hirshleifer (2012), Brettschneider and Burgess (2017), Brettschneider, Burro, and Henderson (2021), Chang, Solomon, and Westerfield (2016), X. He (2020), Heimer (2016), Ivkovich, Poterba, and Weisbenner (2005), Ivkovich and Weisbenner (2005, 2006), Kotomin and Varma (2022), Meng (2010a, 2010b), Peng (2017), Sakaguchi, Stewart, and Walasek (2019), and Strahilevitz, Odean, and Barber (2011). So almost 30 years after the dataset was first collected, it is still being a subject of new research studies.

Weber and Camerer (1998) are the first to study disposition effect incorporating data generated from dedicated experiments. Using student subjects, data are gathered from questionnaires. The difference between the shares of winners (shares trading at a gain) and losers (shares trading at a loss) in sold shares is used as a simple measure for signaling disposition effect. The difference between the overhangs of sold and kept shares also confirms the results. Apart from these aggregate results, an individual measure is also developed as the difference of the number of winner sales and loser sales in the total number of closed positions of each investor. This measure is called disposition coefficient and two thirds of the subjects have values above 0, signaling the existence of the disposition effect. An important point the study finds is that, in an alternative experiment, when positions are forced to be closed at the end of each period, the disposition effect is significantly weaker across subjects. The alternative explanation of belief in mean reversion is rejected since the experiment design is done such that the stock prices are positively correlated, demonstrating momentum effect.

While Odean (1998) provided a stimulus for disposition effect studies, it was not until 2001 that the number of studies increased significantly. Genesove and Mayer (2001) bring two innovations to the field: They carry the study field into a new market, the housing market. And they are the first to apply Cox proportional hazards model, which is a better method to analyze disposition effect at the investor and position level as compared to PGR/PLR ratio, in this field. Cox model is used as a supplementary model to ordinary least squares regression method. While the effect being estimated is basically disposition effect, throughout the study, this term is used nowhere other than the references. *Loss aversion* term is preferred instead. The expected selling price is used for determining the levels of losses with respect to the original purchase price. Findings confirm the existence of loss aversion: Higher prospective losses lead to reduction in sales.

While previous studies utilized datasets that are small samples of the whole corresponding markets, Grinblatt and Keloharju (2001) are the first to include the whole transactions of two years in the Finnish stock market. Logistic regression method is used for analyses. Returns for 11 non-overlapping horizons are also included in order to separate the disposition effect from contrarian strategies (selling winner and buying loser stocks). In both positive and negative past returns, the capital losses overhang variables significantly decrease the propensity to sell, confirming the existence of disposition effect, separate from contrarian investment behavior in response to past returns in different time horizons.

Shapira and Venezia (2001) find evidence for disposition effect across both individual investors and institutional investor in the Israel market. They simply compare the round trip duration of losing and winning positions.

Feng and Seasholes (2005) focus on whether investors' sophistication and trading experience reduce disposition effect and show that existence of both eliminates disposition effect in Chinese capital markets. They use a parametric survival model using Weibull hazard specification. Each record is a unique combination of an investor's position for each day. Independent variable is the dichotomous status for sell or hold for each record. An innovation they bring is to discretize several numeric features, such as age or experience as measured by cumulative number of positions for each investor, into ordinal categories and include those categories as dummies in the regression. Covariates fixed at investor level are combined with time-varying covariates. In this way, no functional shape is imposed on the relationship between the independent and dependent variable. They cite the ability to use all available data when a position is not closed and the interpretability of the hazard ratios as reasons to prefer survival methods to Odean's PGR/PLR approach. They emphasize that the price path information can be incorporated into the analysis by using all daily data between purchase and sale (or the end of the date range of the data in the case of censored positions). This study forms a great portion of the blueprint of survival analysis in disposition effect research.

Shumway and Wu (2005) build upon and progresses the discretization approach of Feng and Seasholes (2005) in survival analysis further for Shanghai Stock Exchange data. They separate the capital gains/losses overhang levels into ordinal categories and include them as dummies in a Cox proportional hazards model. In this way, they also do not impose a shape on the relationship between overhang levels and hazard ratios. The dichotomous overhang variables are time varying covariates in the model. For positions with multiple sales, only the portion up to the first sale is included in the model. The approach in Feng and Seasholes (2005) and Shumway and Wu (2005) is replicated and further developed in many subsequent studies. They also form the basis of the methodological approach followed in this thesis. They find that disposition effect exists among Chinese investors. Disposition prone investors' performance worsens in subsequent years. And interestingly more disposition prone investors diversify their portfolios to a larger extent.

Using Chicago Board of Trade data and a Cox model, Coval and Shumway (2005) find that traders demonstrate disposition effect in afternoon transactions. However, apart from the overall gain or loss on the position, the level of profits in the morning session also helps explain the propensity to close a position.

Dhar and Zhu (2006) show that sophistication or financial literacy of investors, proxied through demographic characteristics such as wealth or professional occupations is inversely related to the extent of disposition effect. They use Odean's measure and also a probit regression in the analysis.

Chen, Kim, Nofsinger, and Rui (2007) extend a similar analysis to Chinese investors. They show that certain traits of investors, such as experience proxied by time since account was first opened, being at middle age, frequent trading and wealth and living in a cosmopolitan area decrease the investors' proneness to disposition effect. Similar to Feng and Seasholes (2005), they use a parametric survival analysis with Weibull hazard specification.

Based on artificially created simulation data and theoretical formulations, Barberis and Xiong (2009) differentiate between annual gains/losses and realized (or paper) gains/losses for the holding period. They show that realized gains/losses can predict the disposition effect more reliably.

Using Finnish stock market data and logit regression, Kaustia (2010) refutes the rational explanations such as portfolio rebalancing or belief in mean reversion for disposition effect. He favors alternative psychological interpretations such as self-justification or self-deception, while no such test is conducted for those interpretations.

Seru, Shumway, and Stoffman (2010) calculate a disposition coefficient for each investor and year using Cox model and PGR/PLR ratio methods. They use these coefficients to explain the return of the investors for the following year. They show that investors in the lower disposition quintiles have higher returns, and this relationship becomes more apparent and monotone in longer return horizons.

Strahilevitz et al. (2011) examine loss aversion from a different perspective. Using a Cox model with discretized return categories for a previously sold position, they try to explain the probability of repurchasing the same stock. Greater degree of a previous loss decreases the propensity to repurchase the stock. However, for the gain region, after a sharp increase in the propensity to repurchase at lower gain levels, the probability of repurchase flattens and even decreases, creating an inverted V-shape.

The analysis is repeated for returns since the previous sale and a V-shaped relationship is revealed between the return bins and the propensity to repurchase with the minimum level observed around zero returns. Their interpretation of the findings is that negative emotions like regret and disappointment influence investors' decision more than positive emotions like rejoicing and elation do. An important point is that they emphasize the simultaneous existence of positive and negative emotions and their offsetting effect. They also cite the emotion of fear that prices may continue to drop.

Hens and Vlcek (2011) create a theoretical argument in which they differentiate between ex-post and ex-ante disposition effects. According to their argument, existing literature can only test ex-post disposition effect and assumes that investors have bought the stocks in the first place. Their analytical model that incorporates a risk-free asset and a risky asset predicts that investors prone to disposition effect with a prospect theoretic explanation would not have invested in those stocks in the first place. So they conclude that prospect theory can explain the ex-post disposition effect but not the ex-ante disposition effect.

In a theoretical and analytical study, Barberis and Xiong (2012) propose the concept of *realization utility* which posits that investors experience a negative (positive) utility from realizing gains (losses) on their investments. In their infinite horizon model, an investor allocates the wealth between a risk-free asset and a risky asset in order to maximize their sum of discounted expected future utilities. So apart from the utility or disutility from realization of a gain or loss, the discount rate provides an additional feature to answer the question of why the investor would realize a gain today instead of realizing a possible larger gain in the future. According to the study, realization utility model provides a view of the disposition effect that is related to but distinct from the view provided by the prospect theory.

Building up on the realization utility view, Ben-David and Hirshleifer (2012) define two types of realization preferences to explain the existence and the form of the disposition effect. Sign realization preference is the preference for realization of gains to losses regardless of the magnitude. This type of preference implies a jump and discontinuity at zero profit. Magnitude realization preference is the preference for realization of a gain or loss with a higher magnitude to those with a lower magnitude. This preference do not require a jump at zero profit level. Using the LDB dataset and a probit regression model, they also confirm the existence of the magnitude realization preference. They find that for short holding periods, investors realize larger losses more readily then smaller losses. And they find no evidence for a jump at zero profits. So they cannot conclude that disposition effect with a static realization preference with respect to a fixed purchase price as a reference point exists among US investors. An important point they mention is that investors may confirm their beliefs towards the larger gain domain and act speculatively to purchase more shares and increase their holdings.

Grinblatt, Keloharju, and Linnainmaa (2012) combine Finnish data on order book and IQ scores of males to find that higher IQ diminishes proneness of investors to disposition effect.

Frydman and Camerer (2016) conduct a neural experiment to detect regret signals in the brains of the subjects as a reaction to foregone purchase opportunities that have recently created positive returns and that would have created further positive returns if they were purchased. A region of the brain that is active in reward processing is focused for collecting signals. The subjects that commonly make stock repurchase mistakes are shown to exhibit disposition effect to a greater extent.

Using data collected from a social networking platform and utilizing Cox regression, Heimer (2016) finds that level of access to investment-specific social networks increases the disposition effect of investors.

In an experimental study, Chang et al. (2016) focus on the effects of delegation of investment decisions and cognitive dissonance on disposition effect. In non-delegated assets, like individual stocks, investors exhibit a strong disposition effect, while in delegated assets like mutual funds, disposition effect is to a lesser extent or even reversed. Cognitive dissonance means having two conflicting cognitions simultaneously. In the context of investments, the belief of an investor in their ability to make successful investment decisions and the observation that the value of an investment declined is an example of cognitive dissonance. The study shows that low level of cognitive dissonance experienced by the investors diminishes the disposition effect that investors demonstrate.

2.3 Main themes in the disposition effect literature

Brief literature review on selected main themes in disposition effect related studies that are instrumental in the development of the hypotheses of this thesis is presented below.

A common finding of V-shaped disposition effect posits that the direction of the relationship between propensity to sell and the level of capital gains/losses overhang changes direction across different regions of the relationship. This phenomenon is taken as a starting point in order to explore whether the observed V-shaped effect is the result of the simultaneous and competing effects of multiple covariates, each of which may have a monotonic relationship with the propensity to sell.

Alternative or multiple reference points argument explored in many studies posits that there may be reference points other than the original purchase cost that the investors may anchor to. These reference points may be fixed throughout the investment horizon or be updated as the market evolves. While in most studies they are mentioned or tested separately, yet few studies explore their combined, simultaneous and interactive effects on the propensity to sell. This theme also forms one of the pillars of the hypotheses development in this thesis.

Highest or lowest prices have commonly been used as alternative reference points in disposition effect studies. They act as reference points due to their salience.

Many studies find reverse disposition effect, a lower propensity to sell towards the extreme gain region and higher propensity to sell towards the extreme loss region under certain circumstances. V-shape evidence also partially documents this effect. Reverse disposition effect observed across overhang levels on alternative reference points is also a major theme that this thesis explores.

Information salience is how visible or prominent information is. The more a reference point or any other information is salient, the more an investor may incorporate this information in their investment decision. This argument is used in this thesis to filter transactions or process data in order to control the salience of price information.

In behavioral studies, emotions and biases help to explain the anomalies in data and certain effects as relationships between observed behaviors and covariates. Especially in the case of combined effect of multiple reference points, competing or reinforcing emotions may act together across multiple dimensions.

2.3.1 V-shaped disposition effect findings

Many studies in the field of disposition effect research present evidence on the V-shaped[4] relationship between the propensity to close a position and explanatory variables' ordinal levels, especially the capital gains/losses overhang with respect to a reference point.

Meng (2010a) documents a V-shaped relationship between the propensity to hold or close a position and the capital gains/losses overhang with respect to investors' reference points. Incorporating the idea of expectations as reference points, he reports that the bottom of the V-shape points at 5.5% capital gains which is interpreted as the average expected gains of investors.

Ben-David and Hirshleifer (2012) propose magnitude realization preference in order to explain the V-shaped probability of selling a position of US investors with a minimum level at zero profits as reported by the study.

Seiler, Seiler, and Lane (2012) report that 7.3% percent of the sample of real estate investors exhibit a V-shaped or inverted V-shaped pattern. In this pattern, the propensity to sell is at a maximum or minimum around the reference price, as compared to extreme gain or loss regions.

[4] The finding is alternatively explained as a U-shaped or hump-shaped relationship across studies.

An and Argyle (2014) find that mutual fund managers exhibit a V-shaped selling schedule across gain/loss levels, complementing the argument of Ben-David and Hirshleifer (2012) for individual investors.

Brown and Yang (2017) find that V-shaped selling pattern exists in the soccer betting market and the introduction of a feature that makes the bettor's gain or loss more salient makes the shape more pronounced.

Peng (2017) proposes that investors act according to the law of small numbers, belief in the representativeness of small samples, and their expectations on continuation or reversal differ in short-term versus long-term trends. The V-shaped selling propensity is shown to be the result of aggregation of the behavior of separate investor groups.

Conducting a Fama-MacBeth regression, Waiyasara and Padungsaksawasdi (2020) find that the strength of the V-shaped disposition effect depends on the investment horizon of Thai investors. The effect is stronger with investors having short-term horizons.

Running a Fama-MacBeth regression on Korean data, Kim, Kim, and Kim (2021) interpret V-shaped disposition effect according to reaction of speculative investors to new information arrivals. When good news arrive at the market, speculative investors with high capital gains overhang realize their positions, since current prices already matches the expected levels. On the contrary, when bad news arrive, speculative investors with high capital losses overhang change their initial beliefs and sell off their losses.

Dai, Jiang, Liu, and Xu (2022) incorporate rational portfolio rebalancing with transaction costs and unknown expected returns to explain disposition effect and the

V-shaped selling pattern. They define two opposing forces that simultaneously exist, namely exposure effect and learning effect. Exposure effect drives investors to follow a contrarian strategy to sell winners and buy losers. Learning effect induces momentum trading strategy to buy winners and sell losers due to revised estimates of expected returns.

2.3.2 Multiple reference points and reference point adaptation

Kahneman (1992) mentions the existence of multiple reference points that an actor can value the outcomes against. He posits that, due to framing effect, people may change the formulation of the same problems using different reference points. And thus gains can become losses and vice versa. He foreshadows that a research topic is to explore how multiple reference points compete and combine, at individual preference level or in negotiations.

Boles and Messick (1995) use norm theory, which suggests that norms, meaning contexts, standards or reference points in evaluation, are calculated after an outcome occurs. And an outcome itself may create its own alternative reference points that could have existed. As a result there may be two reference points, one that is used to evaluate the current situation as a gain or loss, and the other one being from the perspective of the alternative outcome, such as an alternative gamble that is not chosen.

Sullivan and Kida (1995) argue that multiple reference points are determinants of risky decisions by managers. In an experimental design, managers are asked for an investment decision in a context which includes both a current benchmark performance and a target performance. They conclude that managers consider both reference points when a risky option with an expected performance which is between these points is evaluated.

Odean (1998) mentions four alternative reference prices for the calculation of profit or loss overhangs, namely first, last, highest and average purchase prices. However, without any detail, findings for all are reported as the same, taking the average purchase price as the basis for analysis.

Weber and Camerer (1998) find that disposition effect exists for two separate reference points, the purchase price and the previous price. However, how reference points adapt over time and multiple points are balanced are not explored further in the study.

In the field of consumer behavior, Kőszegi and Rabin (2006) mention the possibility of multiple personal equilibria across references. For example, in a consumption decision, a consumer may evaluate the gain or loss of acquiring or not acquiring a new consumer product simultaneously with the gain or loss of keeping or paying the money in the buy or not buy decisions.

In a field survey, Lin, Huang, and Zeelenberg (2006) investigate the regret, as defined by the comparison of "what is" and "what might have been". They test whether multiple reference points such as non-investment outcomes, expectations and outcomes with the highest and lowest performances determine regret. They find that the non-investment or inaction outcome has the strongest impact on regret. The results also confirm that regret is mainly driven by the sign of outcome (gain or loss) relative to each reference point, rather than the magnitude of the gain or loss.

Svedsater, Karlsson, and Garling (2009) show that investors rely on both short-term and long-term reference points for selling or buying decisions. A positive correlation between short-term returns and trading behavior is observed, suggesting momentum trading by investors. However, a negative correlation between long-term returns and selling behavior is found, suggesting disposition effect.

While Baucells, Weber, and Welfens (2011) do not test disposition effect directly, they conduct an experiment with subjects to reveal the dynamics of the formation of reference prices that produce zero experienced utility or emotional neutrality. As the price path evolves in the experiment, purchase price, current price, average price, highest price and lowest price are presented as reference price alternatives. The findings signal that reference price is not updated by averaging new information with old prices across the path, but the most recent price is highly weighted along with the original purchase price. Investors reveal to have experienced a gain if the most recent price is above the purchase price and has moved up in the last period, and vice versa.

Koop and Johnson (2012) conduct an online experimental study with college students who are presented with gambles. In line with the tri-reference point theory, subjects consider three reference points in risky decisions: Minimum requirement, status quo and goal. They show that multiple reference points have joint and differential impact on behavior.

Apart from the purchase price, Quispe-Torreblanca, Gathergood, Loewenstein, and Stewart (2020) show that investors also consider the price viewed during investor's last login into their account through the interface of the security house.

Using UK home sales data, they test and confirm the peak price each investor experiences during their holding period as a reference point additional to the purchase price. Their explanation is that peak price is a salient or prominent reference point that investors pay attention to.

2.3.3 Highest and/or lowest prices as reference points

High and/or low mark of prices have frequently been considered as additional or alternative reference points.

For example 52-week highest or lowest prices are chosen as extreme reference points in Bharandev and Rao (2020) and Della Vedova, Grant, and Westerholm (2016).

Some studies specifically consider the extreme high and/or low prices experienced by investors since their purchases as reference prices. Among those studies are Han (2019), Ivkovich, Pearson, and Weisbenner (2003), Ivkovich, Pearson, and Weisbenner (2007), Kaustia (2010), Quispe-Torreblanca, Hume, Gathergood, Loewenstein, and Stewart (2021), and Zimmermann (2015).

2.3.4 Reverse disposition effect

Apart from a significant number of studies with evidence on disposition effect, reverse disposition effect is also reported in numerous studies. Reverse disposition effect is the lower propensity to close a position towards the higher gain region and a higher propensity to close position towards the higher loss region.

Ranguelova (2001) finds that investors exhibit reverse disposition effect on their positions in small cap stocks. The relationship between firm size and the direction and strength of disposition effect is monotonic: For larger cap stocks, propensity to sell winners (losers) increases (decreases). Hence, investors exhibit less reverse disposition and more disposition effect.

In a similar vein, da Silva Rosa, To, and Walter (2005) confirm that for each size group of stocks, institutional investors, who have a higher sophistication level as compared to individual investors, exhibit reverse disposition effect. The apparent disposition effect in the aggregate level can be due to fund managers' preference for selling larger and more liquid stocks.

In a theoretical argument, Barberis and Xiong (2009) posit that investors may have a higher propensity of selling a stock and exhibit reverse disposition effect through an analytical model with certain parameter settings for loss-aversion and expected returns.

Krause, Wei, and Yang (2009) present evidence for the existence of reverse disposition effect during falling markets with low returns. They also document that for sufficiently short investment horizons, investors exhibit reverse disposition effect. They do not provide a theoretical explanation for the relationship of these factors on the disposition effect.

Dorn and Strobl (2010) posit that disposition effect is a rational response of less informed investors to high information asymmetry. However, when information asymmetry is lower, less informed investors demonstrate reverse disposition effect.

In an experimental setting with college students and a simulated stock market, Wang, Cao, Yuan, and Cheng (2010) show that investors exhibit reverse disposition effect in downward markets.

Talpsepp (2011) contrasts the behavioral characteristics of foreign and local investors in the Estonian stock market using Cox model. Foreign investors who are more momentum driven in their trades exhibit reverse disposition effect, while local investors who pursue contrarian strategies exhibit disposition effect.

Kuo and Chen (2012) propose the idea of maximum loss tolerance and minimum value threshold for each investor. They assume that investors do not tolerate losses larger than the maximum limit, and up to that point investors hold losers. And those investors are not satisfied with gains smaller than the minimum threshold. So gains are realized after that threshold is met. In this setting, when absolute value of the maximum loss tolerance is larger than the minimum value threshold, investors exhibit disposition effect. In the opposite case, reverse disposition effect is observed. The assumptions are confirmed by a survey of Taiwanese investors.

Using Taiwan Futures Exchange data, Cheng, Lee, and Lin (2013) document that investors with low trading volume show the strongest disposition effect. This effect diminishes with higher trading volume and is eventually reversed in the investor groups with the highest trade volume.

In an important study, Gärling, Blomman, and Carle (2017) bring an emotional explanation for disposition and reverse disposition effect by interpreting the relationship between capital gains/losses overhang on two emotional axes:

Elation-disappointment balance which evokes the disposition effect related behavior and hope-fear balance stimulating the reverse-disposition effect related behavior. The relative steepness of these balances determine the observed disposition or reverse disposition effect.

Balkanska (2018) shows that higher information uncertainty drives disposition effect. When information uncertainty is measured by analyst forecast dispersion, investors exhibit disposition effect in stocks with higher analyst forecast dispersion and reverse disposition effect in stocks with the lowest analyst forecast dispersion. The results persist controlling for size, and the relationship is stronger for smaller cap stocks.

In an experimental design, Grosshans and Zeisberger (2018) create three types of simulated price paths separately for winner and loser stocks. Those paths are revealed to subjects who are instructed to assume they bought the stocks at the beginning of the simulated paths. Straight path follows a rather steady decline for loser stocks and rise in winner stocks. Up-down path increases first and decreases later. Down-up path decreases first and increases later. In the straight path scenario without significant turning points, investors believe in trend continuation and disposition effect is not demonstrated. Due to upward updating of the reference point, investors exhibit low propensity to close winning positions and the disposition effect is even reversed. However, in the non-straight path cases, investors show disposition effect.

In experimental studies, Aspara and Hoffmann (2015) and Wierzbitzki and Seidens (2018) show that when investors are given a specific performance goal, they exhibit reverse disposition effect.

da Silva, Mendes, and Abreu (2020) document that both trading experience and trading in a more diversified manner, meaning trading a higher number of securities or funds, reverse the disposition effect.

Haryanto, Subroto, and Ulpah (2020) test the existence of disposition effect in cryptocurrency markets. Using data from cryptocurrency exchange Mt. Gox and using Cox model, they show that cryptocurrency investors exhibit disposition effect in bearish markets, while they exhibit reverse disposition effect during bullish markets.

2.3.5 Information salience

Frydman and Rangel (2014) conduct an experiment in which two conditions are tested. In a high-saliency condition, a trading software makes the purchase price more visible to the subjects. In the low-saliency condition, the purchase price is hidden. Disposition effect is found to be stronger in the high-saliency condition.

Brown and Yang (2017) explore the effect of salience on disposition effect in betting markets. Introduction of cash-outs feature, which continually exhibits updates and makes the profit or loss more salient, increases disposition effect.

Using data from a Chinese brokerage house, Frydman and Wang (2020) show that an increase in the salience of an investor's purchase price in a stock and salience of the respective capital gains/losses make the disposition effect stronger. The salience increase is done through making the purchase price and investor's return information more visually apparent on investor's online account screen. While the data used are actual trade data, the salience shock through the change in user interface is considered an experimental feature. So the data before and after the salience shock are compared for a change in the level of disposition effect. Data are analyzed first by ordinary regression and then the results are confirmed by applying the Cox model.

2.3.6 Emotions

Interpretation of behavioral biases found through empirical studies using theories on emotions in psychology is an important attempt to base the findings on sound theoretical grounds.

Lopes (1987) defines a two-factor theory to understand how individuals act when confronted with risky choices. The *dispositional factor*[5] denotes the motives that dispose people to be security oriented (risk-averse) or potential oriented (risk-seeking). The *situational* factor shows how people respond to immediate needs and opportunities. The dispositional factor is also called as security/potential factor and reflects how individuals respond to or perceive risks. The situational factor is also called as aspiration level and reflects current opportunities or what individuals can get and constraints arising from the environment or what individuals need. The importance of this theory is that individuals make choices according to these two factors simultaneously, and these factors can be in conflict or in concert in different situations. So risk-averse and risk-seeking choices appear at the same time in the behaviors. In subsequent literature, this theory is coined as *security, potential and aspiration theory* or *SP/A theory* for short. The simultaneous effect of multiple behavioral axes which can be in conflict or in consort is a major theme that the findings of this thesis can be based on.

[5] Not to be confused with disposition effect, the main theme of this thesis.

Tversky and Kahneman (1992) develop a new version of their previous prospect theory, namely cumulative prospect theory, to cover more diverse situations. The experimental evidence confirms a fourfold pattern for risk attitudes. Individuals are risk averse (risk seeking) for high (low) probability gains and risk seeking (risk lowing) for low (high) probability gains.

Lopes and Oden (1999) compare the ability of their SP/A theory and cumulative prospect theory in fitting the data collected from experiments involving different kinds of transformed lotteries. They show that SP/A can capture the complex and non-monotonic preferences in some lotteries, while cumulative prospect theory fails to do that.

Lo, Repin, and Steenbarger (2005) contrast automatic emotional responses mediated by amygdala in the brain, such as fear and greed, and higher-level responses mediated by prefrontal cortex, such as complex decision making in investments. Emotional responses can interfere with higher-level responses to result in suboptimal trading performance.

Summers and Duxbury (2007) posit that emotional balance between anticipated regret and rejoicing drives disposition effect. Individuals experience disappointment or elation towards losses or gains for which they are irresponsible, while they also experience regret or rejoicing respectively when they are responsible. So when choice and responsibility are introduced into the framework of gains and losses, rejoicing and regret cause disposition effect.

Mauss and Robinson (2009) summarize two perspectives on conceptualization of emotions. Discrete perspective suggests that each distinct emotion has its own unique factors. The dimensional perspective, however, posits that emotions can be measured along some common factors such as valence, arousal and motivational state. Favoring the dimensional perspective, the study reviews different methods and metrics for measuring emotions.

Lo (2011) reminds the role of fear and greed in boom and bust cycles in recent history. They review evidence from cognitive neuroscience literature for the link between emotions and rational behavior. In the physical world, fear triggers fight or flight response, which is critical for survival. In financial decisions also, moderate fear may induce investors to act diligently, while extreme fear can cause irrational behavior. The human brain has a reward system, and the signal for the rewards is transmitted through dopamine. An imbalance in human brain's dopamine system can induce an investor to take greater risks and this greedy behavior can go into a potentially destructive positive-feedback loop upon positive returns.

Summers and Duxbury (2012) emphasize that while influence of emotions on decision making is generally acknowledged, the pathways from specific emotions to the decisions are underexplored. In an experimental design involving five studies, they show that specific emotions are necessary to activate the responses envisaged by prospect theory, such as elation for selling winners and regret for holding losers in disposition effect. Under alternative scenarios, different emotions are evoked, causing behaviors inconsistent with disposition effect.

X. D. He and Zhou (2016) propose a portfolio choice model involving the role of three emotions: Hope, fear and aspirations. Three indices, namely fear index, hope index and lottery-likeness index are constructed to measure the respective emotions. Fear index is related to exaggeration of small probabilities of highly unfavorable outcomes. Hope index is related to exaggeration of small probabilities of highly favorable outcomes. Lottery-likeliness index reflects the payoffs in high level of aspirations. The model is inspired by SP/A theory and rank dependent utility theory. Rank dependent utility theory serves as the basis for fear and hope indices, whereas SP/A theory serves as the basis for lottery-likeliness index for capturing aspirations.

Gärling et al. (2017) suggest the concept of affect account that explains how potential gains and losses stimulate affect, which determines the propensity to sell. In contrast to prospect theoretic explanation of disposition effect through risk preferences, affect account considers the impact of affect associated with potential gains or losses. Affect is an input to investor decisions. However, sophistication and experience override the impact of affect. Influences of affect are evaluated in two dimensions. In one dimension are the *anticipatory* feelings of hope of earning and fear of losing associated with positive or negative price changes respectively. In the other dimension are the *anticipated* feelings of elation and disappointment associated with decisions to realize gains and losses respectively. Anticipatory feelings are stimulated by expectations of earning or losing not associated with any specific outcome and those feelings are imagined. Anticipated feelings are linked to outcomes of specific choices and those feelings are felt. The hope-fear balance and elation-disappointment balance occur simultaneously as price evolves. The relative changes in these balances determine the investment behavior and whether disposition or reverse-disposition effect is observed. The proposed two-dimensional emotional balance model is critical in the interpretation of the findings in this thesis.

Richards, Fenton-O'Creevy, Rutterford, and Kodwani (2018) use dual process theory to explain disposition effect. This framework defines two types of thinking. System 1 thinking, which involves emotion systems to mediate associative information processing, results in disposition effect. System 2 thinking, which involves effortful and reflective cognitive processes, reduces disposition effect.

2.4 Categorical view of data

Based on the discretization approach in many researches in the field of disposition effect, the numerical variables are also converted into ordinal covariates in this thesis. Categorical perception is the agents' perception of continuous phenomena as discrete categories. Apart from data originally ordinal or nominal in nature, continuous features can also be converted into categorical ones through numerous discretization methods. So a review is also conducted in order to provide the basis for the preference of the categorical approach to numeric processing of raw data.

2.4.1 Categorical perception of continuous phenomena

In this thesis, numerical features, that are capital gains/losses overhangs with respect to reference points, are converted into ordered categorical, or ordinal features to allow for a flexible relationship between the sell or hold status. This approach follows from the method taken by various studies in the field of disposition effect research. Apart from being a choice of data preparation, discretization of numerical features can be thought within a behavioral context.

Categorical perception is a qualitative difference in how things are perceived alike based on whether they are classified in the same category. Categorical perception is a psychophysical phenomenon. Individuals discriminate and identify things by comparing them with a set of stimuli. The stimuli are usually continuous and can be divided into regions using labels (Harnad, 1987). In other words, stimuli from different perceived categories are distinguished better than stimuli from the same category (Goldstone, Steyvers, & Larimer, 1996). This holds true even when the stimuli from different categories are close to the boundary from other sides (Nelson & Marler, 1989). And individuals can create new dimensions and categorize the same objects across these dimensions when they are asked for in experimental studies (Goldstone et al., 1996).

Phenomena may be categorized as distinct memberships or degrees, such as the case of colors, which may be categorized as distinct colors or into labels denoting lightness/darkness. Categories are important since they effect the way the world is perceived and acted upon. With the help of categories, the world is not perceived to be full of blooming and buzzing confusion, but instead perceived to be orderly with discrete objects. In time, perceived differences among the members within the same category are compressed, so that these members are perceived as more similar, while differences between distinct categories are separated so that members of different categories are discriminated better (Harnad, 2003).

VanRullen and Koch (2003) show that temporal dimension of presented stimuli is also perceived as discrete. With neuroscientific methods, this kind of perception is found to be arising from oscillations in different frequency ranges in some parts of the brain.

Categorical perception is also critical in speech cognition. When subjects are presented with a continuum of consonant-vowel type syllables, most subjects identify the sound as a different syllabus at around the same point (Goldstone & Hendrickson, 2010).

Yannakakis, Cowie, and Busso (2021) review methods to evaluate the intensity of emotions using ordinal, interval and nominal scales. In the affective computing field dealing with interpreting, processing or simulating human affect, a common experimental design is to instruct subjects to annotate the intensity of the feelings they perceive in videos. The literature suggests that using ordinal scales in annotation leads to more robust and reliable results.

Dutton (2021) posits that categorization of continuous phenomena is instinctive and an evolutionary adaptation of humans to their environments.

2.4.2 Data discretization

While individuals' perception of continuous phenomena is usually categorical as presented above, in data analysis, numeric data can be discretized into categorical or ordinal scales.

The advantages of discretization are as follows (García, Luengo, & Herrera, 2015; Muhlenbach & Rakotomalala, 2005; Ramírez-Gallego, García, Benítez, & Herrera, 2016; Ramírez-Gallego, García, Mouriño-Talín, et al., 2016):

- Suitability or requirement to work with some machine-learning algorithms like Naive Bayes
- A simpler, more comprehensible and better explainable representation of data

- Reduction and compression of data to be processed more efficiently and accurately in algorithms
- Ability to model non-linear relationships with simple model specifications,

while the downside is the loss of some of the information in the original data. Minimizing the information loss is among the goals of a discretization method.

While several categorization schemes and taxonomies are developed for discretization methods over the last three decades, these taxonomies have some common axes to categorize methods (Bakar, Othman, & Shuib, 2009; Dougherty, Kohavi, & Sahami, 1995; García et al., 2015; Ramírez-Gallego, García, Benítez, & Herrera, 2016; Ramírez-Gallego, García, Mouriño-Talín, et al., 2016; Yang, Webb, & Wu, 2010):

- In hierarchical methods, initial intervals are split into smaller ones iteratively through a top-down approach (splitting) or merged to consolidate into fewer intervals through a bottom-up approach (merging). In non-hierarchical methods the intervals are not split or merged iteratively.
- Supervised methods have target labels for the numeric data to be discretized, while unsupervised methods do not have prior labeling. They are further categorized based on the evaluation metrics, such as error, entropy or statistics based methods.
- Global methods, use the same scheme at once for the whole dataset, while local methods repeat the procedure for sub samples of data.
- Static methods discretize the data before executing a statistical or machine-learning algorithm, while dynamic methods discretize along the execution of models. This categorization is also known as eager vs. lazy methods.

- Univariate methods discretize variables separately, while multivariate methods consider the relationships and complex interactions among variables.
- Incremental methods start with a simple discretization in a certain number of intervals and improve the intervals by iterative improvement in subsequent pass through until a stopping criterion is satisfied, while direct methods discretize into intervals simultaneously.
- Parametric methods require user input such as the number of intervals to conduct the algorithm, while non-parametric methods require no starting parameters.
- Disjoint methods discretize data into intervals that do not overlap, while intervals in non-disjoint methods can overlap each other.
- Time-sensitive methods can classify a quantitative value into different categories as statistical properties of a time-series change in time. Time-insensitive methods use only the stationary properties of time-series data.
- Ordinal methods create categories that keep an ordering information, while nominal methods discard ordering information.
- Fuzzy methods employ a membership function measuring the degree of values being a member of each interval, so a single value can be a part of different intervals at the same time, while non-fuzzy methods employ sharp borders.

The easiest of the methods is binning, in which the data are split into predetermined number of intervals of equal width or equal frequency. The number of intervals is also known as the arity. An alternative method predetermines the frequency of each interval and is known as fixed-frequency discretization. Other discretization methods involve more complex algorithms such as multi interval entropy minimization discretization coupled with minimum description length principle, which uses entropy that measures how consistent a split matches a classifier. ChiMerge, StatDisc and InfoMerge methods are alternatives that employ a hierarchical and bottom-up approach through merging of intervals. Iterative dichotomiser 3 is a local method that creates and organizes classification rules into a decision tree structure (Dougherty et al., 1995; García et al., 2015; Yang et al., 2010).

Most discretization approaches in literature are univariate in nature. However, univariate approaches destroy and miss hidden patterns in the data. In multivariate approaches, the interactions among the variables are considered to determine the intervals to divide the data into categories (Bay, 2001). An example of multivariate methods is cluster-based discretization that work in the n-dimensional space by forming initial clusters and post-processing to minimize the number of clusters (Yang et al., 2010).

Another example of multivariate methods is relative unsupervised discretization for association rule mining, which is a context-sensitive method and reflects the interactions between all attributes. The method is unsupervised in the sense that there are no target classes, while it is supervised in the sense that the intervals are determined relative to other attributes. The method consists of three consecutive steps. In the pre-processing step, all continuous attributes are discretized in an unsupervised manner separately. In the structure projection, the structure of source attributes are projected into target attributes by filtering the data for each value of the source attribute, and a clustering is performed on the target attribute for each partition. Through iterations, all attributes become a target attribute, while other attributes act as source attributes in turn. In the post-processing step, the splits from the iterations in the previous step are merged (Lud & Widmer, 2000). This approach is the one closest to the method created in this thesis, MUHBOD, in the sense that target variables are split for each interval of the source variable. However, in MUHBOD source-target roles are not mutual and instead work in one-way and in a hierarchical[6] and recursive manner and a simple equal frequency method satisfies to create the splits.

Various disposition effect studies involving Cox model incorporate numeric features discretized into categorical variables mostly through the equal width method, including Barber and Odean (2013), Brettschneider and Burgess (2017), Chhabra and De (2012), Chiyachantana and Yang (2013), Choe and Eom (2009), da Silva and Mendes (2021), da Silva, Mendes, and Abreu (2022), De, Gondhi, and Sarkar (2011), Gemayel and Preda (2018), Loos, Meyer, Weber, and Hackethal (2014), Muhl and Talpsepp (2018), Seru et al. (2010), Shumway and Wu (2005), Stoffman (2008), Strahilevitz et al. (2011), Talpsepp (2010, 2011), Talpsepp and Vaarmets (2019), Vaarmets, Liivamägi, and Talpsepp (2019), Wheeler Huttunen (2016), and Zimmermann (2015).

[6] Hierarchical in the sense that variables lower in the hierarchy are discretized along each combination of the categories of variables higher in the hierarchy. This is not to be confused with hierarchical methods which involve splitting and merging intervals across a single variable.

CHAPTER 3

HYPOTHESES DEVELOPMENT

Extant literature on disposition effect as summarized in Section 2.2 (p. 11) and Subsection 2.4.2 (p. 45) frequently mentions five recurring topics:

- A V-shaped or non-monotonic relationship between propensity to sell shares and explanatory variables, especially capital gains/losses overhang with respect to reference prices
- Commonly observed reverse disposition effect, which is basically a higher disposition to sell losing positions than that to sell winning positions
- Multiple alternative reference points other than the purchase price, and their adaptation over time
- Effect of the salience of reference prices on the prevalence of disposition effect
- Inclusion of discretized variables into the model

In studies focusing on V-shape disposition effect findings, emphasis is put on explaining the absolute magnitude of the effect apart from the disposition direction. While few studies report evidence on simultaneous multiple effects acting on opposite directions to create the observed V-shape, this area is largely unexplored.

In some studies reporting reverse disposition effect, the direction of the disposition effect changes based on underlying conditions, especially market states. However, perception of the relative gain or loss in different conditions is not incorporated sufficiently. In case these conditions are incorporated, they might result in a consistent relationship.

In studies focusing on reference points, highest or lowest prices experienced are frequently proposed and included as alternative reference points. However, they are usually included as record prices in a moving window. So in many cases record prices that are reached before positions are opened and as a result that are not experienced by investors may be included in the analysis. Only a few studies take the rigorous challenge to calculate the record prices that each investor experiences in their stock position separately. While a multitude of possible reference prices are mentioned frequently, only few studies include additional reference prices other than the purchase price simultaneously in their models, and the interactions among the reference prices are not considered well enough.

The fact that enhanced salience of the reference points accentuates the disposition effect is well studied. However, in many studies, features of the data, such as multiple purchases and sales or splits or dividends that may veil the salience of the reference point are not sufficiently controlled for.

Continuous features are frequently discretized in order to model non-linear or complex relationships. However, in this case, the significance of the coefficient estimate of a linear relationship is not sufficient to induce observed effects. Rather, the monotonicity of the estimates across categories gets important. The monotonicity is usually probed with visual interpretation of the data, and a robust and systematic test for the monotonicity is mostly lacking. The interactions among categorical features are only done for few dichotomous variables. Interacting multiple categorical variables can inflate the number of features, complicating the computational efforts and the interpretation of the high number of coefficients. A very large model requires special high performance tools to handle the data and model within a reasonable time frame. For these reasons, such an effort in disposition effect studies is hard to implement, and is not encountered in the background research for this thesis as to the author's knowledge, despite a highly exhaustive literature review. Furthermore, while there are univariate methods for testing monotonicity, and ordinal regression methods to explain ordinal responses with numeric covariates exist, there is yet no commonly accepted and ubiquitous method to test the monotonicity of numeric responses across multiple ordinal covariates simultaneously.

This thesis attempts to fulfill the aforementioned gaps in disposition effect literature:

- Data are filtered for a sample of positions in which information on purchase price or other reference points are not blurred or distorted but kept salient, to better isolate the disposition or reverse disposition effect. For this reason only simple positions with a single purchase[7] and either with no sales (right censored data) or a single sale are included. Only the positions with the stocks that had no stock splits or dividends in the year the data were collected are eligible for the study
- Multiple reference points, namely purchase price, maximum price and minimum price are included in the study as separate variables or their triple-way interactions. A rigorous data wrangling algorithm is developed in order to precisely calculate the reference points for each position of an investor in a stock at the seconds level.
- All capital gains/losses overhang measures are discretized into ordered categorical (ordinal) variables. Conflicting or inconsistent findings on the direction of the disposition effect are probably due to the omission of the contextual variables that

[7] Multiple purchases/sales at the same split second and same price are also considered as a single purchase/sale.

may affect the perceived level of the covariates. In order to overcome this, discretization is done so that the investment horizon and market states are taken into consideration in a highly precise manner for all three reference points. A novel discretization approach is developed and introduced here.

- The model specification including three-way interactions of discretized variables involves up to 998 dummy variables to be included in the model. A high performance distributed statistical framework is employed in order to run Cox regression with a large dataset and a complex model.
- The interpretation and multivariate monotonicity testing of high number of coefficients of triple-way interactions among discretized variables are major challenges in this thesis. In a second step, coefficients are fed into a Bayesian model with monotonic effects in order to test whether propensity to sell is monotone along overhang levels with respect to multiple reference points simultaneously. Secondary effects of these monotonic relationships are also tested using the same model with interactions.

So the main hypotheses explored in this thesis are as follows:

$H1_0$: There is no significant relationship between the propensity to sell a position and the capital gains/losses overhang with respect to three reference points separately.

$H1_A$: There is a significant relationship between the propensity to sell a position and the capital gains/losses overhang with respect to three reference points separately.

$H2_0$: There is no significant univariate monotonic relationship between the propensity to sell a position and the capital gains/losses overhang with respect to three reference points separately.

$H2_A$: There is a significant univariate monotonic relationship between the propensity to sell a position and the capital gains/losses overhang with respect to three reference points separately.

$H3_0$: There is no significant relationship between the propensity to sell a position and the capital gains/losses overhang with respect to three reference points included as triple-way interactions.

$H3_A$: There is a significant relationship between the propensity to sell a position and the capital gains/losses overhang with respect to three reference points included as triple-way interactions.

$H4_0$: There is no significant multivariate monotonic relationship between the propensity to sell a position and the capital gains/losses overhang with respect to three reference points included as triple-way interactions.

$H4_A$: There is a significant multivariate monotonic relationship between the propensity to sell a position and the capital gains/losses overhang with respect to three reference points included as triple-way interactions.

$H5_0$: There is no significant interaction between the monotonic relationships found in Hypothesis 4.

$H5_A$: There is a significant interaction between the monotonic relationships found in Hypothesis 4.

$H6_0$: Investors do not incorporate holding period and market state information between the time a reference price is formed or updated and the current time in their decisions.

$H6_A$: Investors incorporate holding period and market state information between the time a reference price is formed or updated and the current time in their decisions.

Upon the rejection of all null hypotheses:

- A monotonically increasing (isotonic) relationship between the propensity to sell and overhang measures is interpreted as the existence of disposition effect (monotonically increasing propensity to sell from extreme losses to extreme gains).
- A monotonically decreasing (antitonic) relationship is interpreted as the existence of reverse disposition effect (monotonically decreasing propensity to sell from extreme losses to extreme gains).

And also, the alternative hypothesis for H6 posits that, apart from the level of the difference between the current price and the reference price, namely the overhang, investors' perception of gain and loss is context-sensitive in the sense that holding period and market state between the time of the formation or update of a reference price and the current time determine how the overhang is perceived.

The monotonicity for the relationships with respect to each reference point (purchase, maximum or minimum) is not guaranteed to be in the same direction. Existence of simultaneous relationships in opposite directions can provide an explanation to the commonly observed V-shaped pattern.

In case significant simultaneous monotonic relationships are found across different reference points, the multidimensional emotions argument as reviewed in Subsection 2.3.6 (p. 39) can be used as a possible explanation in the discussion. Shortly, when making decisions, each reference point may evoke a different emotion, such as regret, elation, fear and greed or hope, in individuals and these emotions may appear simultaneously and in interaction. These interactions may amplify or understate the effect of different reference points and their respective emotions so that the net effect may not be understood easily without taking this complexity into consideration.

CHAPTER 4

SOFTWARE TOOLS USED

Due to dealing with investor-transaction level stock market data, a major challenge of this thesis is the big data aspect:

- Although transactions are filtered according to the rules explained in detail in Chapter 5 (p. 61), the panel format for survival analysis requires that for each single transaction a separate row is created for every day until either a position is totally closed or last date of data is reached (in which case the transaction is deemed as censored in survival analysis). So the number of rows is inflated especially for long holding periods or positions that are not closed within the time frame of the data.
- Overhangs with respect to multiple reference points are converted to ordinal variables of three, five, or ten ordered categories each. In some models, triple-way interactions of all those ordered categories are included as separate covariates, a situation which widens the model matrix of the data up to 998 columns.

Since Cox model works on an iterative procedure maximizing likelihood and Bayesian models for monotonic effects is based on iterative sampling of parameters and respective recalculations on the data, in-memory storage and computing requirements of the models increase.

The tools utilized in this thesis are selected in order to maximize computational performance and reproducibility, and they are mostly free software and open source projects and tools.

All data importing, wrangling, transformation, modeling, reporting and visualizations are made on a PC running Arch Linux distribution (“Arch Linux”, n.d.) of GNU/Linux operating system (“Linux”, n.d.). GNU/Linux is a part of GNU project (“GNU”, n.d.), a collection of free software led by Richard Stallman.

The main computational and programming environment is R programming language and environment for statistical computing and graphics (R Core Team, 2021). R is a project of GNU.

Data importing from csv and xls formats are done using data.table (Dowle et al., 2021) and XLConnect (Studer, GmbH, Foundation, Builder, & Woolridge, 2021) R packages.

Data wrangling, transformation and visualizations are done using data.table (Dowle et al., 2021) R package, tidyverse suite of R packages (Wickham & RStudio, 2021) and ggpubr and ggh4x R packages, which are extensions to tidyverse (Kassambara, 2022; van den Brand, 2022).

Tables are created using kable function from knitr (Xie, Sarma, et al., 2022) R package, and they are enhanced with kableExtra (Zhu et al., 2021) R package.

Formatting of in-text numbers and percents are done by scales (Wickham, Seidel, & RStudio, 2022) R package. Integers are automatically converted into words by english (Fox, Venables, & Salverda, 2021) R package.

Fast and parallel serialization and deserialization of large R data objects between major subsequent steps are done using fst (Klik, 2020) and qs (Ching et al., 2021) R packages.

Basic R package for survival analysis and Cox regression is survival (Therneau, Lumley, Elizabeth, & Cynthia, 2021) and for visualization of survival curves is survminer (Kassambara, Kosinski, Biecek, & Fabian, 2021). While these packages are used in the study to calculate and visualize Kaplan-Meier (KM) survival curves, the computation times of Cox models on survival package multiply exponentially as the number of covariates increases. To benefit from more efficient optimization algorithms for maximum likelihood estimation and leveraging the power of multiple cores for parallel execution, h2o R package (LeDell et al., 2021) is used as the main tool for running Cox regressions. h2o package is an R interface to highly performant H2O.ai machine learning library ("H2O.Ai", n.d.).

To test for the monotonicity of coefficients from Cox models and ratios of Odean's measure, brms R package (Bürkner et al., 2021) is used. brms is an R interface to Bayesian computing environment Stan ("Stan", n.d.). CmdStanR package ("CmdStanR", 2021) is also used as a backend for parallel execution of chains.

A total of 13 models are created and tested using primary (Cox regression, KM curves and Odean's measure) and secondary (Kendall's Tau and Bayesian regression model with monotonic effects) methods. The models are created in a declarative and tabular manner on Google Sheets, and the created model metadata is accessed from R through Google Sheets API using googlesheets4 R package (Bryan & RStudio, 2021).

The models, declarative metadata of which is created on Google Sheets and imported into R are executed in batch mode and the results are collected in large list objects and are serialized. rlist (Ren, 2021) and pipeR (Ren, 2016) R packages are utilized for manipulating highly nested list objects holding model results.

All code is written inside reproducible Jupyter notebooks on JupyterLab environment ("Project Jupyter", n.d.), and the notebooks are connected to R through IRkernel R package (Kluyver, Angerer, Schulz, & Ram, 2021). Whole analytical workflow is easily reproduced by re-running Jupyter notebooks and data provenance from the source files to the final datasets can easily be traced. The code versioning is done using Git software ("Git", n.d.), and the code is also pushed to a private repository on Github ("GitHub", n.d.).

The authoring of this thesis is done in Rstudio integrated development environment ("RStudio", n.d.), and it is mostly based on a template by thesisdown (Ismay, 2022) R package, which is built upon bookdown (Xie, Allaire, et al., 2022) R package. bookdown itself is powered by rmarkdown (Allaire et al., 2022) R package, knitr (Xie, Sarma, et al., 2022) R package and pandoc ("Pandoc", n.d.), a document conversion system. The material is first converted into LaTeX ("LaTeX", n.d.), a document preparation software based on typesetting system TeX, and then it is converted to portable document format. The LaTeX engine used is xelatex, the LaTeX compiler of XeTeX project, a unicode based version of TeX ("XeTeX", n.d.).

All cited references are stored in Zotero reference management system ("Zotero", n.d.), and they are automatically exported into BibLaTeX ("BibLaTeX", n.d.) format to be imported and used by authoring packages in R. Collected information on cited references are stored in a SQLite ("SQLite", n.d.) database, and the database connection between R is done through RSQLite (Müller et al., 2022) R package. Connection between R and Zotero is done through zoteror (Comai, 2017, January 24/2022), RefManageR (McLean, 2022) and citr (Aust, Xie, Lovelace, & Aden-Buie, 2019) R packages. jsonlite (Ooms, Lang, & Hilaiel, 2022) and httr (Wickham & RStudio, 2022) R packages are also utilized to manipulate Zotero collection data through its web API.

CHAPTER 5

DATA

5.1 Data definition and sources

Data used in the study consist of four sources:

- Transaction data: All stock transactions made on BIST[8] between 1 February 2005 and 30 December 2005
- Capital increases: Capital increase dates, amounts and types for stocks trading on BIST
- Dividend payments: Dividend payment dates, amounts and types for stocks trading on BIST
- BIST Index values: BIST price index values calculated once in every ten seconds

5.1.1 Transaction data

Disposition effect is a topic of behavioral finance and requires disaggregated investor level data on stock transactions. Investor level transaction data are accumulated, stored and served by mainly three types of institutions:

- Istanbul Clearing, Settlement and Custody Bank Inc. (*Takasbank*)
- Central Securities Depository (*Merkezi Kayıt Kuruluşu*)
- Investment houses

Since each investment house holds only a portion of the total investor universe, the data may not be complete and in case the institution targets only certain investment segments, the data might be biased from the onset.

[8] BIST was named Istanbul Stock Exchange (ISE) as at the date of data.

The detailed data on stock transactions made between 1 February 2005 and 30 December 2005 on BIST are exhaustive as they include all data made by all investors and on all stocks. The data were collected by Takasbank and were previously used in Ozsoylev, Walden, Yavuz, and Bildik (2014).

The fields in the original dataset are as follows:

- tarih: Date of transaction
- zaman: Time of transaction
- m_kod: Ticker code of stock
- fiyat: Unit price of transaction
- adet: Quantity of transaction
- alan_uye: The institutional code of the buyer
- alan_hesap_no: Account no of the buyer
- satan_uye: Institutional code of the seller
- satan_hesap_no: Account no of the buyer

The generating SQL code of the dataset which is included as the first row of each source file in csv format reveals that a *where* clause was used for filtering only E value for the p_kod field. E stands for old (*eski*) stocks while the other value for the field is Y that stands for new (*yeni*) stocks. The default code is E. The bisection of the stock of a company into E and new Y occurs when a capital increase is made after the beginning of the new year but before the distribution of the dividends on previous year's profits. While the old stocks are eligible to receive dividends from previous year's profits, the new stocks are not. For this reason old type shares of the same stock are expected to trade at a premium over new type shares. After the ex-date of dividends, both codes merge into old type shares. The dataset includes trades for only old type shares.

The raw dataset has a total of 39,749,152 transactions made in 235 transaction dates between 1 February 2005 and 30 December 2005 on 313 separate stock codes. Daily transaction number is between 67,210 and 292,125 with an average of 169,145.

Daily transaction numbers across the dates are plotted in Figure 1.

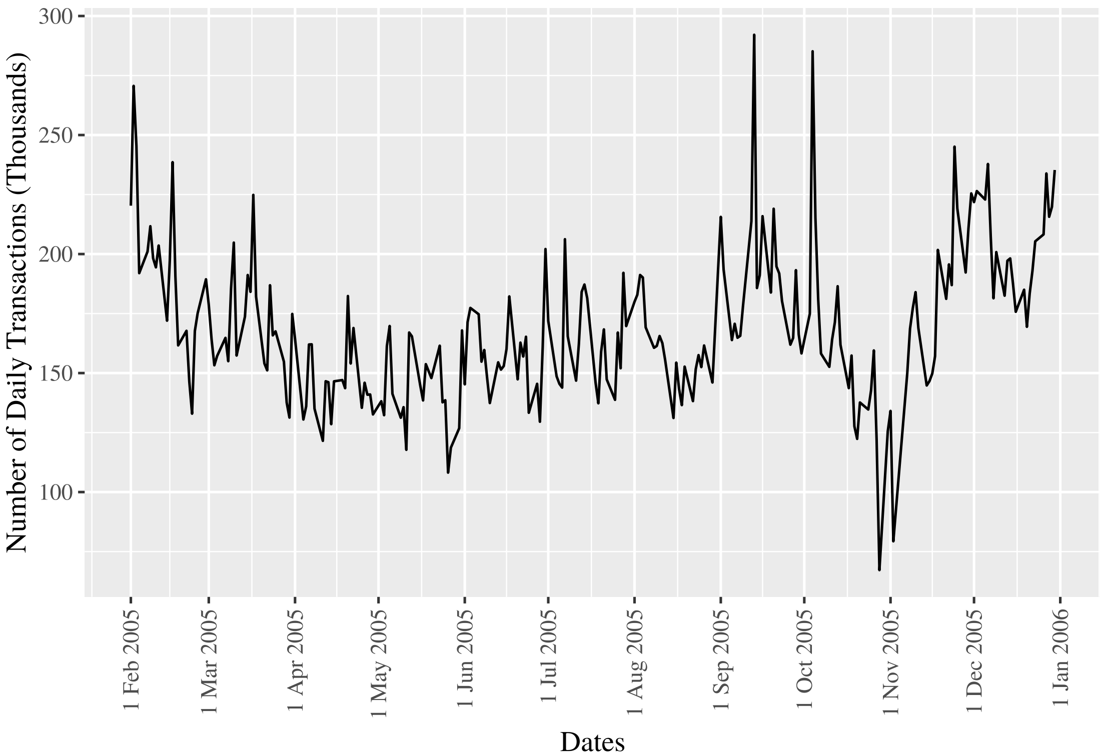

Figure 1. Number of daily transactions by date

The histogram of daily transaction numbers are plotted in Figure 2.

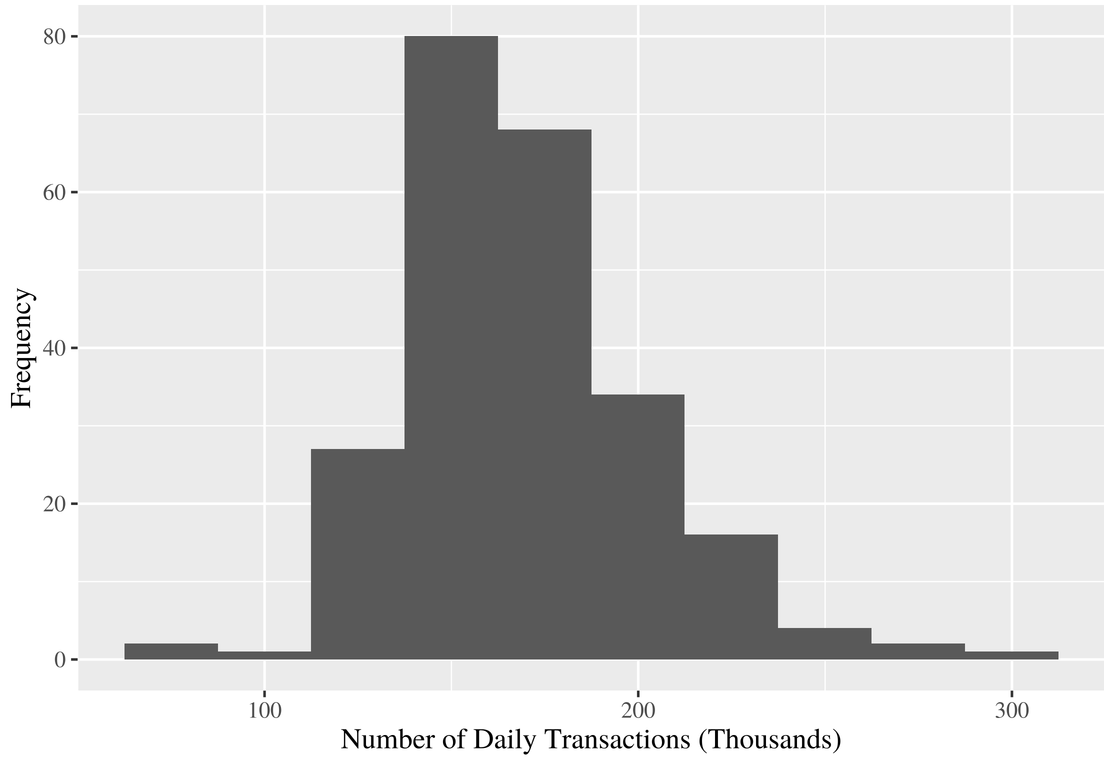

Figure 2. Histogram of daily transaction numbers

Daily total transaction value is between TL 371 million and TL 2,817 million with an average of TL 1,029 million. Daily total transaction values across the dates are plotted in Figure 3. The histogram of daily total transaction values is plotted in Figure 4.

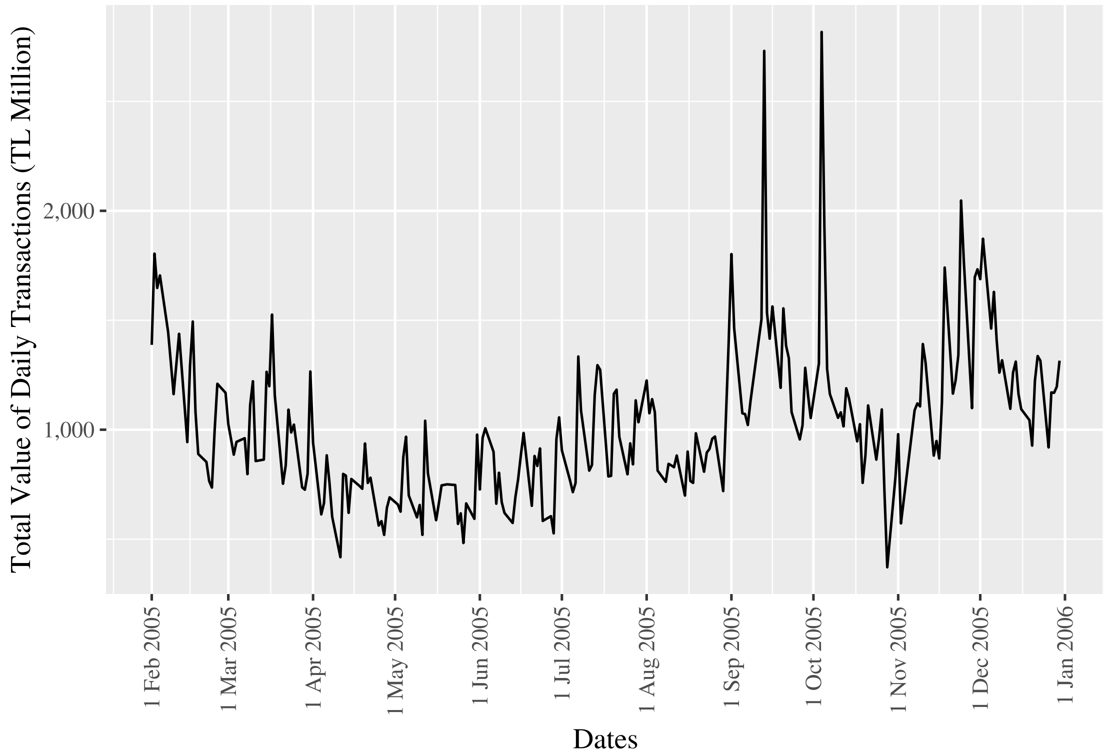

Figure 3. Total transaction values by date

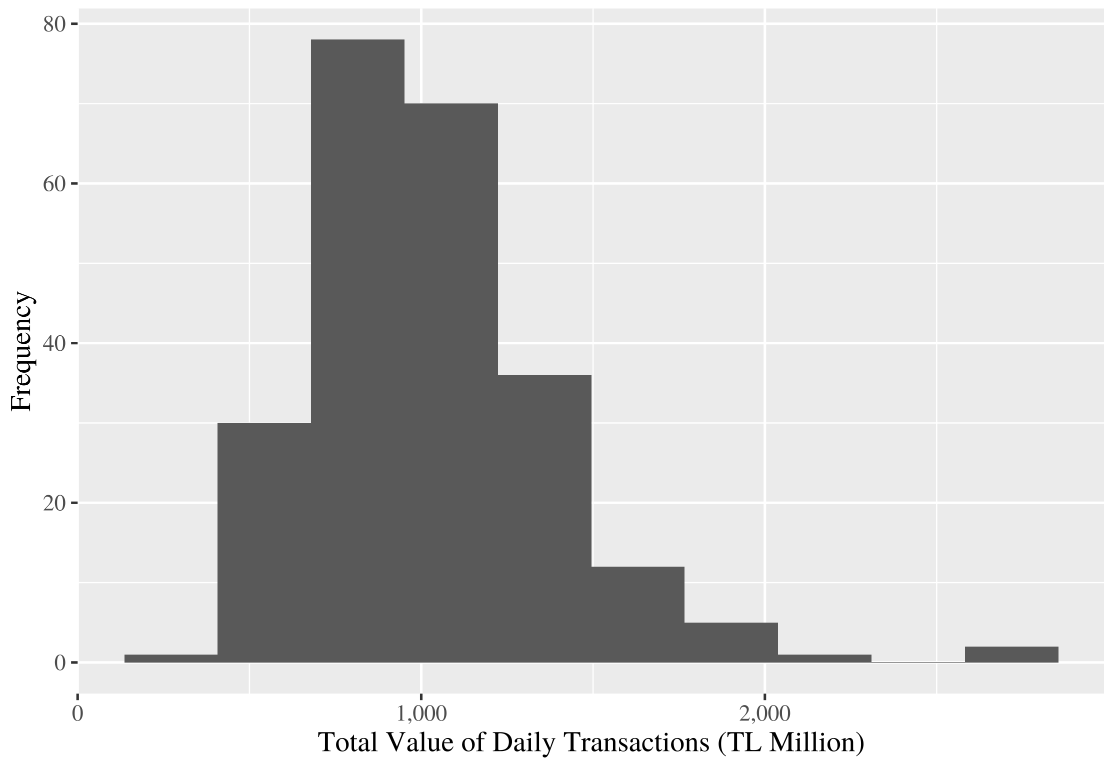

Figure 4. Histogram of daily total transaction values

5.1.2 Capital increases and dividends

Capital increase and dividend datasets are compiled by BIST to summarize information on capital increase and dividend dates, types, ratios and amounts. These datasets are utilized here in order to filter out those stock codes that have gone through any of these actions during the range of transaction dates in the main dataset.

The capital increase dataset covers years between 1986 and 2009 and has information across 377 stock codes. The dividend dataset covers years between 1985 and 2009 and has information across 378 stock codes.

5.1.3 BIST index values

Historic values of BIST indices by ten second intervals are provided by BIST. While BIST TUM[9] is a broader index including all companies regardless of the size, BIST 100 index is the most salient one that is tracked by the investors as the measure of market state and movement.

The market indices on BIST are calculated on two bases: Price indices do not take into account the cumulative effect of capital increases and dividend payments on the values, while return indices do incorporate this information. So return indices are better proxies of the wealth change of investors from market movements. However, return indices are usually not salient on bulletins and media and can only be accessed through statistics. Investors track the market states through price indices. So although BIST TUM return index is the best proxy for market values on BIST, BIST 100 price index is selected as the market proxy since investors' market judgments are based on this index.

[9] While BIST TUM and BIST 100 are the better known names of the corresponding indices, the internal codes used by BIST are XU TUM and XU 100 respectively. They are used interchangeably in this thesis.

Index value is between TL 22,888 and TL 40,012 with an average of TL 29,629 within the date range of the transaction data. The candlestick chart of the open, low, high and close values of the BIST 100 price index is provided in Figure 5.

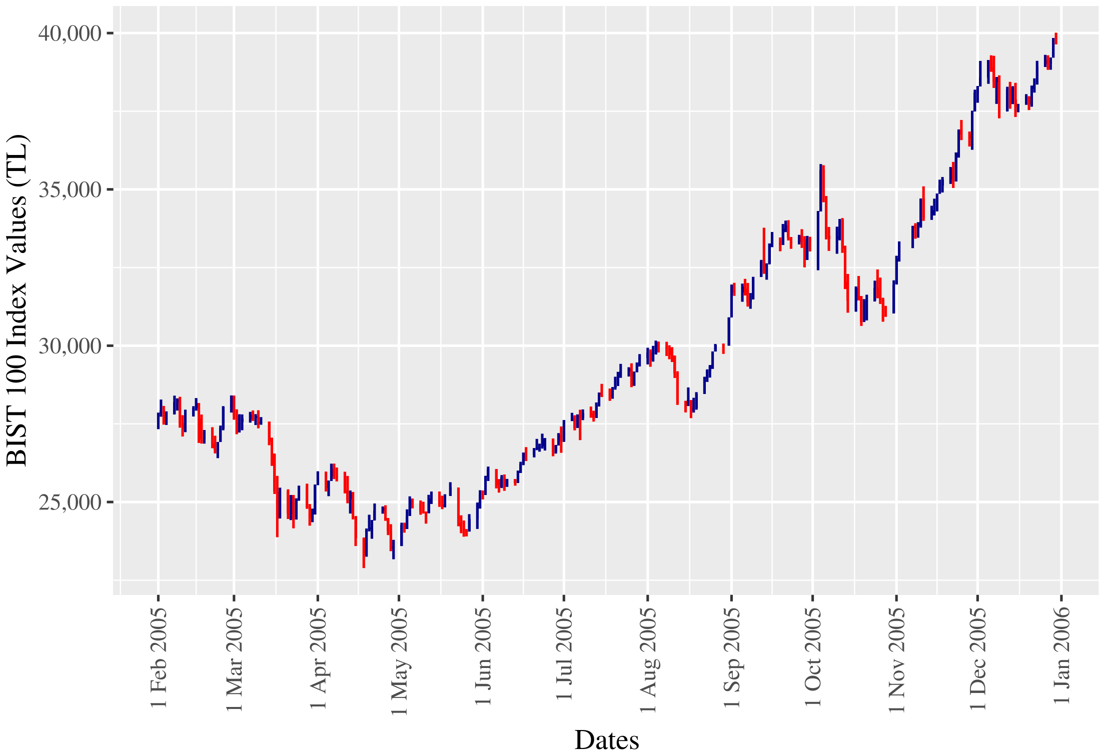

Figure 5. Candlestick chart of daily BIST 100 price index values (TL)

While BIST was mostly bullish within the date range, there were also periods of decline throughout the period. The market return for each investor-stock position and for each reference price is evaluated separately and on a daily basis. Out of the 27,495 start and end date combinations within the date range, the market had a positive return on closing values in 22,957 date combinations, while it had a negative return in 4,538 ones.

5.2 Summary of data wrangling steps

The raw datasets explained in Section 5.1 (p. 61) have to go through several rigorous steps to be suitable for subsequent modeling. Major types of actions are as follows:

- ID indexing across several variables for different purposes
- Summarization of numeric variables across time or categorical variables
- Exact or rolling joins among multiple datasets
- Filtering on multiple complex criteria
- Expansion of a single stock position across all dates within the holding period
- Calculation of daily high-low-close-average prices for each stock ticker
- Cost reference calculation
- Maximum and minimum reference price calculation
- Calculation of index values corresponding to reference prices
- Calculation of multiple overhang[10] measures
- Ordinal discretization of numeric variables in a hierarchical and multivariate manner

5.3 Splitting transactions

The preliminary wrangling steps for the transaction dataset are as follows:

- Creating an index for the transactions order
- Roll joining by intraday BIST 100 index values
- Splitting the buy and sell side of each transaction into separate records

The timing of the transactions are given by seconds accuracy along with date. Any subsequent chronological sorting using the data itself can be done using date, time

[10] "Overhang" is used interchangeably for "capital gains/losses overhang".

and stock ticker code fields. However, even in the seconds accuracy there may be multiple transactions per second. As can be seen in Table 1, the cases where a single transaction occur per second per stock code make up only 63% of all unique second-stock code combinations present in the data and only 31% of all transactions. So any later sorting of the dataset using existing columns will lose its original order. For this reason a unique ID column named trans_no1 is added, and a unique sequential number is given to each transaction.

Intraday BIST 100 index values in ten second intervals are left joined to the transaction dataset by date and time. Last observation carried forward method is used, so that the last record at or before the exact time is used as the corresponding index value.

All records in the dataset are copied into buy and sell side objects so that the date, time, stock code, quantity, price and transaction id fields are common. While buyer side member and account fields are used in the buy side data, sell side member and account fields are used in the sell side data. A member means an investment institution that holds the account for the investor. Four new fields are added:

- buy_sell: B and S for buy or sell transaction
- buy: Boolean field. TRUE for all buy, FALSE for all sell side transactions
- sell: Boolean field. FALSE for all buy, TRUE for all sell side transactions
- trans_no2: For each unique transaction ID in the raw dataset, two new transactions IDs are created for the split dataset by doubling the indices and decrementing the index for buy side of the original transaction, so that the ID of the buy comes first.

The two datasets are combined into a single dataset with a total row count of 79,498,304.

Table 1. Cumulative Percentage of Total Seconds and Transaction Count by Number of Transactions per Second

Number of Transactions per Second	Seconds Count	Cumulative Percent in Seconds	Cumulative Percent in Transactions
1	12,314,106	62.93	30.98
2	3,838,743	82.54	50.29
3	1,271,521	89.04	59.89
4	761,788	92.93	67.56
5	395,193	94.95	72.53
10	62,744	98.51	85.32
50	577	99.94	97.72
100	56	99.99	99.33
501	1	100.00	100.00

While each record shows one side of a unique transaction, the main entity of analysis is *position*. A position is a collection of transactions that a certain investor makes in a certain stock code. So the reference prices, overhang measures and sell status for each day a transaction is open will be calculated across positions.

To track the transactions across positions, three additional unique and sequential ID columns are added after the merge:

- investor: Unique investor ID for each member and account combination
- pos_no2: Unique position ID for each investor and stock combination
- pos_transaction_no: After the data are sorted for pos_no2 first and trans_no2 later, a unique sequential ID is given for each transaction.

A single real person or legal entity can have multiple accounts across different member institutions. No identifier for the person or entity is present in the data. Furthermore, mental accounting concept posits that investors track segregated accounts for the overhang (unrealized gains or losses) of their investments without aggregating the gains and losses. Treating each account separately is in line with mental accounting.

There are 552,389 unique investor accounts held by 101 members. These investors opened a total of 4,761,393 positions between 1 February 2005 and 30 December 2005. Beginning balances on 1 February 2005 are not known. Any previous transactions for the positions are outside the scope of the data. So there is no information to left truncate those positions that start before the beginning date. Furthermore, there might be positions that are opened before the beginning date and do not have any trading activity during the period of the data. So it is assumed that a position is opened at the time an investor makes the first transaction for a stock within the data period. An investor can completely close their balance in a stock several times and reopen again with subsequent transactions. Since only simple positions with a single buy transaction and at most one sell transaction will be included for further analysis, this kind of complication is not addressed and all transactions of an investor in a stock are considered to be made inside a single position.

5.4 Creating daily and intraday historic prices

For each position, for each day the position is open, the overhang will be reevaluated by comparing the current price with three reference bases:

- Cost price
- Maximum price encountered so far after the position is opened
- Minimum price encountered so far after the position is opened

The return on the BIST 100 index as experienced in each position and day combination with respect to the indices corresponding to the exact time the references are formed or updated should also be calculated separately.

For each position, trading days are the days for which a position is traded with either a buy or sell transaction, and non-trading days are those days without any buy or sell transaction by the investor. Trading and non-trading days should also be treated differently for calculations. Non-trading days can be treated the same across all positions, for calculating the maximum and minimum prices and corresponding market indices for overhang measures and reference price updating. Since the ask price is always above the bid price and the most likely instant an investor considers closing a position is the point where the price is at a maximum, the daily maximum price of a stock is taken as the current price basis.

Daily minimum and maximum prices are also used for updating the cumulative minimum and maximum prices for each position on a non-trading day. The daily minimum and maximum prices for each stock ticker are calculated from the whole transaction data. An advantage of not using a ready dataset for daily prices but utilizing the intraday nature of the transaction dataset lies in its ability to combine with BIST 100 values. The timing of intraday extreme values of indices and stock prices are not guaranteed to be at the same instant. Since the overhang is calculated by comparing the current price and a reference price (cost, maximum or minimum), the index values corresponding to those extreme prices should also be updated respectively. So the maximum or minimum index values for a stock-day combination are the index values that correspond to the exact time that stock price reached a daily minimum or maximum.

Stock-day level minimum and maximum prices are joined with the original transaction dataset. So the index values corresponding to the first instance a minimum and maximum price is encountered in each day is taken as the maximum and minimum index values for the stock-day combination. If the same minimum or maximum price is encountered again, it is assumed that the existing reference price is not updated, so the index value is not updated either. The idea behind the process can be illustrated in Figure 6.

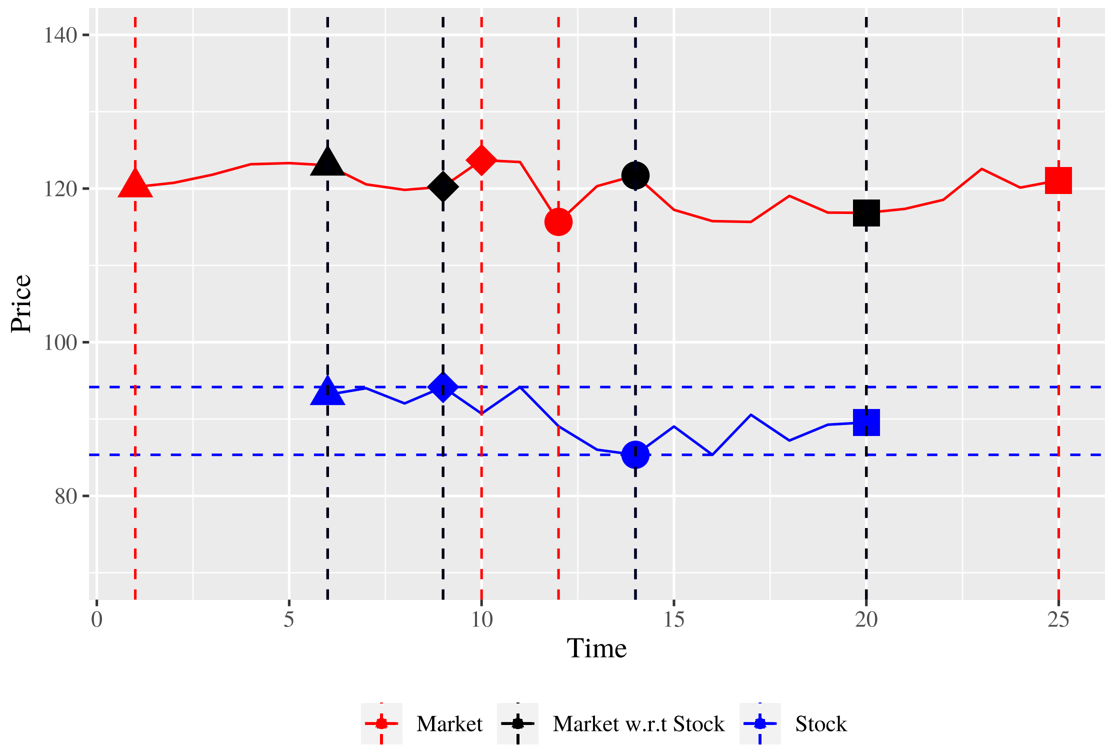

Figure 6. Determination of market index values corresponding to stock price references

Blue series represent hypothetical stock prices for a position, while red series represent hypothetical market index values, both during the course of a unit period, e.g. a day. Triangles mark the initial values of the series, squares mark the ending values, circles mark the minimum values and diamonds mark the maximum values. The markers are also colored blue and red for the stock position and market index, respectively. The stock position is opened after the market is opened, and it is closed before the market is closed[11]. The market index values corresponding to the times of starting, minimum, maximum and ending values of the stock position are marked black. The black diamond mark, representing the index value corresponding to the time the maximum of the stock position is first reached, is before and lower than the red diamond mark, representing the maximum index value of the day. Similarly, the black circle mark, representing the index value corresponding to the time the minimum of the stock position is first reached, is after and higher than the red circle mark, representing the minimum index value of the day. Furthermore, in this hypothetical example, the market index value with respect to the minimum stock price of position, marked by the black circle, is higher than the market index value with respect to the maximum stock price of position, marked by the black diamond.

The blue dashed horizontal lines show the maximum and minimum prices of the stock position. Both the maximum and minimum prices are encountered twice during the day. However, the corresponding index values are not updated unless a new minimum or maximum stock price is experienced.

A dataset labeled as historic data is created from these calculations, holding stock-day level minimum and maximum prices and corresponding minimum and maximum BIST 100 index values for each stock-day. The dataset has 71,614 rows.

The harder case is for addressing the minimum and maximum stock prices and their corresponding index values for trading days of each position. The daily minimum or maximum prices are not guaranteed to be applicable to all positions as reference prices. A daily minimum or maximum price reached before a buy transaction is made cannot be a reference price for the corresponding position. So for

[11] For illustration purposes, a hypothetical intraday position is created in this example. However, in the actual dataset, intraday positions are filtered out.

a day a position is opened with a buy transaction, the maximum and minimum prices until the end of the day are the ones that occur at the time or after that position is opened. The price trajectories for each date, second and stock are extracted from all transactions in the dataset. This idea is illustrated in Figure 7.

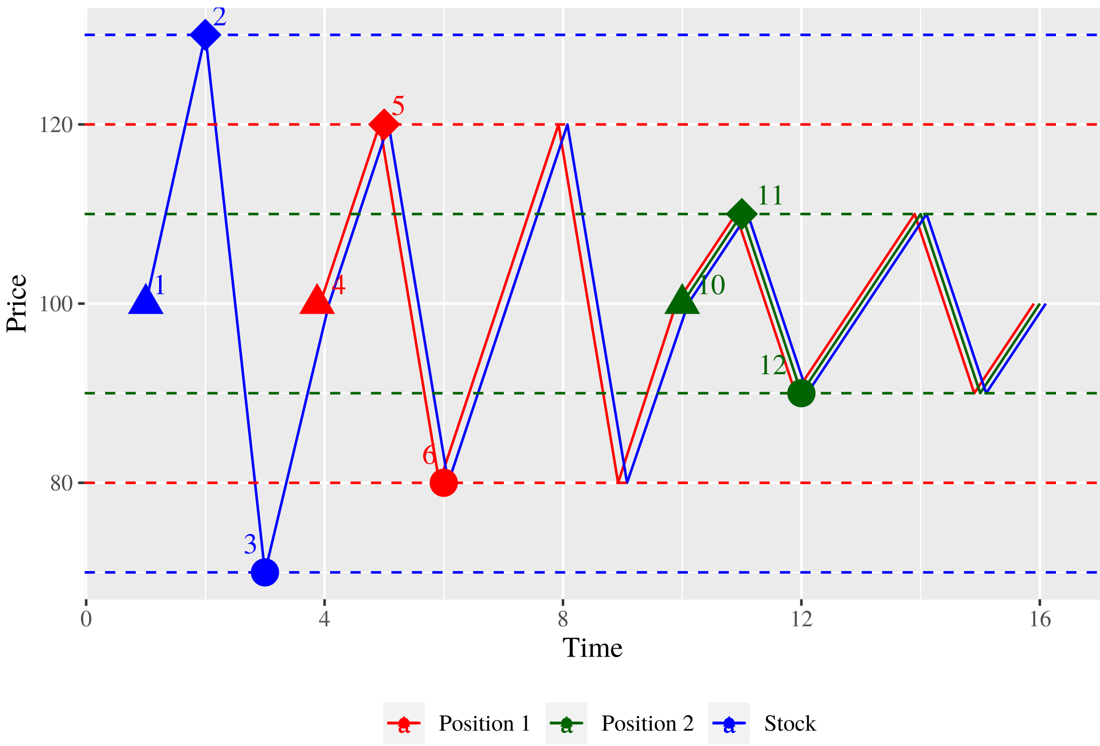

Figure 7. Determination of maximum and minimum references on the opening day of a position

The blue series represent the prices of a stock throughout a day, red series represent the prices for a position opened at time 4 and green series represent the prices for a position opened at time 10. Since the positions are opened for the same stock, the price trajectories overlap. The triangles mark the initial prices for the stock and the positions, while the circles mark the minimum prices and diamonds mark the maximum prices. The daily maximum and minimum prices for the stock occur at times 2 and 3 respectively. However, these extreme levels are experienced by neither of the two positions that are opened at a later time. So these prices cannot serve as references for those investors. For Position 1, maximum and minimum prices are experienced at times 5 and 6 respectively. The maximum of Position 1 is lower than the daily maximum and the minimum of the position is higher than the daily minimum. The extreme prices experienced by the investor of Position 1 are not experienced by the investor of Position 2. For Position 2, maximum and minimum prices are experienced at times 11 and 12 respectively. The maximum of Position 2 is lower than both the daily maximum and Position 1's maximum. Similarly, the minimum of the position is higher than both the daily minimum and Position 1's minimum.

For the purposes of this calculation, for each stock, date and second level time combination, backward cumulative maximum and minimum price is calculated by reversing the order of transactions temporarily and tracking the cumulative extreme prices. When the order is restored, the intraday maximum (minimum) is non-increasing (non-decreasing) with respect to time in seconds. So for the time ranges until the end of a day, the ones that start at a later time cannot have a maximum (minimum) price within that time range higher (lower) than maximum (minimum) price within time ranges starting at previous times on the same day. For each time in seconds, the corresponding index values for all instances an intraday maximum or minimum is equal to the current price is also joined. This is different from the behavior in forward cumulative maximum and minimum price calculations. Different investors opening a position at different times during a day can encounter the same backward looking maximum/minimum price levels visited multiple times during a day. But the corresponding index value should be the one that is applicable when the maximum/minimum is encountered by the investor at the exact time. This subtle detail is illustrated in Figure 8.

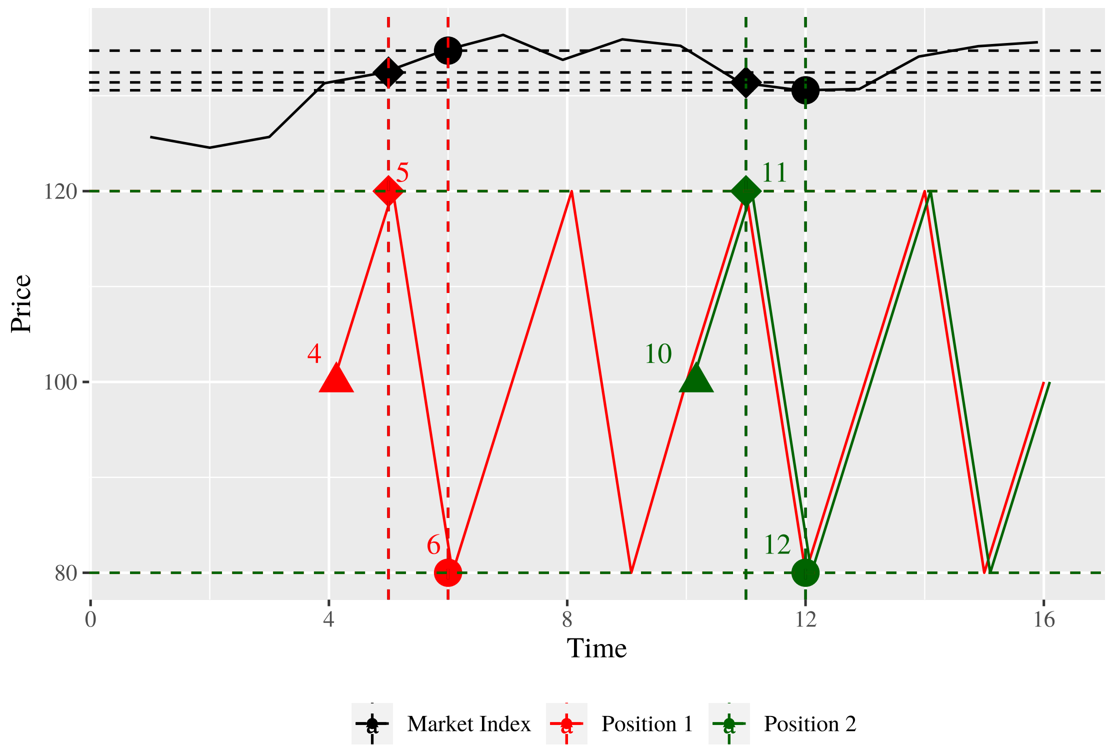

Figure 8. Determination of corresponding market indices for maximum and minimum references on the opening day of a position

Position 1 (red) is opened before Position 2 (green) on the same day. Triangles mark the beginning, circles mark the minimum and diamonds mark the maximum prices for each position and the corresponding market index. By observing the red and green circles and diamonds and the intersecting horizontal lines, it is apparent that experienced maximum and minimum prices of two positions have the same value, albeit they are encountered first at different times. However, the corresponding market indices for the first times maximum and minimum prices are encountered by each investor are different, as can be observed from the black circles and diamonds and the intersecting horizontal lines. Furthermore, each investor encounters these maximum and minimum prices multiple times. However, the corresponding market index is determined with respect to the first encounter of each investor and is not updated unless a new maximum or minimum price is experienced by the investor.

The backward cumulative maximum and minimum prices and corresponding index values are used for calculating the minimum and maximum reference prices on the day a position is opened with a buy transaction.

Similarly, a daily minimum and maximum may occur after a position is closed with a sell transaction. In that case, those daily extreme prices cannot serve as reference prices for that position. For the day a position is closed, the actual selling price is taken as the current price. The maximum or minimum reference price should be updated if a new extreme is reached until the time in seconds the sell transaction is realized. So the forward cumulative maximum and minimum prices for each stock, date and time in seconds combination is calculated. This idea is illustrated in Figure 9.

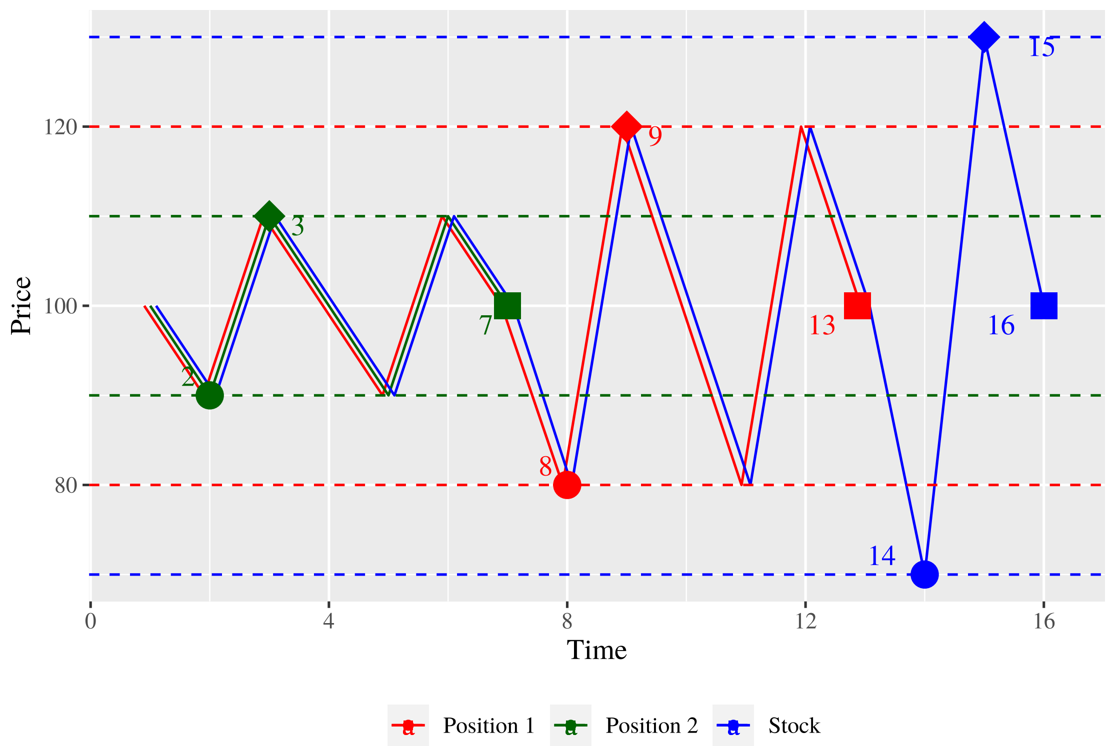

Figure 9. Determination of maximum and minimum references on the closing day of a position

The blue series represent the prices of a stock throughout a day, red series represent the prices for a position closed at time 13 and green series represent the prices for a position closed at time 7. Since the positions are for the same stock, the price trajectories overlap. The squares mark the ending prices for the stock and the positions, while the circles mark the minimum prices and diamonds mark the maximum prices. The daily minimum and maximum prices for the stock occur towards the end of the day at times 14 and 15 respectively. However, these extreme levels are experienced by neither of the two positions that are closed at an earlier time. So they cannot serve as reference prices for those investors. For Position 1, minimum and maximum prices are experienced at times 8 and 9 respectively. The maximum of Position 1 is lower than the daily maximum and the minimum of the position is higher than the daily minimum. The extreme prices experienced by the investor of Position 1 are not experienced by the investor of Position 2, which is closed at an earlier time. For Position 2, minimum and maximum prices are experienced at times 2 and 3 respectively. The maximum of Position 2 is lower than both the daily maximum and Position 1's maximum. Similarly, the minimum of the position is higher than both the daily minimum and Position 1's minimum.

The index values for each new maximum and minimum prices when they are first encountered are also joined. For different investors closing a position at different times of a day, a new maximum/minimum price level encountered happens at the same time. So for forward cumulative maximum and minimum price calculations, the index value at the first instance of a new maximum/minimum level is applicable to all investors. This feature is illustrated in Figure 10.

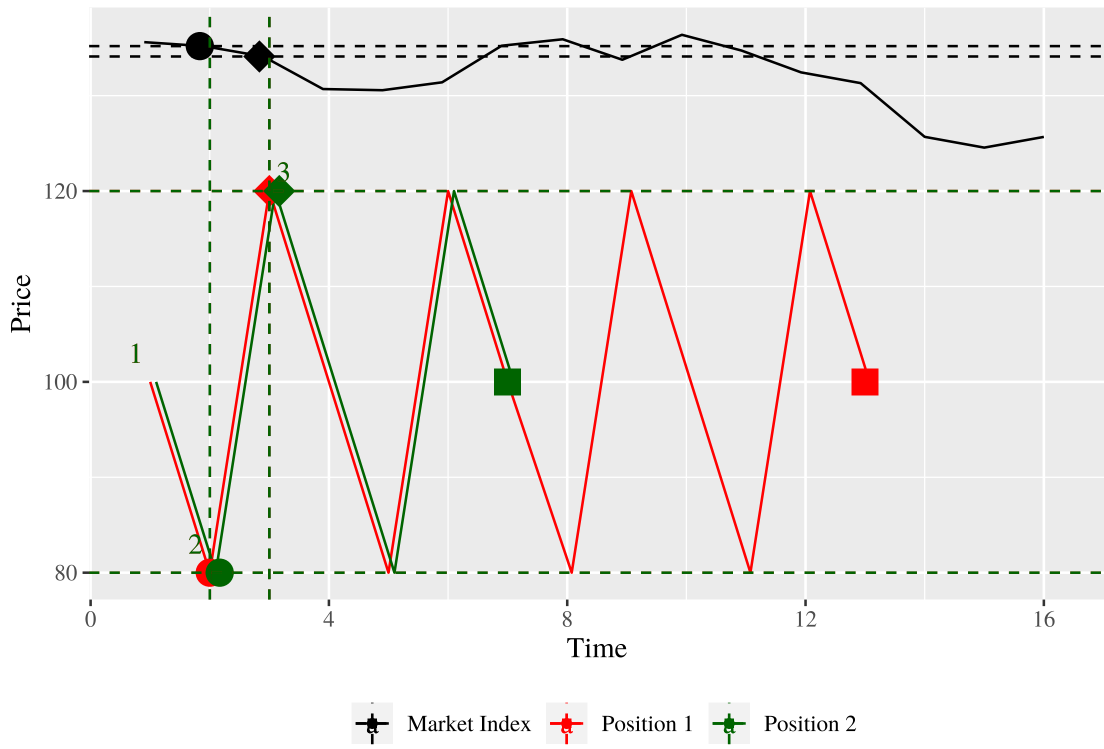

Figure 10. Determination of corresponding market indices for maximum and minimum references on the closing day of a position

Position 1 (red) is closed after Position 2 (green) on the same day. Squares mark the closing, circles mark the minimum and diamonds mark the maximum prices for each position and the corresponding market index. By observing the red and green circles and diamonds and the intersecting horizontal lines, it is apparent that experienced maximum and minimum prices of two positions have the same value, albeit they are encountered first at different times. In contrast with the case of the opening day, the corresponding market indices for the first times maximum and minimum prices are encountered by each investor are the same. This feature can be observed from the black circle and diamond and the intersecting horizontal lines. Similar to the previous argument, while each investor may encounter these maximum and minimum prices multiple times, the corresponding market index is determined with respect to the first encounter of each investor. Hence, it is not updated unless a new maximum or minimum price is experienced by the investor.

These forward cumulative maximum and minimum reference prices and corresponding index values are used for calculating the minimum and maximum reference prices on the day a position is closed with a sell transaction.

While the minimum and maximum prices are calculated daily for each position, the cumulative minimum and maximum prices tracked from the beginning of the position are taken as the bases for maximum and minimum references. These references are updated only when a new cumulative minimum or maximum price is reached.

There are 19,569,458 rows in the calculated intraday maximum and minimum price dataset and for each stock, date and times in second. Eight separate values are calculated for backward/forward, maximum/minimum and price/index combinations.

It is assumed that the buy or sell decision of an investor occurs exactly at the moment the transaction is realized in the above calculations. In reality, for most transactions there is a limit order explicitly given for a price level, and that waits in the order book for a matching market order. A sell (buy) market order instantly matches a buy (sell) limit order with the highest (lowest) bid (ask) price awaiting on the order book. The transaction dataset offers no way to figure out when the orders are placed into the order book if it is a limit order. Even when the order is a market order that is executed instantly matching the prevailing bid or ask price level, the exact moment that the investor decides the action may be some time before the actual transaction is realized. Since there is no way to figure out the length of this lag, it is assumed that the decision and realization of the transactions are simultaneous.

5.5 Filtering the positions

Since the simultaneous impact of multiple reference points on disposition effect is explored in this thesis, the perception of the reference prices and the overhang measure must be salient and unambiguous to the investor. For this reason, several filters are applied to the overall dataset, to arrive at the smaller set of positions. For this set of positions, the exact reference prices and the corresponding overhang measures can be calculated with precision for each day and the decision to close the position or not should also be clear. The main filters are as follows:

- Leaving only stocks with no splits during the data period
- Filtering simple positions with a single buy transaction first, and a single sell transaction if the position is closed
- Controlling for non-negative ending balances

- Requiring at least one day after the buy transaction if the position is closed, so intraday positions are filtered out

5.5.1 Ticker level capital increase and dividend filters

Capital increases and dividend payments blur the exact cost of the purchase to the investor. A capital increase may be done in one of the two methods or combinations of both:

- New cash proceeds may be requested from the investors using preemptive rights so that whatever the current market price is, the investor acquires the new shares from a predetermined nominal price. The new prices are adjusted for the split considering the new number of shares and the proceeds from the preemptive rights. The preemptive rights may be restricted in the case of public offerings. In this case, existing shareholders do not have the right to acquire the shares from a nominal price and shares are offered to public from the prevailing market price or a new offer price.
- Capital increase is just an accounting record from the capital reserves, including items such as revaluation funds, in the shareholders' equity. No new proceeds are received by the company. However, the number of shares increase, so the price is adjusted for the split.

The investor may use the preemptive rights, or sell those rights in the market separately or not use the rights so that they expire. In any case, it is impossible to determine from the transaction data which case has occurred for a position. Furthermore, the position becomes a complex one when new capital is invested after the initial buy transaction. And because of the split, it is ambiguous for most investors to track the exact cost base of the position.

The similar case is true for dividend payments. For cash dividend payments, new proceeds are received from the company after the initial buy transaction, a fact which complicates the calculation of the cost base. As a result, the position becomes a complex one, making the relationship between the overhang and the sell decision more blurred. Stock dividends are similar to stock splits with no preemptive rights. In this case, the capital increase is made from the retained earnings account. The cost base should be adjusted, a fact which investors might not adopt well.

To keep the cost base simple throughout the holding period, only those stocks that did not go through any capital increases or dividend payments during the period of the data are filtered.

There are 92 stocks that went through a capital increase between 1 February 2005 and 30 December 2005. And 105 stocks received cash or stock dividends within the same period. Some of these stocks are common, and the union of the two sets reveal that 172 stocks either had capital increase or a dividend payment within the period. There are 313 unique stock codes in the transaction database.

To make an anti join to find the set of stocks without any splits or dividend payments in the transaction dataset, another adjustment must be made. Some companies have multiple groups of shares traded simultaneously and separately on

BIST. The transaction dataset has separate codes for these shares, while capital increase and dividend payments datasets use a single code for a company by selecting one of the share groups. So first, the stock codes from the transactions are matched by their counterparts in the capital increase and dividend payments dataset:

- Adana Cement company's ADANA, ADBGR and ADNAC tickers in the transaction dataset are matched with ADANA ticker in the capital increase and dividend payment datasets
- Isbank's ISATR, ISBTR, ISCTR and ISKUR tickers are matched with ISBTR
- Karabuk Iron and Steel's KRDMA, KRDMB and KRDMD tickers are matched with KRDMD.

There are other companies with multiple share groups trading on BIST. However, the ones listed above are the only companies with multiple share groups and stocks of which appear both in the data codes from transactions and the codes from the capital increase and dividend datasets with any of the actions during the period of the dataset.

After the anti join, a control is made for any possible code change based on whether there are stock codes that appear only in one dataset and not the other, again through anti joins. There are several codes in the capital increase and dividend datasets that go through any of those actions during the data period, but do not appear on the transaction dataset. And there are also few codes in the transaction dataset that never appear in the whole date range of the capital increase and dividend datasets. The transaction dataset was compiled in 2005, while the capital increase and dividend datasets were last downloaded in 2009. So it is possible to have stock code changes during the period. Upon inspection, nine codes in the transaction dataset are changed in the other datasets so these old codes are matched with the new ones in order to make a complete anti join. The code changes are presented in Table 2.

After the anti join, there are 143 stock codes that appear in the transaction dataset and are free of any capital increase or dividend actions during the period. The transaction dataset is filtered to include only these stock codes.

The stock codes in the transaction dataset are further checked for any possible code changes within the date period. For this, from the historic prices dataset extracted from the transactions, those stock codes for which first date of transaction is later than 1 February 2005 or last date of transaction is before 30 December 2005 are filtered. They are reported with the corresponding first and last dates. The filtered ones are probed whether there is a relationship between pairs of codes in which one of them is introduced after the other ceases to trade. The results are provided in Table 3. Upon inspection, no code changes that require merging of codes exist. The stocks in

Table 2. Code Changes Across Transaction Dataset and Capital Increase and Dividend Datasets

Old Code	New Code
BSPRO	BSHEV
DISBA	FORTS
DOKTS	COMDO
DUROF	DURDO
EVREN	EGCYO
IHGYO	YYGYO
KOZAD	KOZAA
MMART	MARTI
PERYO	PEGYO

Table 3. First and Last Dates That Stocks Appear in The Transaction Dataset

Stock Code	First Date	Last Date
RAKSE	1 February 2005	30 May 2005
RKSEV	1 February 2005	30 May 2005
KCHOL	2 February 2005	30 December 2005
TCELL	2 February 2005	30 December 2005
YKBNK	2 February 2005	30 December 2005
YKFIN	2 February 2005	30 December 2005
YKGYO	2 February 2005	30 December 2005
YKRYO	2 February 2005	30 December 2005
YKSGR	2 February 2005	30 December 2005
ISATR	3 February 2005	19 December 2005
AKMGY	15 April 2005	30 December 2005
TSPOR	15 April 2005	30 December 2005
IBTYO	9 May 2005	30 December 2005
EVNYO	11 July 2005	30 December 2005
BIMAS	15 July 2005	30 December 2005
ANELT	13 September 2005	30 December 2005
HDFYO	6 October 2005	30 December 2005
VAKBN	18 November 2005	30 December 2005
MRTGG	16 December 2005	30 December 2005

the table appear after the first date of the data period probably due to a regulator action to limit abnormal price movements or being listed on a later date, and they disappear before the last date of the data period because of delisting.

5.5.2 Position level filters

For the study only simple positions where all stocks are bought at once at the same price and sold at once at the same price if the position is ever closed during the data period are filtered. In these simple positions, the cost, maximum or minimum reference prices are clear, salient and unambiguous. And the reference prices are easy to compare with the current prices to get the overhang measures. The rules for the position level filters are listed below:

- All transactions are made in at most two separate days.
- The transactions made in a day are either all buy or all sell transactions, so they are completely in the same direction, for each day in the transactions. This filter also rules out intraday positions that are opened and closed on the same day, since intraday positions' underlying dynamics might be different from the dynamics of positions with a duration longer than a day.
- The transactions on the first day of the position are all buy transactions.
- If the position has a later day of transactions, they are all sell transactions.
- The balance after the sell transactions is non-negative. A negative balance means either the position is short or it is started before the period and should be left truncated.
- If the transactions are made in only one day, and as a result the position is not closed, that day is not the last day of the data period or not the last day a stock is traded if it ceased to trade or was delisted before the end date.
- All transactions in a day are made in the same second and at the same price. A single buy or sell order when matched with a counter[12] order is not guaranteed to be with the same quantity. In that case orders are split so that orders of different sizes can be matched to clear the market. The time and price restrictions ensure that the transactions come from a single order and are the result of the decision to buy or sell at once.

After the filters, there are 317,931 positions left in the final set. Total number of separate transactions is 725,969. However, once the transactions made in a day are aggregated, total number of transactions drops to 574,447. Number of positions that

[12] A market buy/sell order is matched with a limit sell/buy order.

are opened with a buy transaction and closed with a sell transaction is 256,516. 61,415 positions are opened with a buy transaction but have no subsequent sell transactions. Hence, they stay open by the end of the data period. For the purposes of survival analysis, they are treated as right censored: Their data until the end of the period are included in the analysis while they do not have any sell status. The sell status of closed positions corresponds to the *failure* concept in the survival analysis literature.

In closed positions with an initial buy and a later sell transaction, it is possible that not all shares are sold off. However, when 256,516 closed positions are explored, 98.5% of the total positions are completely closed, so that number of sold shares are exactly equal to number of shares previously bought. 0.9% of the total positions are less than completely and at least half closed. And just 0.6% of the total positions are less than half closed. So a great majority of the total positions are completely closed.

Since unclosed positions do not encounter any sell transactions as at the last date of the data period, they by definition occupy a longer time span in the final dataset than closed positions do. As can be seen from Table 4 and Figure 11, closed positions with two transactions concentrate on earlier holding periods while unclosed positions with a single transaction are dispersed along longer holding periods: 25%, 50% and 90% of all positions are reached in two, five and 44 days in closed positions. The same marks are reached in eight, 29 and 205 days in unclosed positions respectively.

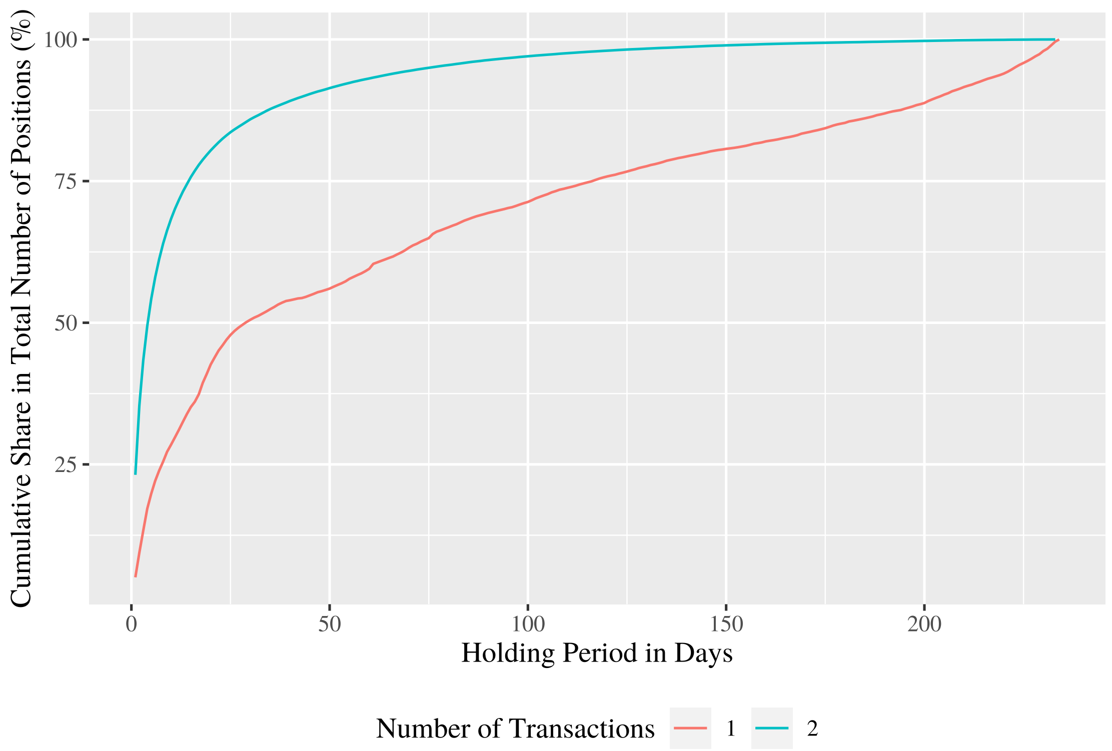

Figure 11. Cumulative share of positions across holding periods by number of transactions per position

For this reason unclosed positions dominate longer holding periods. As can be seen from Figure 12 share of closed positions go down from 95% at day one to 79%, 55%, 15% and 0% on days 50, 100, 200 and 234 respectively.

Table 4. Cumulative Share of Positions Across Selected Holding Periods by Number of Transactions per Position

Holding Period in Days	Positions With One Transaction	Positions With Two Transactions
1	5.07	23.16
2	9.40	35.35
3	13.33	43.41
4	17.18	49.42
5	19.83	54.21
6	22.09	58.01
7	23.90	61.22
8	25.49	63.96
9	27.22	66.29
10	28.48	68.33
20	42.62	80.41
30	50.53	85.91
40	53.97	89.15
50	56.03	91.41
100	71.32	97.03
150	80.69	98.96
200	88.79	99.74
234	100.00	100.00

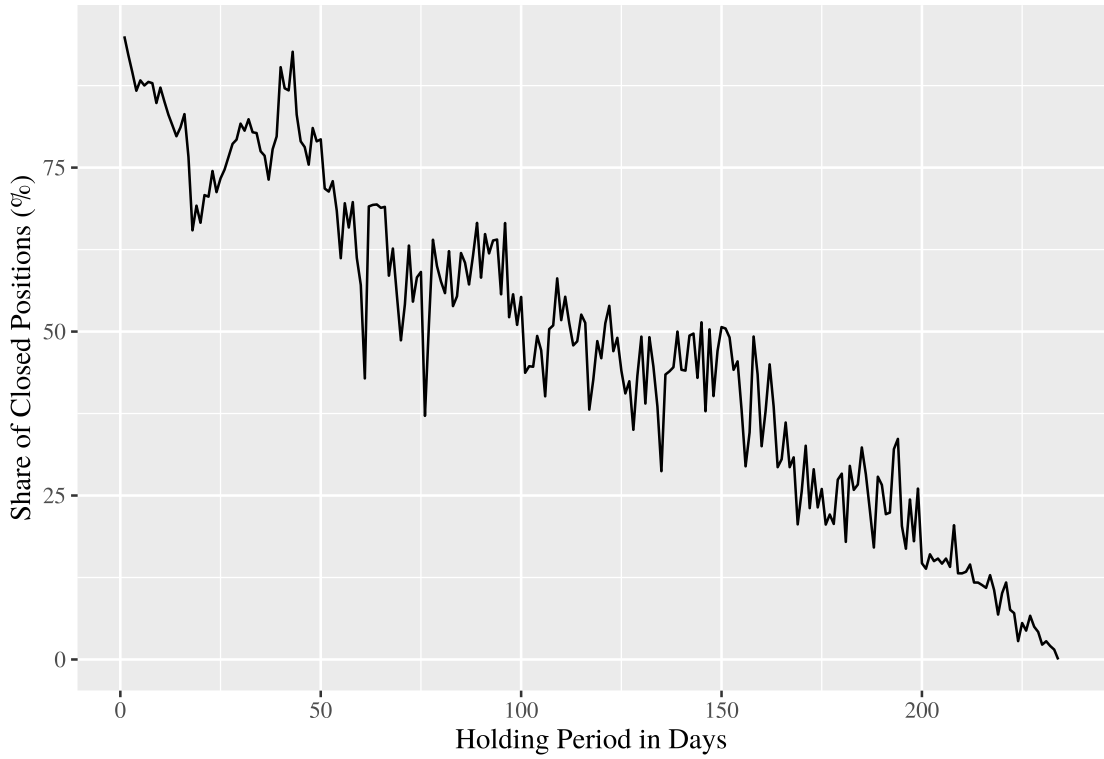

Figure 12. Share of closed positions across holding periods

5.6 Filling in dates

The data format suitable for further disposition effect analysis using survival methods is the panel format. In this format, all dates between the opening and closing transactions of the positions are filled for closed positions. All dates between the opening transaction and the end of the data period or the last trading day of the stock are filled for unclosed positions. The corresponding number of days since the opening of the position is also tracked. This figure counts the trading days and not the calendar days.

The transactions should be aggregated per position per day, by their quantities and total values for the cost prices and the prevailing BIST index values at the exact time of purchase separately. However, since only positions in which all buy or sell transactions occur at the same second and at the same prices are filtered, the weighted averaging has no effect. This stock cost and the corresponding BIST index value act as the reference cost price and index value at cost for further overhang calculations. If the position is closed, selling price and the prevailing BIST index at the exact time of sale are filled in for the closing date. The minimum and maximum prices for each date and the BIST index values at the exact time those extreme prices occur in each day are also added.

At the filtering phase, the last date that each stock trades is figured out and the date fill-ins are made accordingly. Even before the last trading days, for each position, there are days for which the underlying stock has never traded in the market at all. So aggregated historic prices of those stocks for these dates are missing. They are also deleted from the dataset. The regulator may close the trading for a stock when some unusual price movements or any undisclosed developments occur. The missing dates could also be completed using last observation carried forward method. That may be done in event studies not to lose observations due to missing prices. But in a behavioral study to explore disposition effect, there is no point in including a day for a position with missing price information since the investor cannot act upon the decision to close the position when the stock is closed to trading. When the price information is imputed, it is implicitly assumed that the investor did not choose to close the position at that date. So only days with active trading for the stock are left in the dataset. 13,919 rows are deleted this way.

For the dates with a buy or sell transaction, the intraday aggregated historic dataset with seconds precision is joined with the main panel dataset to match the intraday maximum and minimum stock prices encountered and the BIST index value exactly at the time that minimum and maximum prices occurs, separately for each position. Backward (forward) calculated index values are used to match the maximum and minimum stock prices and corresponding index values that occur on the day and at the exact time or after (before) a stock is bought (sold), for each position.

After minimum and maximum prices on transaction dates are updated from the values from intraday data, cumulative minimum and maximum prices that serve as alternative reference points are calculated for each position and each day.

The panel format dataset with filled in values for all dates for a position has a total of 8,608,331 rows.

5.7 Calculating the basic variables

In the next step, some basic and important variables are calculated. Overhang measure, the unrealized gain or loss, is the percent that current price is above the reference price. For the position-days combinations with no buy or sell transactions, the current price measure can be calculated from two options: Maximum intraday price and weighted average price. Using the maximum intraday price as the current price measure is more realistic, since an investor is supposed to seek the best intraday opportunity to close a position. The tests for all models are done on the variables calculated using maximum intraday price as the current price. For the corresponding index values, the maximum index value is the index value exactly at the time maximum price is experienced.

An advantage of calculating the overhang with respect to a fixed reference point[13] over the returns calculated within a fixed length moving time window is its

[13] Maximum and minimum price references for an investor in a position can be updated as new extremes are experienced. However, these prices are not automatically updated within sliding time ranges as in the case of fixed length time window return calculation.

robustness. In the case of fixed length moving time window, the same current price may yield a positive or negative return depending on the choice of the width of time window, e.g. daily, weekly, or monthly returns. Hence, results may not be robust to different window length choices. This fact can be illustrated with simulated data in Figure 13.

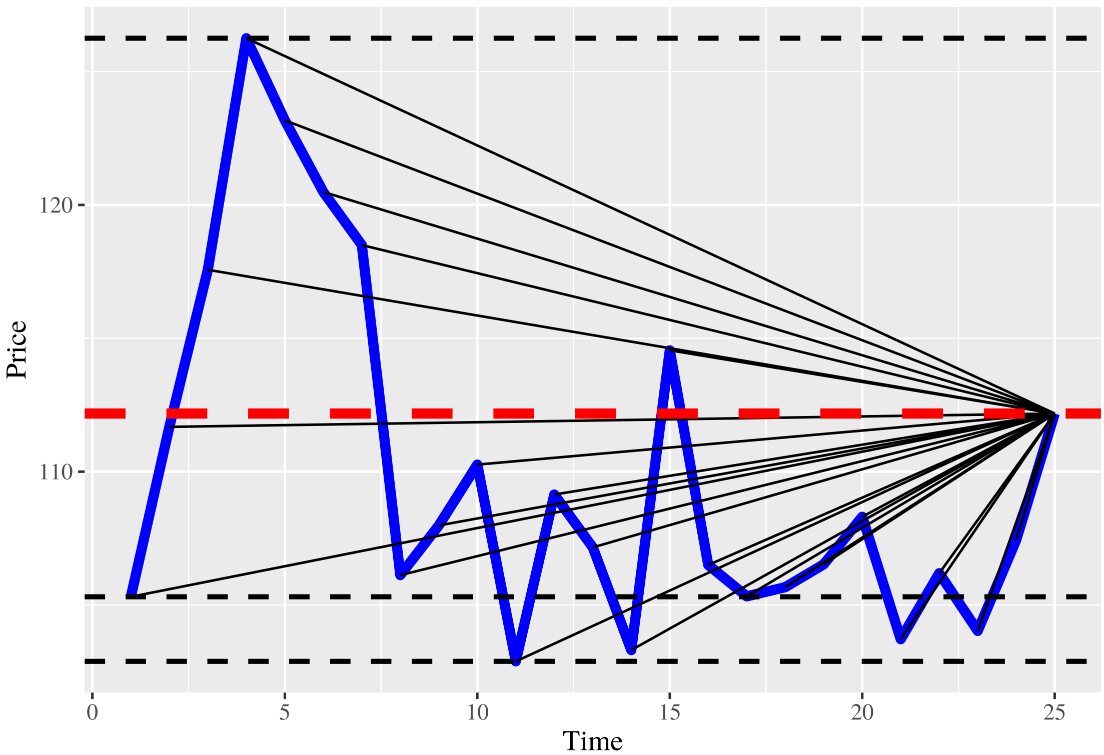

Figure 13. Stock returns over different horizons in a fluctuating market

In this simulated fluctuating time series, new maximum and minimum levels are attained over the course with alternating trends. Last price is above the purchase price which is the first price in the series. Dashed horizontal lines mark the levels of original cost, maximum and minimum prices. Red bold dashed horizontal line marks the level of the current or last price. The thinner sloped lines show the trend and return between each of the previous prices and the current price. In six of the total 24 possible horizons, the current price level is associated with a negative return while for 18 horizons, the current price level is associated with a positive return. For investors who consider past returns calculated within a certain time window in their decisions, that return horizon is not guaranteed to be consistent across their personal choices. However, in the case of calculating overhang measures, once the reference prices are determined properly, there can be no discretion that may change the direction and magnitude of these overhang measures.

For the reference prices, there are three simultaneous alternatives tracked together:

- The cost price
- Cumulative minimum price since the position is opened
- Cumulative maximum price since the position is opened

While the overhang can be positive, zero or negative with respect to cost price, it is always negative or at most zero against the maximum price and always positive or at least zero against the minimum price.

On the day a position is closed, the current price is the actual selling price, and the corresponding BIST index value is the value exactly at the time of the transaction.

All return measures for overhang calculations are done using the logarithmic method as in Equation (1).

$$\ln \frac{P}{R} \tag{1}$$

where,

- P is the current price.
- R is the reference price.

The calculated overhang measures are denoted with an oh prefix as variable names. Three basic overhang measures, with respect to cost, (cumulative) reference minimum and (cumulative) reference maximum prices are calculated.

A status variable, which is the dependent variable Cox proportional hazards model tries to predict, is added, which takes the value of 1 whenever a position is closed on a certain day with a sell transaction and 0 otherwise.

In order to control overhang measures with respect to reference minimum and maximum prices with holding period and market returns, the corresponding BIST index values and days since the most recent minimum and maximum price is encountered are also added. Whenever a new minimum or maximum price level is reached for a position, the number of days since the minimum or maximum references are reset and the BIST index corresponding to those reference prices are updated. So, three reference prices are tracked:

- Cost price, fixed at the actual level on buy transaction
- Maximum price, cumulative maximum since the exact second the position is opened, until the exact second the position is closed (if any sell transactions occur) and calculated for each day
- Minimum price, cumulative minimum since the exact second the position is opened, until the exact second the position is closed (if any sell transactions occur) and calculated for each day.

For each of those reference prices:

- Number of days since the purchase or last maximum or minimum reference price update

- BIST index price that exactly corresponds to the instant the cost, minimum and maximum reference prices occur or are updated
- The overhang measures of current prices with respect to those reference prices

are also tracked and calculated.

5.8 Calculating additional numeric variables

Overhang measures for all current and references prices and corresponding index values are calculated separately. Overhang is simply the logarithm of the ratio between the current price and respective reference price (cost, maximum or minimum prices). The index returns corresponding to each overhang measure are calculated as the logarithm of the ratio between the current index value and respective index value at the exact time the reference price is formed or updated.

Some variables are intermediate variables to complete the wrangling steps and calculate the variables to be discretized and included in models. All variables that appear in the original data and kept in the final dataset and numeric variables calculated in this thesis are detailed in Appendix A (p. 217).

After all the numeric variables are calculated, rows with the zero days since purchase, that denote the days in which a position is opened are excluded from the final dataset to be entered into the models. The reason is that intraday positions are also excluded and the sell status in the first day will not be evaluated. A total of 8,290,400 rows are left for modeling.

5.9 Ordinal discretization of numeric variables

Following Shumway and Wu (2005), the selected covariates are introduced in to the Cox model not as numeric variables, but discretized ordinals. In this way, no shape is imposed on the relationship between the covariates and the status variable. Especially extreme values in tails are treated better with ordinal variables. Furthermore, the interactions between ordinal variables allow for modeling more complex relationships in a highly flexible manner. Also from a psychological perspective, individuals have a tendency to perceive continuous features as categorized phenomena as detailed in Subsection 2.4.1 (p. 43).

An innovation brought by this thesis is the discretization of variables controlling for other ordinal variables in a multivariate and hierarchical setting. This allows for a balanced distribution of variables especially across the extreme categories of other ordinal variables. The effect of the method is detailed below in steps using a set of selected variables treated together to create balanced categories.

In this thesis, it is hypothesized that the perception of the level of overhang, the return of the current price with respect to a reference point, is formed within the contexts of holding period and the market return as experienced by the investor during the course of the position. The hypothesis can be confirmed in case models incorporating variables accounting for these contexts predict investor behavior better than models with naive versions of the same variables.

In order to demonstrate the effect of MUHBOD, the results of a naive approach for discretization are contrasted with those of the controlled one. In each case, the level of quantiles and the distribution of the case counts across controlling ordinals are compared.

5.9.1 Naive and imbalanced ordinal discretization

As a first example, the main numeric variable oh_cost, overhang with respect to the cost reference price, is discretized naively into five bins from quantiles as the ordinal variable oh_cost_05. The four quantile values used as cutting points for five bins are -0.13, -0.04, 0.01 and 0.08 and the resulting number of cases across ordinal bins are 1,658,782, 1,659,438, 1,656,913, 1,657,558 and 1,657,709 respectively as shown in Figure 14. The minimum and maximum points for the bins are a result of the calculated quantiles. In the plot, ranges for only the second to the fourth bins of the total five bins are shown. The first bin ends at where second bin starts and the last bin starts at where the fourth bin ends.

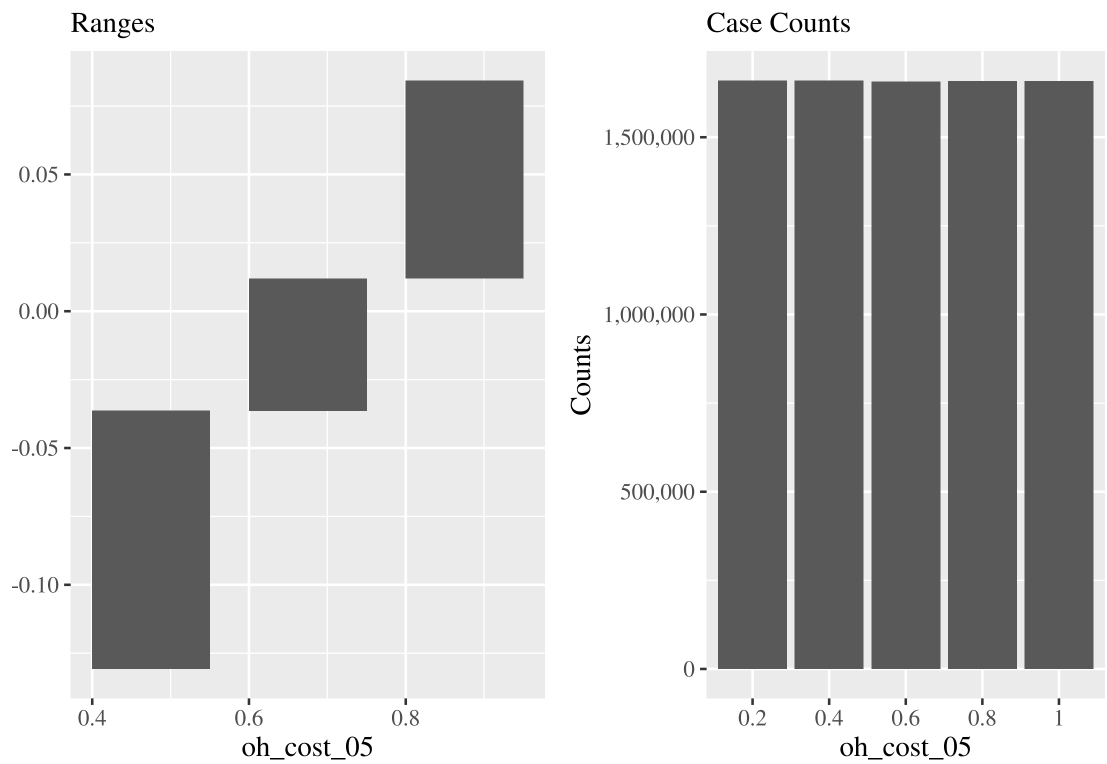

Figure 14. Ranges and counts of five ordinal categories of oh_cost_05

Across a single variable, for the categories created within quantiles for the same variable, the distribution of counts is uniform, by definition.

Next, the min-max ranges and the counts for oh_cost_05 are repeated for the rows which correspond to each of the categories of days_cost_05 as shown in Figure 15. days_cost_05 is the discretized version of the days since the first trading date of the position, initial day excluded from the dataset.

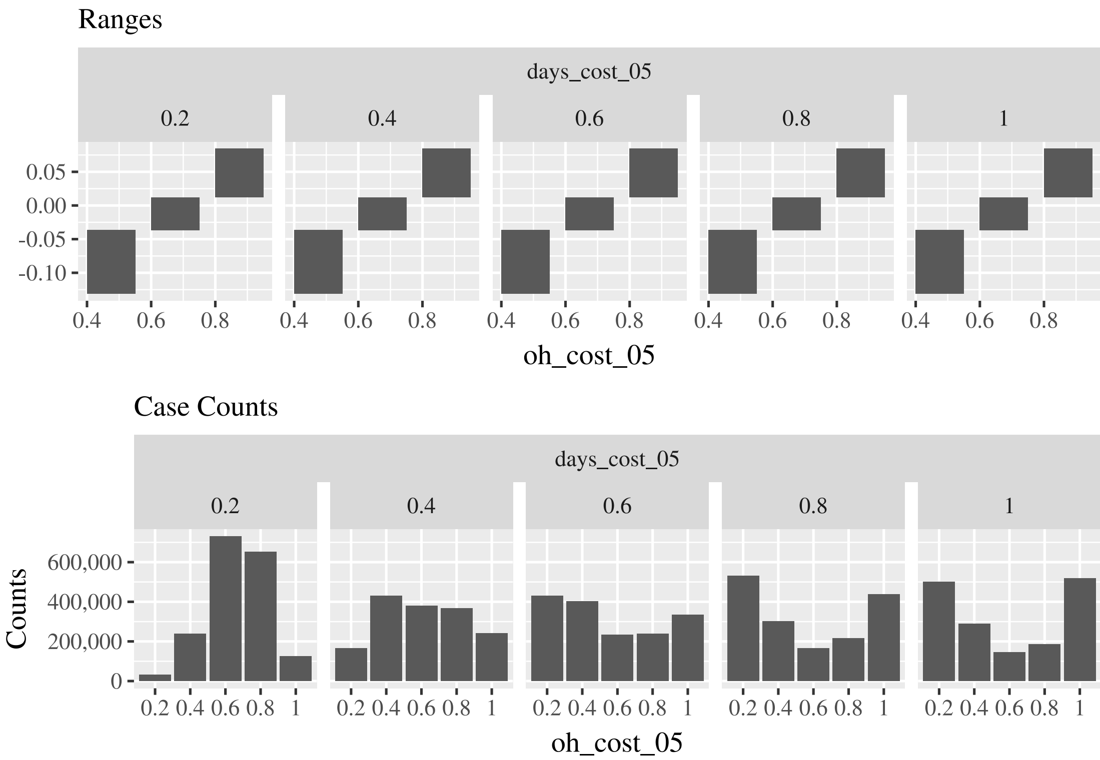

Figure 15. Ranges and counts of five ordinal categories of oh_cost_05 across days_cost_05

While, the min-max ranges stay the same across days_cost_05 categories, the distributions of the counts vary widely: In the leftmost category for the most recent days, the middle oh_cost_05 categories have more weight, while in the rightmost category for the most distant days, extreme oh_cost_05 categories in left and right ends have more weight. An overhang value close to the median is common, while an extreme overhang value is rare in recent days. However, an overhang value close to the median is rare while an extreme overhang value is common in distant days. A certain fixed overhang value can be a member of different ordinal categories for different time horizons. In order to correct for this imbalance, the perception of high or low overhang values should be a function of the horizon since the opening of the position.

In the next example, the min-max ranges and distribution of case counts of oh_cost_05 are calculated across two ordinal variables: days_cost_05 and oh_xu_cost_05, the discretized market (BIST 100) return since the opening of the position, as shown in Figure 16 and Figure 17.

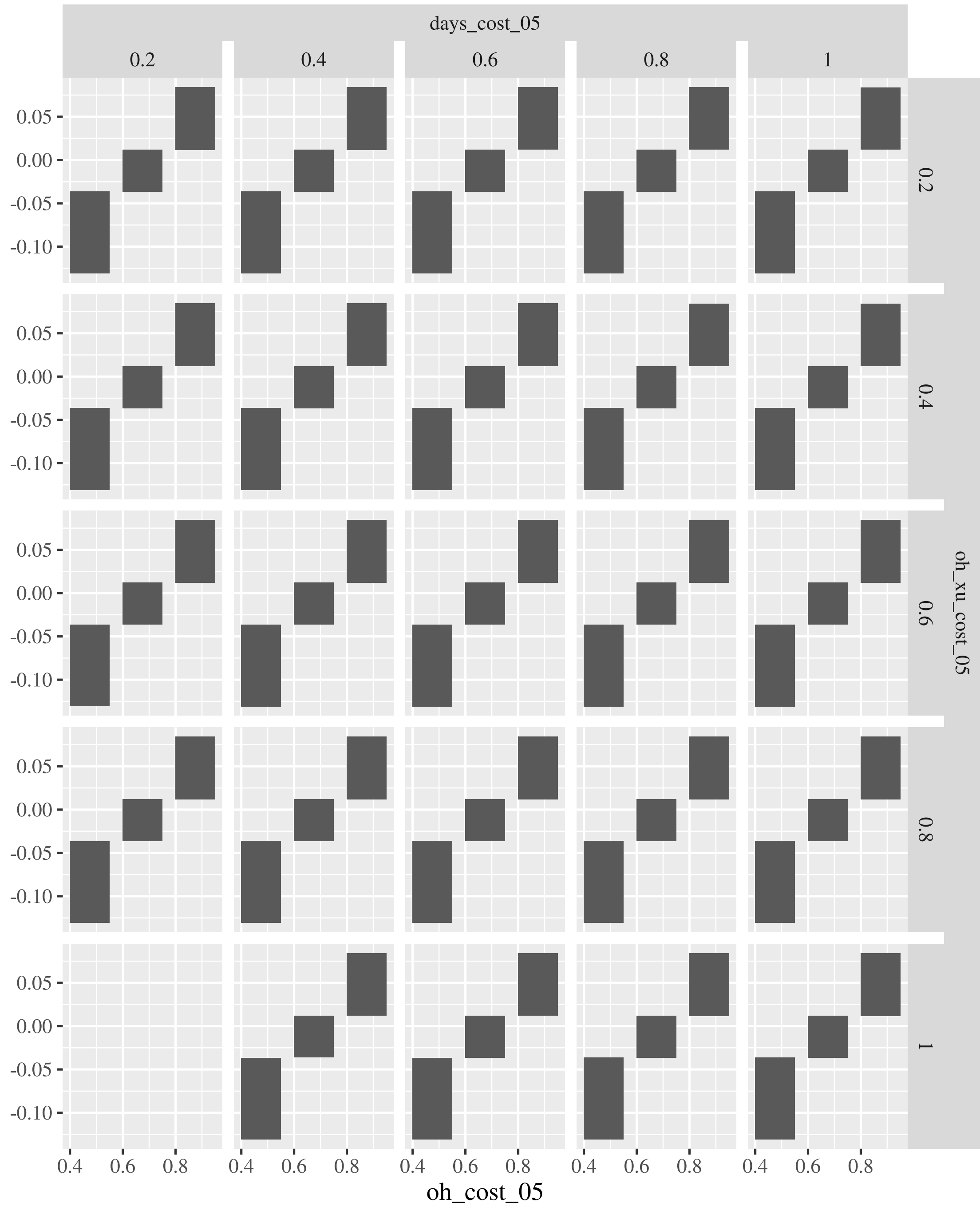

Figure 16. Ranges of five ordinal categories of oh_cost_05 across oh_xu_cost_05 and days_cost_05

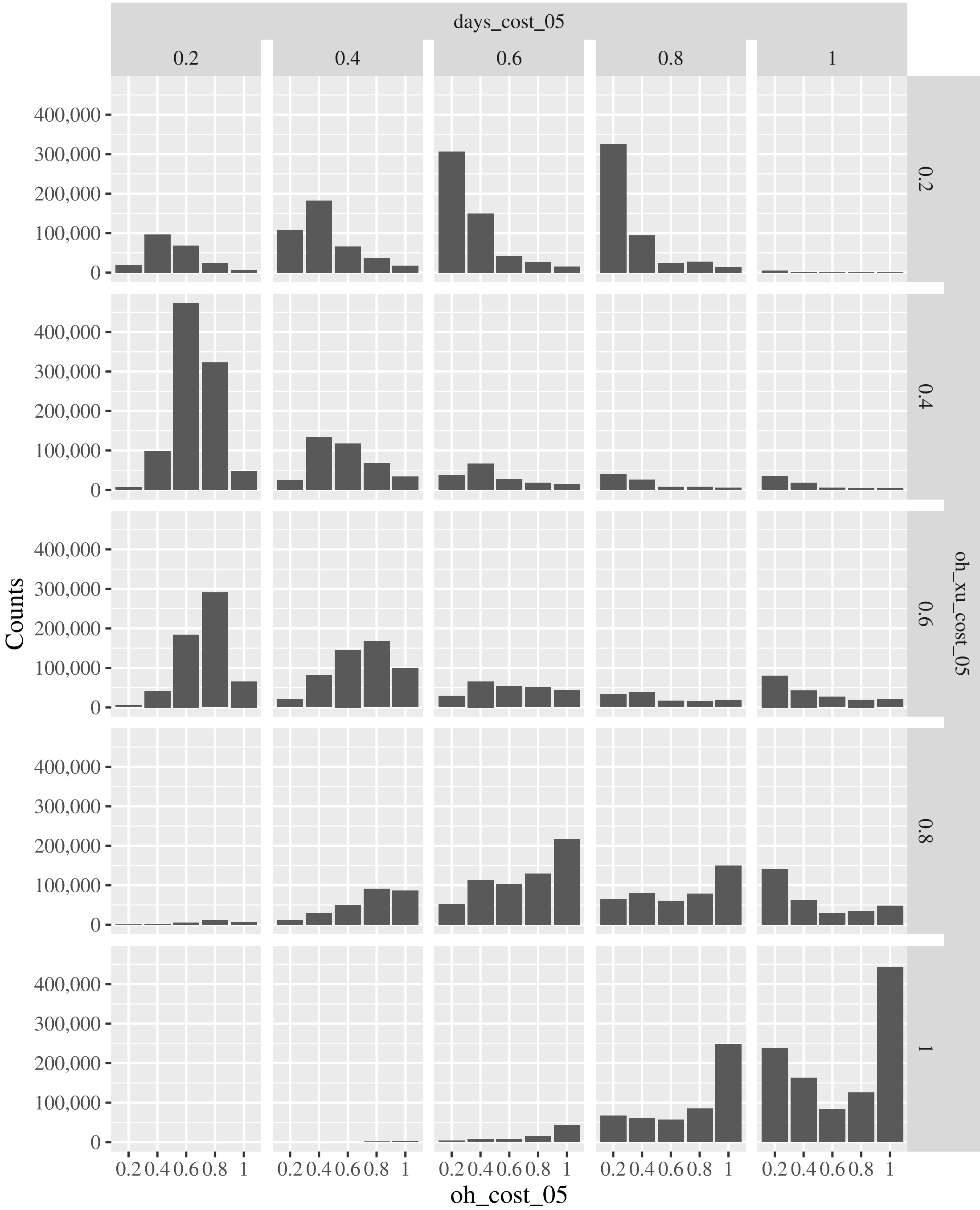

Figure 17. Counts of five ordinal categories of oh_cost_05 across oh_xu_cost_05 and days_cost_05

All min-max ranges for oh_cost_05 are the same. Hence, the cut-off points for oh_cost_05 are set as universal across all levels of the other two ordinal variables. However, the distribution of the overhang values shift from left (categories for lower values) to right (categories for higher values) as one moves from oh_xu_cost_05 categories with lower market returns to categories with higher market returns. So lower (higher) overhang values are more common in lower (higher) market returns. As a result, a certain fixed overhang value might be perceived as high as a lower market return is experienced, while it might be perceived as low in a higher market return setting. When one moves from left days_cost_05 categories (closer horizons) to right days_cost_05 categories (distant horizons), the same effect as detailed above, the shift of density from the center to the extremes, also occurs simultaneously.

In order to correct for this imbalance, the cut points for oh_cost_05 categories should be set separately for each combination of categories for the controlling ordinal variables. With this correction, the distribution of case counts can still be balanced across oh_cost_05 categories as in the univariate case.

5.9.2 Multivariate hierarchical balanced ordinal discretization

Multivariate hierarchical balanced ordinal discretization, or MUHBOD for short, is a term coined in this thesis. In MUHBOD, multiple numeric variables are discretized in a hierarchical manner to ensure that the main target ordinal variable has balanced counts across categories for all level combinations of the other ordinal variables.

Through MUHBOD, the ordinal levels of overhang are calculated to be context-sensitive, so that the time horizon and market return between the time of the reference points and the current prices are also taken into consideration. To understand how the categorical perception of numeric overhang levels can be dependent on the time horizon and market return, a simple example with simulated data is illustrated in Figure 18.

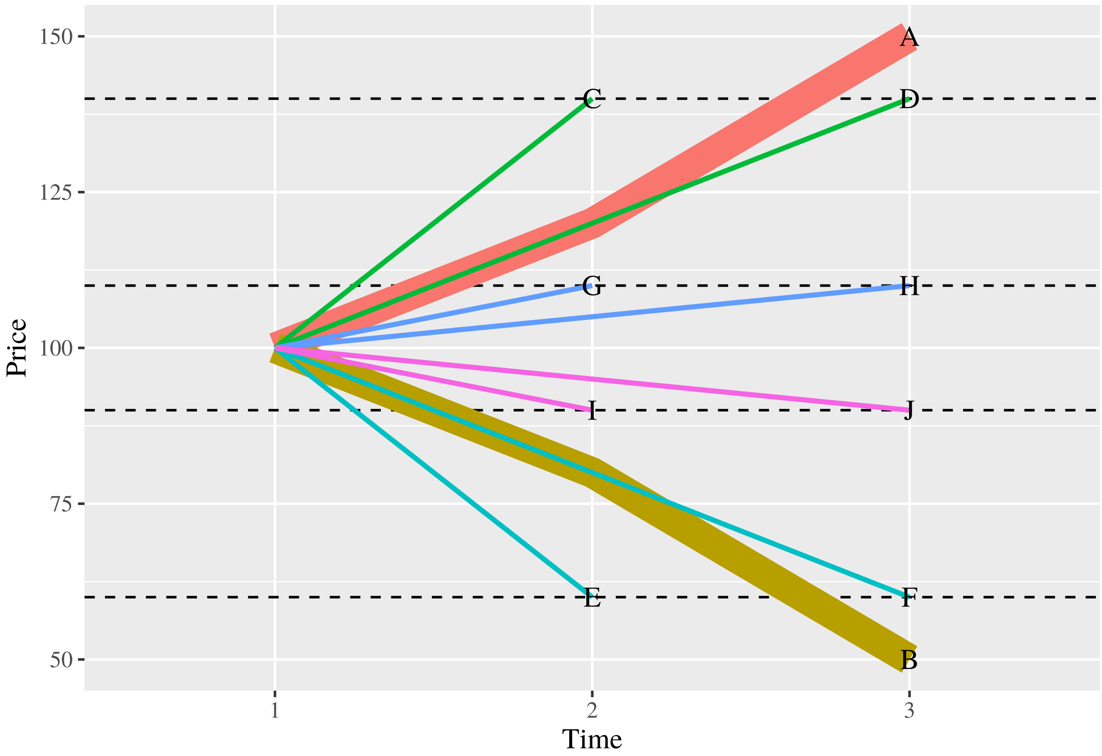

Figure 18. Comparison of stock returns with market returns across time

X axis shows the time horizon where 1 is the starting date of the positions, 2 represents a shorter time horizon while 3 represents a longer time horizon. The bold lines A and B are up and down market states. C and D show two positions with the same positive level of return but with shorter and longer time horizons respectively. While return of C is above that of the up market A in the short term, the same level of return of D is below that of the up market in the long term. Similarly, E and F show two positions with the same negative level of return in the shorter and longer time horizons respectively. While return of E is below that of the down market B in the short term, the same level of return of F is above that of the down market in the long term. So the same level of return can be perceived as higher or lower in different time horizons.

G and H show two positions with a moderate positive level of return in the shorter and longer horizons, while I and J are mirror counterparts with a negative return level. G and H positions with a positive return will be perceived as high return positions when compared with the return of the down market state B. However, the positive return level for both positions is below that of the up market state A. Hence, that return level can be considered as low in the relative sense. I and J positions with a negative return will be perceived as low return positions when compared with the return of the up market state A. However, the negative return level for both positions is above that of the down market state B. Hence, that return level can be considered as high in the relative sense.

To achieve context-sensitive discretization for cost reference and with three variables, oh_cost_05 being the target variable, the following steps should be taken:

- Most basic variable is univariately discretized across quantiles. In the following example, days numeric variable, number of days since opening of the position, is discretized into days_cost_05 with five ordinal categories
- Next level variable is discretized separately for each level of the discretized basic variable. In the following example, oh_xu_cost numeric variable, the market return since the opening of the position, is discretized using separate quantiles for each level of days_cost_05. The resulting ordinal variable is oh_xu_cost_by_days_05. To ensure balanced distribution across the categories of controlling variable, the quantiles as cutting points for discretization may differ for each category
- Last and target variable is discretized separately for each level combination of the discretized basic variable and the discretized second level variable. In the following example, oh_cost numeric variable, overhang on cost of position, is discretized with

separate quantiles for each level combination of days_cost_05 and oh_xu_cost_by_days_05. The resulting variable is oh_cost_by_xu_by_days_05. In this way the count distribution of all levels of oh_cost_by_xu_by_days_05 is balanced across both controlling variables, while the cut-off points vary.

Following these steps, we start with days_cost_05 variable as seen in Figure 19.

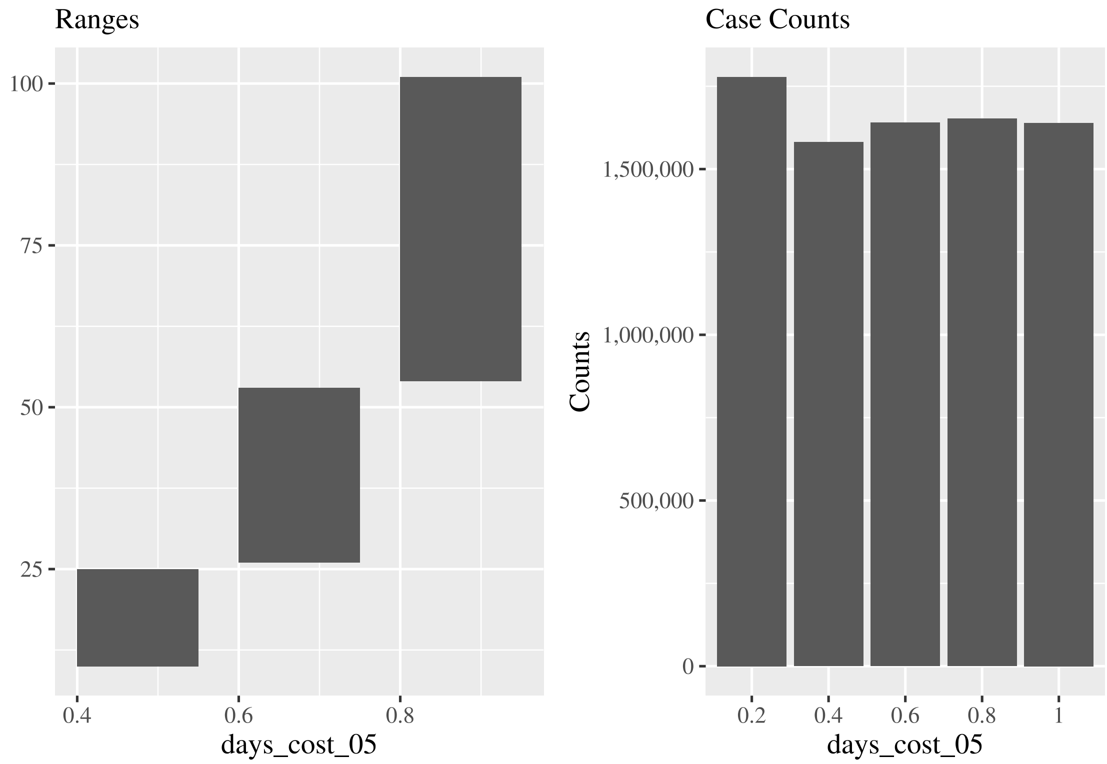

Figure 19. Ranges and counts of five ordinal categories of days_cost_05

Then we go on to the second step to discretize oh_xu_cost variable into oh_xu_cost_by_days_05 for each category of days_cost_05. As seen in Figure 20, the min-max ranges for categories of oh_xu_cost_by_days_05 become wider and higher from closer to more distant horizons of days_cost_05 variable. However, the distribution of counts of oh_xu_cost_by_days_05 stays balanced for all levels of days_cost_05.

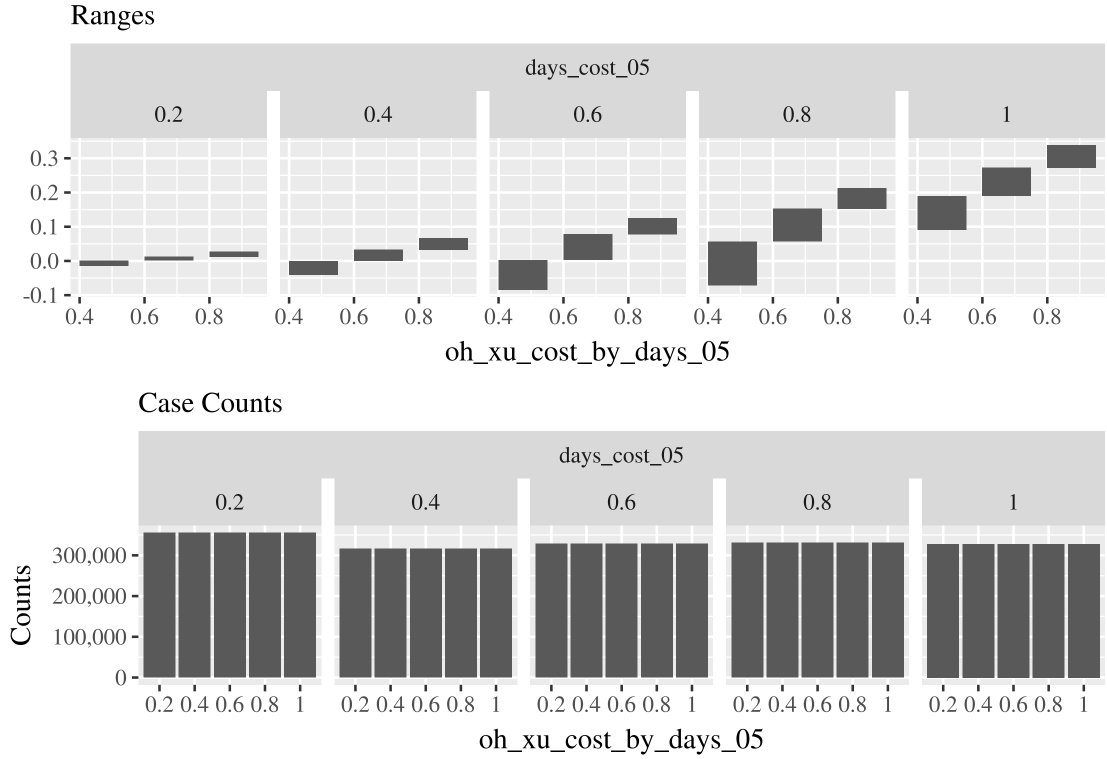

Figure 20. Ranges and counts of five ordinal categories of oh_xu_cost_by_days_05 across days_cost_05

For the last step, oh_cost is discretized into oh_cost_by_xu_by_days_05 for all level combinations of oh_xu_cost_by_days_05 and days_cost_05.

As shown in Figure 21, as we go from closer to more distant holding periods of days_cost_05 left to right, the min-max ranges become wider. So for a certain ordinal category of oh_cost_by_xu_by_days_05, when the holding period is longer, more extreme absolute values are members of that category. As we go from lower to higher market returns of oh_xu_cost_by_days_05 variable from the top to the bottom, min-max ranges become higher. So for a certain ordinal category of oh_cost_by_xu_by_days_05, when the market return is higher, higher values are members of that category.

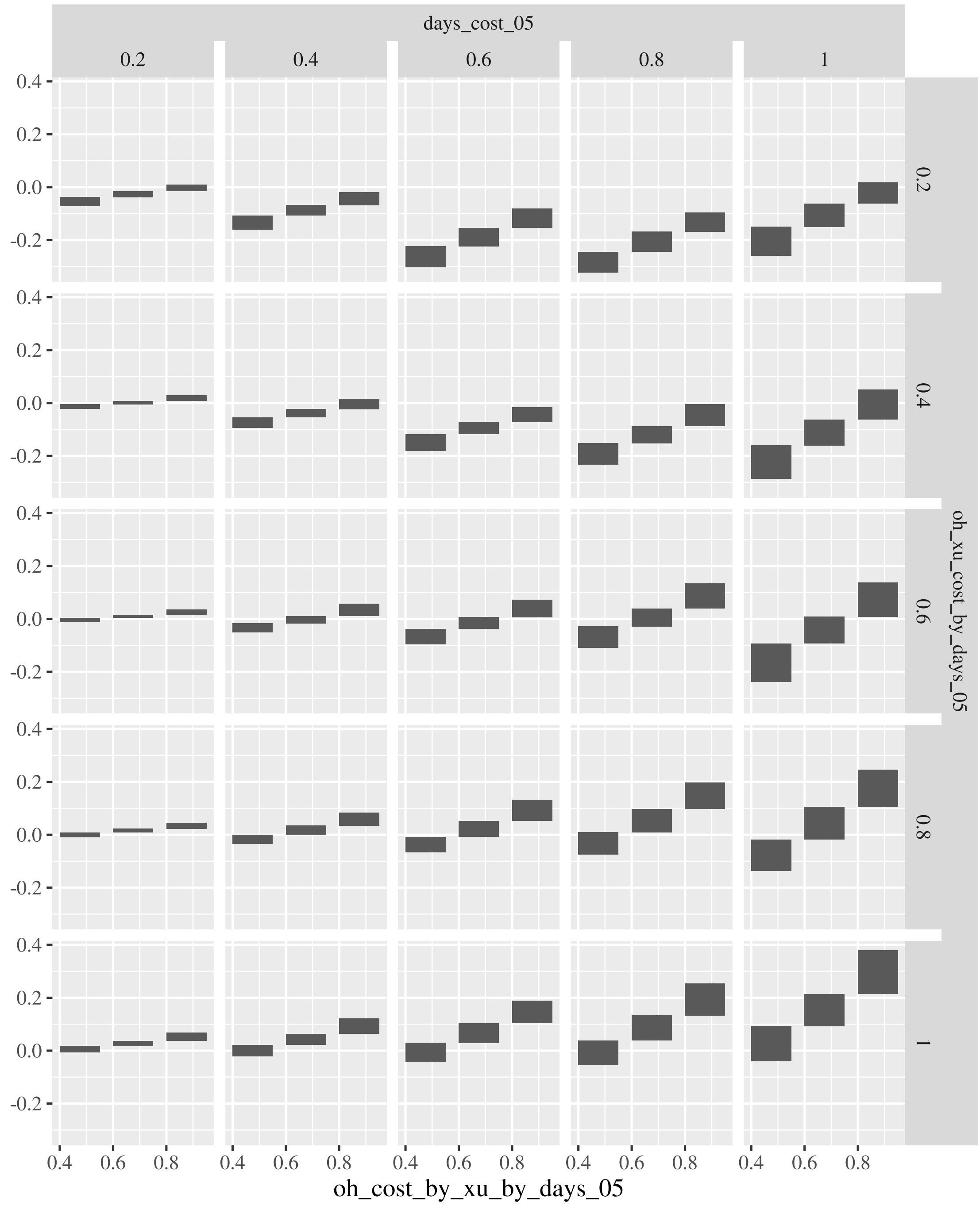

Figure 21. Ranges of five ordinal categories of oh_cost_by_xu_by_days_05 across oh_xu_cost_by_days_05 and days_cost_05

As seen in Figure 22, with this hierarchical approach, the distribution of counts of oh_cost_by_xu_by_days_05 levels remains balanced across all levels of oh_xu_cost_by_days_05 and days_cost_05. So the cut points for determining ordinal levels of overhang on cost is controlled for holding period and market returns.

This hierarchical and balanced ordinal discretization is repeated for minimum and maximum reference points using their corresponding variables for holding periods and market returns. All ordinal variables are calculated in three, five and ten bins separately, indicated by the suffixes of the variable names.

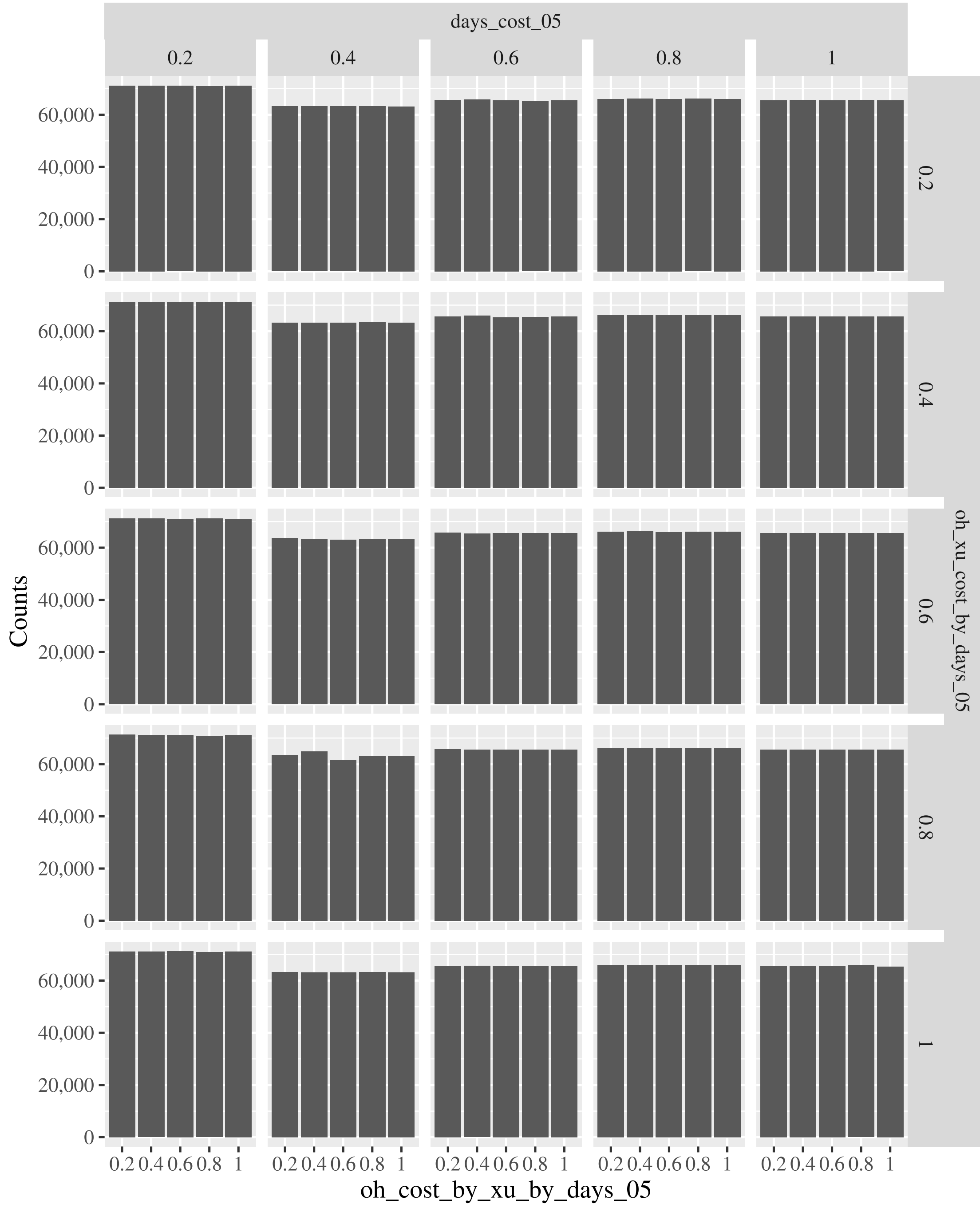

Figure 22. Counts of five ordinal categories of oh_cost_by_xu_by_days_05 across oh_xu_cost_by_days_05 and days_cost_05

Reference levels for each ordinal variable can be either minimum, median or maximum category of that variable. The reference level for a variable determines the baseline for the Cox model. The baseline is the case in which all covariates have zero values. When categorical variables are included in a model, the first levels of each categorical variable are automatically dropped during the model matrix preparation phase in order to prevent linear dependence. In order to ensure that the true reference level is the dropped level and represents the baseline, that reference level is brought to front as the first level. For example,

- For overhang on cost reference price related ordinals, the median level is the reference one. Because after the position is opened, the overhang can go lower or higher.
- For the overhang on maximum reference price related ordinals, the maximum level is the reference one. Because the current price can at best be equal to the most recent maximum price and can only go lower than that level.
- For the overhang on minimum reference price related ordinals, the minimum level is the reference one. Because the current price can at worst be equal to the most recent minimum price and can only go higher than that level.

A total of 14 ordinal variables are calculated from numeric variables. Each ordinal variable is created for three, five and ten equal sized bins separated by quantiles.

All ordinal variables are detailed in Table 5. Factor Prefix column shows the prefix of the name for the variable and is followed by _03, _05 or _10 based on the number of bins. Numeric Variable column shows the numeric variable that serves as the basis for the ordinal variable through discretization. Grouping Factor Prefixes column lists the prefix of the ordinal variables for each combination of which the numeric variable is discretized across selected number of equal sized bins. They serve as the controlling variables for hierarchical discretization.

Table 5. List of Factorized Variables

Factor Prefix	Numeric Variable	Grouping Factor Prefixes	Definition
oh_cost	oh_cost	-	Overhang over cost
oh_xu_cost	oh_xu_cost	-	Market return between position opening and current date (market return since cost)
days_cost	days	-	Days since the opening of position (days since cost)
oh_cost_by_days	oh_cost	days_cost	Overhang on cost for each investment horizon category (overhang on cost, controlling for days since cost)
oh_xu_cost_by_days	oh_xu_cost	days_cost	Market return since cost for each investment horizon category (market return since cost, controlling for days since cost)
oh_cost_by_xu_by_days	oh_cost	oh_xu_cost_by_days, days_cost	Overhang on cost for each 1) market return since cost, controlling for days since cost category 2) days since cost category (overhang on cost, controlling for market return since cost and days)
oh_min	oh_min	-	Overhang over latest minimum
days_min	days_min	-	Days since the latest minimum (days since minimum)
oh_xu_min_by_days	oh_xu_min	days_min	Market return since minimum for each investment horizon category (market return since minimum, controlling for days since minimum)
oh_min_by_xu_by_days	oh_min	oh_xu_min_by_days, days_min	Overhang on minimum for each 1) market return since minimum, controlling for days since minimum category 2) days since minimum category (overhang on minimum, controlling for market return since minimum and days)
oh_max	oh_max	-	Overhang over latest maximum
days_max	days_max	-	Days since the latest maximum (days since maximum)
oh_xu_max_by_days	oh_xu_max	days_max	Market return since maximum for each investment horizon category (market return since maximum, controlling for days since maximum)
oh_max_by_xu_by_days	oh_max	oh_xu_max_by_days, days_max	Overhang on maximum for each 1) market return since maximum, controlling for days since maximum category 2) days since maximum category (overhang on maximum, controlling for market return since maximum and days)

CHAPTER 6

METHODOLOGY

Multiple methods are conducted together in the analysis. In the first level of modeling, for revealing the relationship between the disposition to close a position and the ordinal covariates for overhang with respect to three separate reference points, the methods used are as follows:

- Cox proportional hazards model and extended Cox model
- KM survival curves
- Enhanced Odean's measure

In the second level of modeling, for testing the monotonicity of the coefficients estimated in Cox model and ratios calculated in enhanced Odean's measure, the methods used are as follows:

- Kendall's Tau
- Bayesian regression with monotonic effects

6.1 Cox proportional hazards model and extended Cox model

Cox proportional hazards model is a part of survival analysis and a regression method to model the hazard rate of failure events. Hazard rate is the instantaneous probability that a failure event may happen, such as death from a disease, at a certain time with the condition that the event has not happened until that time. Hazard ratio is the ratio of the hazard rate to the baseline hazard rate, the hazard rate when all covariates have a value of zero. Cox model is especially popular in medical studies. The model is used to understand the relationship between the hazard rates and time-varying covariates, such as readings of medical test or covariates that are fixed for a certain subject across time, such as demographic factors.

Decision of an investor to sell-off a position in a stock is similar to the failure events that Cox model is utilized for in medical studies. For this reason Cox regression has been conducted extensively in the disposition effect literature since Genesove and Mayer (2001)[14]. This study mostly includes numeric variables. Feng and Seasholes (2005) use a parametric Weibull hazard function instead of Cox proportional hazards model. They are intuitive in the sense that they include mostly dichotomous variables and especially numeric variables discretized into multiple dummies.

Shumway and Wu (2005) are the first to discretize the overhang measure into dummies of 1% return steps and model hazard rate with these dummies[15]. The power of this construct comes from its flexibility, since it is possible to estimate different hazard rates across different overhang levels and thus reveal non-linear and complex type of relationships. The method of the mentioned study is the main inspiration of discretization followed here in this thesis.

Since then, many other studies utilized Cox regression incorporating discretization of numeric features, especially overhang, for disposition effect including Barber and Odean (2013), Brettschneider and Burgess (2017), Chhabra and De (2012), Chiyachantana and Yang (2013), Choe and Eom (2009), da Silva and

[14] While Genesove and Mayer (2001) study disposition effect, the term disposition effect is not used throughout the study except for references. Instead, loss aversion is preferred.

[15] Although Feng and Seasholes (2005) use a parametric Weibull hazard function, Shumway and Wu (2005) falsely mention this study to have used the Cox proportional hazards model.

Mendes (2021), da Silva et al. (2022), De et al. (2011), Gemayel and Preda (2018), Loos et al. (2014), Muhl and Talpsepp (2018), Seru et al. (2010), Stoffman (2008), Strahilevitz et al. (2011), Talpsepp (2010, 2011), Talpsepp and Vaarmets (2019), Vaarmets et al. (2019), Wheeler Huttunen (2016), and Zimmermann (2015).

Other studies that utilize Cox regression for disposition effect include Akay (2008), Bansal and Jacob (2022), Bradley, Pantzalis, and Yuan (2016), Chiu, Shu, and Jian (2009), Coval and Shumway (2005), Crane and Hartzell (2010), De Winne (2020), de Groot (2017), Engelberg, Henriksson, and Williams (2018), Frydman and Wang (2020), Haryanto et al. (2020), Heimer (2016), Ivkovich et al. (2007), Ivkovich and Weisbenner (2009), Ivkovich and Weisbenner (2005), Li, Massa, Zhang, and Zhang (2019), Linnainmaa (2011), Richards (2013), and Richards, Rutterford, Kodwani, and Fenton-O'Creevy (2017).

Cox regression[16] is first defined in Cox (1972). A more detailed account is found in Cox and Oakes (1984). More comprehensive narratives of Cox model along with other survival analysis methods are found in Therneau and Grambsch (2000) and Kleinbaum and Klein (2012) among others. "Cox Proportional Hazards" (n.d.) also includes concise information on the method and related formulations.

The basic concepts to understand survival analysis methods are survivor and hazard functions.

Survivor function $S(t)$ shows the probability that a subject survives past a certain time t:

$$S(t) = P(T > t) \tag{2}$$

where T is a random variable of a subject's survival time, and t is a certain value in T.

[16] Cox model and Cox regression are used interchangeably for both Cox proportional hazards model and extended Cox model, the differences of which will be explained here.

$S(0) = 1$, since at the beginning of the study all subjects are alive and $S(\infty) = 0$, since eventually no subjects survive.

The *hazard function* $h(t)$ is the instantaneous rate at which the subject's failure event may occur on the condition that the subject already survived until time t. It is alternatively called as conditional failure rate. So hazard function is negative of the derivative of $S(t)$ with respect to t divided by $S(t)$:

$$h(t) = -\frac{dS(t)/dt}{S(t)} \tag{3}$$

An advantage of survival analysis lies in the usage of censored data. Although there are different types of censoring, the one most relevant for the study is right-censoring. Right-censored subjects are known not to have the event already at time 0, so patients are diagnosed for a disease or an investor opens a position in a stock at time t, similar to uncensored subjects. However, the track of the right-censored record is lost before the observation ends. So it is not known whether the subject has survived after the last recorded time.

In survival analysis, the data of the subjects until the time they are right-censored are incorporated in the model. So the whole data of the right-censored subjects are not omitted, unlike the treatment of those subjects in alternative methodologies. In a disposition effect study, a right-censored observation is a position of an investor in a stock that is still not closed by the end of observed time frame of the data or the last trading day of the stock. Censoring is not done manually, but is a feature of the algorithm in many implementations. The data must be coded so that the time a subject is right-censored should be distinguished for censored subjects apart from the failure time of failed subject, through a status variable[17].

Left-censoring occurs when time of failure is before some determined time, but it is not exactly known. In a disposition effect study, the failure times are the times a position is closed, and they are exactly known from the transaction data. So left-censoring is not relevant for this thesis.

Another feature of survival data is left-truncation. Unlike censoring, a truncated observation is one in which the subject has acquired or developed the event before time 0, but the time is not known. A patient having developed the disease at an unknown time before the time frame of the study or a position of an investor in a stock that has already been opened at an indeterminate time before the time frame of the data should be truncated from the dataset.

The hazard function for a subject i is defined as:

$$h_i(t) = \lambda(t)e^{\mathbf{x}_i \cdot \beta} \quad (4)$$

where

- $\lambda(t)$ is the baseline hazard function for all subjects.
- $e^{\mathbf{x}_i \cdot \beta}$ is the exponentiated dot product of the covariate vector $\mathbf{x}_i$ and the coefficient vector β, or the sum of the linear combination of the covariate vector using the coefficient vector.

[17] In the extended version of the model with time-varying covariates, status variable takes a value of 0 when the position is not closed on a day and 1 when the position is closed on that day. In a right-censored subject, the value of the status variable for the last record is 0 while for an uncensored subject, the last value of the status variable is 1.

Baseline hazard function shows the hazard rate when all covariates take the value of 0. Hence, in this case the exponentiated values are 1. The interpretation of the coefficients is such that for a certain increase of x_i in a covariate and its coefficient β_i, the hazard rate is multiplied by $e^{x_i\beta_i}$.

In the general sense, hazard ratio is the ratio of the hazard rates of two subjects, each distinguished by a set of the covariate values. The ratio of the product of $e^{x_i\beta_i}$ terms for all covariate values for two subjects yields the hazard ratio between them. Or rearranging the terms, the hazard ratio between two subjects 1 and 2 becomes:

$$HR_{1,2} = e^{\beta\cdot(\mathbf{x}_1 - \mathbf{x}_2)} \tag{5}$$

In the simple Cox model with a time-independent covariate vector, proportional hazards assumption should hold. Proportional hazards assumption posits that while baseline hazard is a function of t and as a result varies across time, the vector of coefficients β does not vary across time. Hence, the hazard ratios are proportional across time.

The Cox model is semi-parametric in the sense that the distribution of the outcome or baseline hazard function is not specified, while the vector β of regression parameters for the risk scores are estimated. In contrast, in a parametric survival model, the survival time is assumed to follow a specific distribution, such as Weibull, exponential, generalized gamma, log-normal or log-logistic.

Cox model is considered to be quite robust, since estimates closely approximate the correct parametric model, although Cox model does not specify the baseline function. The absence of the need to specify the baseline hazard function makes Cox model easier to implement. And it prevents the risk of misspecification of the parametric model. Furthermore, the censoring feature as described above makes the Cox model utilize a greater portion of the dataset, as compared to alternative methods such as logistic regression.

The coefficient vector β is estimated by maximum likelihood. Since the formula incorporates only failed subjects and excludes subjects at the time they were censored, the method is called as *partial maximum likelihood*.

In extended Cox model, time-varying covariates are introduced and proportional hazards assumption is relaxed. This is a natural outcome since covariates are allowed to change across time, the hazard rate between two subjects need not be proportional for different points in time. While in the simple model, a single record for each subject is sufficient, in the extended Cox model, the time-varying nature of the covariates necessitates multiple records for different covariate values for each subject. This data format allowing for multiple records per subject and time-varying covariates for survival analysis is known as *counting process* format. In this format, for each subject, each time horizon for which the covariate values are fixed for that subject is a separate record with specified starting and ending times. In the extreme case where covariate values change across each unit of time in data, a separate record for each time unit can be opened. In the case of disposition effect studies, each subject is a position, which is a combination of a unique investor and a unique stock that the investor buys. And for the counting process format, each record shows the closed or not closed status of a position per each time unit, usually a day.

The hazard function specification allowing for time-varying covariates for subject i at time t in an extended Cox model becomes:

$$h_i(t) = \lambda(t) e^{\mathbf{x}_{i,t} \cdot \beta} \tag{6}$$

where,

- $\lambda(t)$ is the baseline hazard function for all subjects.
- $e^{\mathbf{x}_{i,t} \cdot \beta}$ is the exponentiated dot product of the covariate vector $\mathbf{x}_{i,t}$ for a subject i at time t and the coefficient vector β, or the sum of the linear combination of the covariate vector using the coefficient vector.

Coefficient vector β is fixed across subjects and time.

In the extended Cox model, the hazard ratio between two subjects 1 and 2 at time t is formalized as:

$$HR_{1,2}(t) = e^{\beta \cdot (\mathbf{x}_{1,t} - \mathbf{x}_{2,t})} \tag{7}$$

Considering the partial likelihood of the null hypothesis in which all coefficients are zero is denoted by $l(\beta^{(0)})$ and the partial likelihood of the model is $l(\hat{\beta})$, there are three methods to test the significance of a model. In all of them $l(\hat{\beta}) = l(\beta^{(0)}))$ follows a chi-square distribution with degrees of freedom equal to the number of coefficients:

- The likelihood ratio test is $2\Big(l(\hat{\beta}) - l(\beta^{(0)})\Big)$.
- The Wald test is $\left(\hat{\beta} - \beta^{(0)}\right)^{\intercal} I\left(\hat{\beta}\right)\left(\hat{\beta} - \beta^{(0)}\right)$.
- Score (Log-Rank) test is $U\left(\beta^{(0)}\right)^{\intercal} \hat{I}\left(\beta^{0}\right)^{-1} U\left(\beta^{(0)}\right)$.

where,

- $U(\beta)$ is the derivative (gradient) of the log partial likelihood.

- $H(\beta)$ is the second derivative (hessian) of the log partial likelihood.
- $I(\beta) = -H(\beta)$ is the observed information matrix.

The original name coined for the method is Cox proportional hazards model. However, the models that incorporate time-varying covariates and as a result that do not satisfy the proportional hazards assumption should be called as extended Cox model. In this thesis, all covariates are time-varying, so the method used is the extended Cox model version. However, Cox proportional hazards model, Cox model and Cox regression are used interchangeably for extended Cox model.

The implementation of Cox model in the h2o R package will be utilized. h2o R package is an interface to H2O.ai, a scalable open machine learning platform. h2o is intended for leveraging multi-core parallelism and/or cluster computing for many statistics and machine learning algorithms, including the Cox model ("Cox Proportional Hazards", n.d.; "H2O.Ai", n.d.; LeDell et al., 2021). h2o implementation can easily scale up to hundreds of covariates over several million records within a reasonable execution time while the legacy single-threaded implementation in survival R package (Therneau et al., 2021) cannot handle large number of covariates.

In this thesis, Cox models are run by either including ordinal variables as main effects, or including triple-way interactions of three ordinal variables. In the main effects case, the coefficients for each ordinal variable is tested for monotonicity using Kendall's Tau test, as detailed in Section 6.4 (p. 129) . For the triple-way interaction case, the coefficients are first arranged visually across three dimensions, with interpretations of the monotonic patterns. And the monotonicity of the coefficients simultaneously across three ordinal dimensions are further tested with Bayesian regression model with monotonic effects as described in Section 6.5 (p. 131) .

6.2 KM survival curves

KM survival curves are empirical plots of survival probabilities for each ordered failure time (Kleinbaum & Klein, 2012). A typical KM survival curve is a step function that starts at a survival probability of 1 since all subjects are alive at the beginning. And then the curve gradually descends towards 0 in steps since eventually all subjects will either fail or get censored. KM formula involves cumulative product of conditional probability terms. Each term gives the probability of surviving past a certain failure time $t_{(f)}$ on the condition that the subject has survived until that failure time.

KM formula is a product-limit formula since it is limited to the product of the terms up to a certain survival time as shown in Equation (8) (Therneau & Grambsch, 2000):

$$\hat{S}_{KM}(t) = \prod_{i:\ t_i \leq t} \left(1 - \frac{N_i}{Y_i}\right) \tag{8}$$

where

- $\hat{S}_{KM}(t)$ is the KM estimate of the survival probability at time t.
- N_i is the number of observed failures at time i.
- Y_i is the indicator function that shows the number of subjects that have survived until time i.

This formula shows the product of the probability of survival past certain failure time t_i (conditional probability of surviving past t_i given survival to at least t_i) times the probability of survival until certain failure time t_i (or survival past previous failure time t_{i-1}).

In this thesis, KM survival curves are used as a supplementary method to confirm the results of Cox model visually. However, no further testing for equality of curves will be done, due to the multitude of curves yielded through stratification across ordinal levels of multiple categorical variables.

The implementation of KM survival curves in the survival R package is used (Therneau et al., 2021). survminer R package (Kassambara et al., 2021) is also utilized as a wrapper around the functions for calculating KM survival curves in the survival package and visualizing those curves.

6.3 Enhanced Odean's measure

Disposition effect was first mentioned in Schlarbaum et al. (1978) and first analyzed in Shefrin and Statman (1985). However, it was not until Odean (1998) that a specific method for testing the existence and extent of the disposition effect was developed. In the aforementioned study, hypothesis that investors sell stock positions trading at a gain more readily than those trading at a loss is tested.

In Odean (1998), over a specified period, each position, defined as an investor's holdings in a certain stock, is determined whether it is trading at a gain or loss with respect to the purchase price and whether the position is sold by the investor. If the position is closed, it is classified as a realized gain or loss. Otherwise, it is classified as a paper gain or loss. So each position can be classified across four categories: Realized gains, paper gains, realized losses, paper losses.

Two related measures are developed for this purpose:

$$\text{PGR} = \frac{\text{Realized gains}}{\text{Realized gains + Paper gains}} \tag{9}$$

$$\text{PLR} = \frac{\text{Realized losses}}{\text{Realized losses + Paper losses}} \tag{10}$$

The ratio of these two measures is later coined as Odean's measure in subsequent disposition effect studies such as Da Costa, Goulart, Cupertino, Macedo, and Da Silva (2013), Ingersoll and Jin (2013) and De Winne (2020):

$$\text{Odean's measure} = \frac{\text{PGR}}{\text{PLR}} \tag{11}$$

A value above 1 means investors close a greater portion of the positions in gain than the portion of the positions in loss. Hence, they display disposition effect.

While the measure is used commonplace in disposition effect studies mostly due to its ease to calculate, there are shortcomings to this measure. First of all, the measure is an aggregated ratio. Hence, it does not contain information across positions and/or investors. Additionally, the measure does not have the ability to take into account the price path across the period. Hence, it cannot use the temporal information. Furthermore, it is not very suitable to interact the measure with time-varying covariates that may affect the level of the measure, or conduct parametric tests for significance as in methods like Cox regression, logistic regression or ordinary least squares regression. Only a simple t-test can be conducted by calculating the standard error for the difference in PGR and PLR ratios using the binomial distribution's variance formula. Feng and Seasholes (2005) also point at the shortcomings of Odean's approach and compare it with survival methodology.

Despite its shortcomings, some attempts have been made to calculate the ratio across some categorical variables to generate more insight on the interaction of the level of disposition effect with those variables. Odean's ratio is calculated across following dimensions:

- Absolute return bins (Frino, Lepone, & Wright, 2015)
- Genders, investor experience levels, age ranges, wealth bins and geographic regions of the investors' domiciles separately (Tekçe, Yılmaz, & Bildik, 2016)
- Deciles of fund flow (Chiang & Huang, 2017)
- Analyst forecast dispersion and market capitalization quintiles (Balkanska, 2018)
- Trade volume deciles (Bouteska & Regaieg, 2018)
- Expected return intervals (Meng & Weng, 2018)

In this thesis, Odean's measure will be conducted in a more enhanced way, taking into account all days and considering interactions with multiple variables synchronously.

The calculation is done for each set of rows corresponding to a unique combination of the levels of the ordinal variables included in the Cox regression. The ordinal variables are discretized versions of overhang with respect to alternative reference prices, namely, cost, maximum and minimum price references. Section 5.9 (p. 98) details these variables and how they are discretized controlling for holding period and market returns through MUHBOD method. The discretized days variable standing for an ordinal version of days since reference price is attained is also included as a fourth dimension to mimic the baseline realization rate for each time horizon. This time horizon is calculated automatically in Cox model as a feature of the algorithm. For four variables of five ordinal categories each, the whole dataset is partitioned into $5^4 = 625$ separate smaller sets. Each record in the dataset is a day of a position (an investor's holding in a certain stock). Each record has one of the two statuses: Position is closed (realized) or not closed. Original Odean's measure considers the profit/loss state with respect to the reference price as a dichotomous variable. In this enhanced version, a separate percent realized is calculated for each smaller data partition, as the ratio of records with a realized status to the number of total records in the partition. Instead of the basic PGR/PLR ratio for detecting the disposition effect, the monotonicity of the ratios are tested simultaneously across the four dimensions using the Bayesian regression model with monotonic effects as detailed in Section 6.5 (p. 131). An isotonic, or monotonically increasing, pattern of ratios across the ordinal categories of a dimension denotes a disposition effect with respect to the reference price that is the basis for the overhang calculation of that dimension. In that case, for higher categories (which may stand for higher gains, or lower losses), the disposition to close a position is higher. An antitonic, or monotonically decreasing, pattern of ratios across the ordinal categories of a dimension denotes a reverse-disposition effect with respect to the reference price. In the reverse-disposition case, for higher categories, the disposition to close a position is lower.

As a confirmatory measure, a visual arrangement of the ratios are also presented with detailed interpretation of the pattern of ratios across dimensions.

Application of enhanced Odean's measure is done through usual data wrangling packages of R, without a need for any specialized statistical extensions.

6.4 Kendall's Tau

The coefficients for the levels of each ordinal variable in main effects Cox models are tested for monotonicity in the next step. So the rank correlation between the coefficients and the ordinal levels is to be calculated.

In the analysis of ordinal variables, there are two extreme approaches (Agresti, 2010):

- In one approach, the categorical nature of an ordinal variable is completely ignored, and numerical scores are assigned to ordinal levels, as a result treating them as numeric variables.
- In the other extreme approach, only the ordering information about the levels of variables are used.

Methods for calculating the rank correlation coefficient are Gamma, Kendall's Tau, Spearman's Rho and Somers' D (Agresti, 2018). The common features of these methods are as follows:

- They take values between -1 and $+1$, the sign telling the direction of the association, similar to the case of Pearson's correlation.
- When the variables are completely independent, the degree of association is 0.
- The stronger the association is, the larger the absolute value of correlation is.

Of these methods, Spearman's Rho treats the ranks of the values in each variable as integer values and applies the Pearson correlation method on two paired integer variables, albeit with a simplified formulation applicable to integer valued variables. However, Spearman's Rho discards the ordinal nature of the data by assuming that ordinal levels represent equidistant intervals.

Other methods are based on the concordance and the discordance of matched values for all pairs of subjects:

- For n observations or subjects in a dataset of two variables X and Y with matched values, there are a total of $\frac{n(n-1)}{2}$ pairs of subjects.
- A pair of subjects are *concordant* if the subject with the higher (lower) rank in X also has the higher (lower) rank in Y. Number of concordant pairs is denoted by C.
- A pair of subjects are *discordant* if the subject with the higher (lower) rank in X has the lower (higher) rank in Y. Number of discordant pairs is denoted by D.

First defined in Kendall (1938), Kendall's Tau calculates the share of the difference in the number of concordant and discordant pairs in the total number of pairs:

$$\hat{\tau} = \frac{C - D}{n(n-1)/2} \tag{12}$$

For testing the null hypothesis for $\hat{\tau} = 0$, the variance formula is used (Kendall, 1970):

$$var(\hat{\tau}) = \frac{2(2n+5)}{9n(n-1)} \tag{13}$$

A $\hat{\tau}$ value significantly different from 0 does not automatically imply a strong association. $\hat{\tau}$ with an absolute value close to 1 is considered a sign of strong rank correlation.

In this thesis, coefficients for the ordinal levels of a certain overhang variable constitute X, and the rank of those ordinal levels constitute Y. In the case of large number of observations, number of pairs grows with n^2, making the algorithm computationally demanding and in the case of a large portion of ties, pairs of subjects with same values in X and Y variables, the algorithm has to be modified to account for ties. However, in this thesis, the number of cases for each ordinal variable is at most ten and there are no ties. So the basic Kendall's Tau formulation is sufficient.

Kendall's Tau rank correlation coefficient and its p-value are calculated using the cor.test function of the base stats package of R. The function is invoked with the kendall value for the method argument, instead of pearson or spearman alternatives.

6.5 Bayesian regression with monotonic effects

In the Cox models incorporating only the main effects for the ordinal variables, corresponding coefficient values are estimated across the ordinal levels of a single categorical variable. And Kendall's Tau is an easy method to test this univariate monotonicity.

However, in the Cox models incorporating the triple-way interactions of three ordinal variables, a separate coefficient is generated for each triple combination of levels of the ordinal variables. The test for monotonicity requires a multivariate model with the coefficient values as the dependent variables, and ordinal variables as covariates. In the case of the Cox model, the ordinal variables and their interactions are converted into dichotomous dummies taking values of 0 and 1 only and Cox model is a linear model. In the test for monotonicity, the requirement is that the covariates are treated as ordinals and not integer valued variables. While there are isotonic regression methods incorporating numeric valued variables as response and covariates and trying to fit a monotonic step function instead of a line, they do not allow for ordinal covariates.

The only solution for such a combination is provided by the Bayesian regression models from R package brms (Bürkner et al., 2021). The monotonic effects as defined by Bürkner and Charpentier (2020) and included in Bayesian regression models offers a solution, in which ordinal variables are neither included as numeric (as in the case of Spearman's rho) or nominal (as in the case of a regression model where categorical or nominal variables with n unique categories are converted into $n-1$ dummy variables in the model matrix) variables.

brms stands for *Bayesian regression models using Stan* and implements Bayesian models in R using Stan (Bürkner et al., 2021). Stan is a programming language for statistical analysis including full Bayesian statistical inference with Markov chain Monte Carlo (MCMC) methods ("Stan", n.d.). The main use for the language is defining a log density function conditioned on data ("Stan", n.d.).

Bayesian data analysis is basically counting the number of ways a given data can happen, according to assumptions (McElreath, 2020). Probability theory is used as a general method to represent plausibility. Bayesian analysis has two foundational ideas (Kruschke, 2015):

- Bayesian inference is the reallocation of plausibility or credibility across possible alternatives.
- The possible alternatives or possibilities over which plausibility is allocated are parameter values in mathematical models.

Main tool for Bayesian analysis is Bayesian updating (McElreath, 2020):

- A Bayesian model is initiated with a set of alternative possibilities and assigned plausibilities to each, known as *prior plausibilities*.
- Then these priors are updated in the light of data, to produce posterior plausibilities.

Parameters are ways of indexing possible explanations of given data. They are unobserved variables to be inferred and can be thought of as unknown variables in a mathematical formula. *Likelihood* is the relative number of ways that a parameter value can generate the data at hand. For every unique combination of data, likelihood, parameters and prior, a unique posterior distribution is generated. This posterior includes the relative plausibility of possible parameter values, conditional on the given data and the model used. However, instead of computing or approximating the posterior distribution directly, MCMC methods sample from the posterior. At the end of a sampling chain, various parameter values are collected, frequencies of which are interpreted as posterior plausibilities.

Several algorithms to implement MCMC exist, including Metropolis-Hastings updates and Gibss-Sampling (Bürkner, 2017). However, these algorithms are slow to converge[18]. Stan implements Hamiltonian Monte Carlo algorithm and its extension the No-U-Turn-Sampler, with the advantage of faster convergence.

Monotonic effects feature is proposed by Bürkner and Charpentier (2020). In monotonic effects, ordinal discrete predictors can be handled (Bürkner, 2022). The categories for such a predictor are allowed not to be equidistant with respect to their effect on the response. The distance between adjacent categories is estimated from data and may vary. There are two parameters describing this effect, b and ζ:

- b represents the size and direction of the monotonic effect and is interpreted as the expected average difference between two adjacent categories.
- ζ is a vector of estimated normalized distances between adjacent categories, summing up to 1.

[18] Convergence means that chains with different starting positions converge around the same region of high probability in the posterior distribution (McElreath, 2020).

For the case of a single monotonic predictor x, the linear predictor term of an observation n is as follows:

$$\eta_n = bD \sum_{i=1}^{x_n} \zeta_i \tag{14}$$

D is the length of ζ and is equal to number of categories of the ordinal predictor minus 1. Since b is the expected average distance between adjacent categories, bD is the expected total distance between the largest and smallest ordinal categories. $\sum_{i=1}^{x_n} \zeta_i$ shows the normalized distance as a percent from the smallest category up to the category that observation n is a member of.

b can have any real value. The simplex parameter is a vector of positive percent values between 0 and 1 and summing up to 1, thus having the properties $\zeta_i \in [0, 1]$ and $\sum_{i=1}^{D} \zeta_i = 1$.

The default prior for ζ is the equality of all distances. However, a Dirichlet prior with parameter α, the canonical distribution for simplex parameters, can also be used. In this thesis, equal distant default prior is used as the starting point.

In monotonic effects models, one or more ordinal predictors can be used with continuous predictors, and the interactions between any of them can also be included. The functional form of such an example model is as follows:

$$\eta_n = b_0 + b_1 z_n + b_2 mo(x_n, \zeta_2) + b_3 z_n mo(x_n, \zeta_3) \tag{15}$$

where x is a monotonic predictor, z is a continuous predictor and ζ_2 and ζ_3 are two simplex parameters for the main and the interaction terms of x.

In this thesis, two separate kinds of models are run using monotonic effects. A simpler model only includes three monotonic predictors x, y and z as main terms:

$$\eta_n = b_0 + b_1 mo(x_n, \zeta_1) + b_2 mo(y_n, \zeta_2) + b_3 mo(z_n, \zeta_3) \tag{16}$$

A more complex model also adds the two two-way interactions between these main terms, one between x and y, the other between x and z as such:

$$\begin{aligned} \eta_n =& b_0 + b_1 mo(x_n, \zeta_1) + b_2 mo(y_n, \zeta_2) + b_3 mo(z_n, \zeta_3) + \\ & b_4 mo(x_n, \zeta_4) mo(y_n, \zeta_5) + b_5 mo(x_n, \zeta_6) mo(z_n, \zeta_7) \end{aligned} \tag{17}$$

In brms, every parameter is summarized using the mean estimate, standard deviation of the posterior distribution and two-sided 95% credible intervals based on quantiles (Bürkner, 2017). Apart from the significance of the estimate, convergence of the algorithm should also be checked. Rhat ($\hat{R}$) is the Gelman-Rubin convergence diagnostic and should be at 1.00 for a fully converged chain (McElreath, 2020). Stan documentation recommends using the results when $\hat{R}$ is less than 1.05 (Guo, Gabry, Goodrich, & Weber, n.d.).

In this thesis, monotonic effects models are conducted in order to test the monotonicity of the estimates from two models:

- The coefficients of the Cox model with triple-way interaction terms among the levels of three ordinal variables representing overhang with respect to alternative reference points.

- The enhanced Odean's measures across the levels of three ordinal variables representing overhang with respect to alternative reference points and also the levels of a fourth ordinal variable for the holding period horizons.

There are a total of 124 coefficients in the Cox model with interactions across five category ordinals, while the number of coefficients is 998 for the ten category version. There are a total of 625 Odean's measures for the five category version. In line with the main philosophy of Bayesian analysis, which considers the whole distribution rather than point estimates, a random sampling approach is introduced for the coefficients and ratios. Each coefficient in the Cox model has a mean estimate and also a standard error, as with case of all similar regression approaches. A predetermined number of values are sampled from each of these coefficients' distribution, using their mean estimates and standard errors and assuming a normal distribution. rnorm function from base R package stats is used for this purpose.

Accordingly, each ratio in the Odean's measure is the total number of records with a status of 1, meaning a position is closed on that day, divided by the total number of records having a status of either 1 or 0, standing for positions not closed on that day. A predetermined number of values are sampled from a binomial distribution with a probability of success equal to the ratio and number or trials equal to the total number of records corresponding to a combination of levels for each of the four ordinal variables included. rbinom function from base R package stats is used for this purpose. The sampled values are number of successes, so they are divided by total number of records to arrive at measures as ratios.

So a synthetic dataset that conforms with the distributions of the original coefficients or ratios is generated for each monotonic effects model. The predicted response variable consists of the sampled values for coefficients or ratios, while the predictors are the ordinal variables included in the original model generating the coefficients or ratios. The ordinal variables are included as monotonic effects as described above.

CHAPTER 7

MODELS AND RESULTS

A total of 13 Cox regression models are run in this thesis. The variables used in the models, number of predictors as degrees of freedom and p-values of significance tests are listed in Table 6. All models are significant at 5% level across all three test types.

Table 6. Cox Models Run in the Study

Model Name	Main Effect Variables	Multiway Interacting Variables	Log Likelihood Test DF	Log Likelihood Test P-Value	Score Test P-Value	Wald Test P-Value
cost10	oh_cost_10		9	0	0	0
cost_day_xu_10_only	oh_cost_by_xu_by_days_10		9	0	0	0
max10	oh_max_10		9	0	0	0
max_xu_day10	oh_max_by_xu_by_days_10		9	0	0	0
min10	oh_min_10		9	0	0	0
min_xu_day10	oh_min_by_xu_by_days_10		9	0	0	0
cost_max_min_5	oh_cost_by_xu_by_days_05 oh_max_by_xu_by_days_05 oh_min_by_xu_by_days_05		12	0	0	0
cost_max_min_ia_only_5		oh_cost_by_xu_by_days_05 oh_max_by_xu_by_days_05 oh_min_by_xu_by_days_05	124	0	0	0
cost_max_min_10	oh_cost_by_xu_by_days_10 oh_max_by_xu_by_days_10 oh_min_by_xu_by_days_10		27	0	0	0
cost_max_min_ia_only_10		oh_cost_by_xu_by_days_10 oh_max_by_xu_by_days_10 oh_min_by_xu_by_days_10	998	0	0	0
cost_max_min_simple_5	oh_cost_05 oh_max_05 oh_min_05		12	0	0	0
cost_max_min_simple_ia_only_5		oh_cost_05 oh_max_05 oh_min_05	99	0	0	0
cost_max_min_simple_10	oh_cost_10 oh_max_10 oh_min_10		27	0	0	0

7.1 Univariate models

Six simple models include measures for a single type of reference price each. Two simple models use cost based reference measures. Next two simple models use maximum based reference measures, and remaining two simple models use minimum based reference measures.

In the models, reference level for each ordinal variable is automatically excluded from the model matrix in order to prevent linear dependence. These levels are deemed to represent the baseline in which the models assume the coefficients to be zero. For the case of cost reference based ordinal variables, the reference level is the middle category. For maximum price reference based ordinal variables, the reference level is the highest category. And for minimum price reference based ordinal variables, the reference level is the lowest category. In the below analyses, reference levels automatically omitted in the univariate models are added back with zero values in order to test for monotonicity.

When coefficients of ten category ordinal variables for overhang on cost based reference points are plotted across categories in Figure 23 and presented in Table 7, all coefficients of simple oh_cost_10 variable, and all but one coefficient of oh_cost_by_xu_by_days_10 variable are significant at 5% level. However, none of the variables demonstrate a significant or strong monotonic pattern in the univariate setting.

Table 7. Ordinal Levels, Coefficients and P-Values of Ordinal Variables in Univariate Models With Ten Category Cost Based Overhang Variables

Variable	Ordinal Level	Coefficient	P-Value	Significance
oh_cost_10	0.1	0.06	0.00	Significant
oh_cost_10	0.2	0.13	0.00	Significant
oh_cost_10	0.3	0.24	0.00	Significant
oh_cost_10	0.4	0.35	0.00	Significant
oh_cost_10	0.5	0.29	0.00	Significant
oh_cost_10	0.7	0.15	0.00	Significant
oh_cost_10	0.8	0.25	0.00	Significant
oh_cost_10	0.9	0.32	0.00	Significant
oh_cost_10	1.0	0.31	0.00	Significant
oh_cost_by_xu_by_days_10	0.1	0.66	0.00	Significant
oh_cost_by_xu_by_days_10	0.2	0.36	0.00	Significant
oh_cost_by_xu_by_days_10	0.3	0.10	0.00	Significant
oh_cost_by_xu_by_days_10	0.4	0.04	0.00	Significant
oh_cost_by_xu_by_days_10	0.5	0.01	0.19	Insignificant
oh_cost_by_xu_by_days_10	0.7	0.03	0.00	Significant
oh_cost_by_xu_by_days_10	0.8	0.06	0.00	Significant
oh_cost_by_xu_by_days_10	0.9	0.14	0.00	Significant
oh_cost_by_xu_by_days_10	1.0	0.20	0.00	Significant

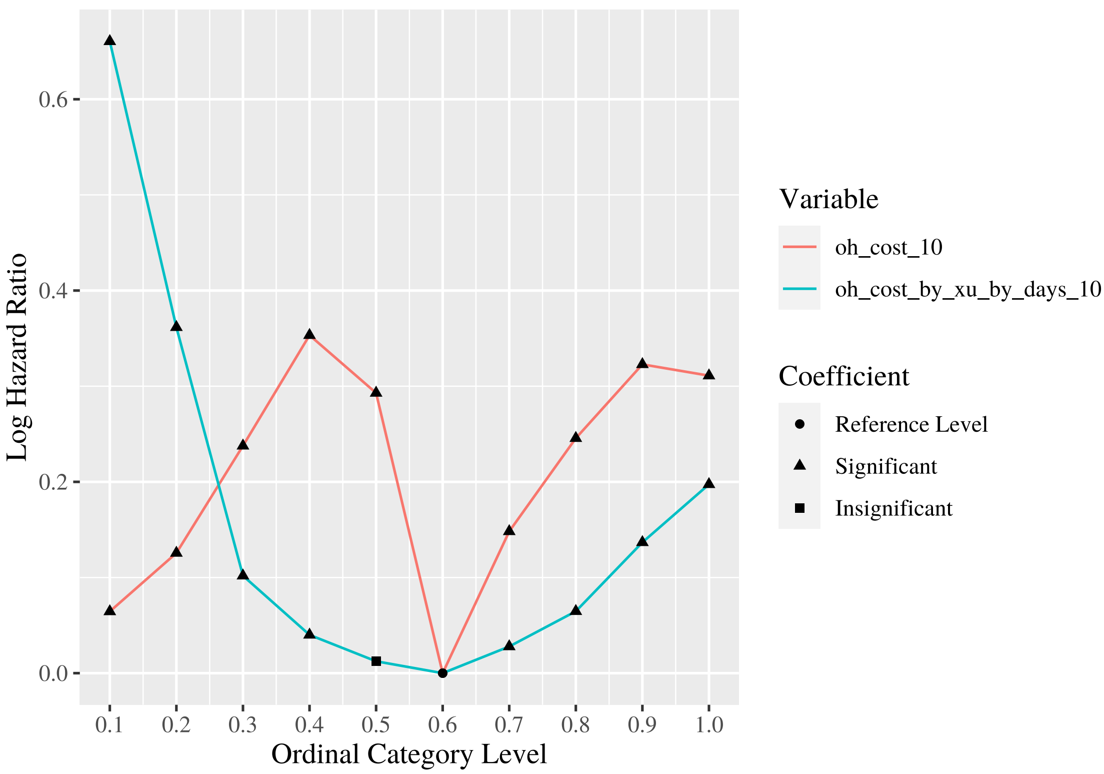

Figure 23. Coefficients of ten category cost based overhang variables

Kendall's Tau values are also between -0.16 and 0.38 and none of them are significant as shown in Table 8.

For the case of simple models incorporating maximum price based overhang measures, all coefficients are significant at 5% level, as shown in Table 9. oh_max_by_xu_by_days_10 variable, which is controlled for both days and market returns since last maximum price is attained, is perfectly monotonic and decreasing, showing a reverse disposition effect as shown in Figure 24 and Table 10. The Tau value of oh_max_10 is insignificant and the level of Tau value does not show a strong monotonicity, as confirmed by the sharp decrease between the coefficient of 0.9 level and 0 coefficient of the reference level at 1.0.

Table 8. Kendall's Tau Values and Corresponding P-Values for Coefficients of Ten Category Cost Based Overhang Variables

Variables	Tau	P-value
oh_cost_by_xu_by_days_10	-0.16	0.60
oh_cost_10	0.38	0.16

Table 9. Ordinal Levels, Coefficients and P-Values of Ordinal Variables in Univariate Models With Ten Category Maximum Price Based Overhang Variables

Variable	Ordinal Level	Coefficient	P-Value	Significance
oh_max_10	0.1	1.34	0	Significant
oh_max_10	0.2	1.38	0	Significant
oh_max_10	0.3	1.48	0	Significant
oh_max_10	0.4	1.59	0	Significant
oh_max_10	0.5	1.67	0	Significant
oh_max_10	0.6	1.74	0	Significant
oh_max_10	0.7	1.80	0	Significant
oh_max_10	0.8	1.74	0	Significant
oh_max_10	0.9	1.67	0	Significant
oh_max_by_xu_by_days_10	0.1	2.44	0	Significant
oh_max_by_xu_by_days_10	0.2	1.70	0	Significant
oh_max_by_xu_by_days_10	0.3	1.59	0	Significant
oh_max_by_xu_by_days_10	0.4	1.49	0	Significant
oh_max_by_xu_by_days_10	0.5	1.31	0	Significant
oh_max_by_xu_by_days_10	0.6	1.18	0	Significant
oh_max_by_xu_by_days_10	0.7	1.09	0	Significant
oh_max_by_xu_by_days_10	0.8	0.87	0	Significant
oh_max_by_xu_by_days_10	0.9	0.75	0	Significant

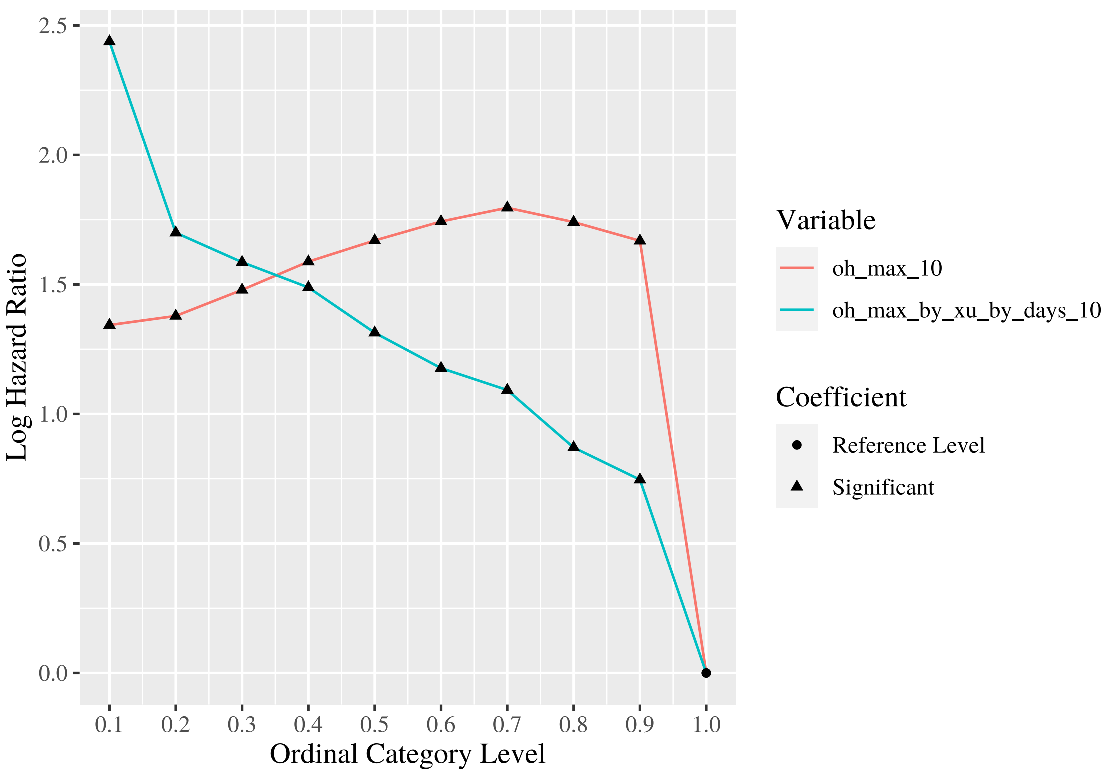

Figure 24. Coefficients of ten category maximum price based overhang variables

For the case of simple models incorporating minimum price based overhang measures, all coefficients are significant at 5% level as shown in Table 11. Coefficients of each of the two variables present V-shapes as shown in Figure 25. According to Table 12, oh_min_by_xu_by_days_10 variable's Tau value is insignificant, while oh_min_10 variable's Tau value is significant. However, the rank correlation coefficient demonstrates only a mildly strong positive monotonicity, as also confirmed by the sharp decrease between the 0 value of the reference level at 0.1 and the coefficient of the next level.

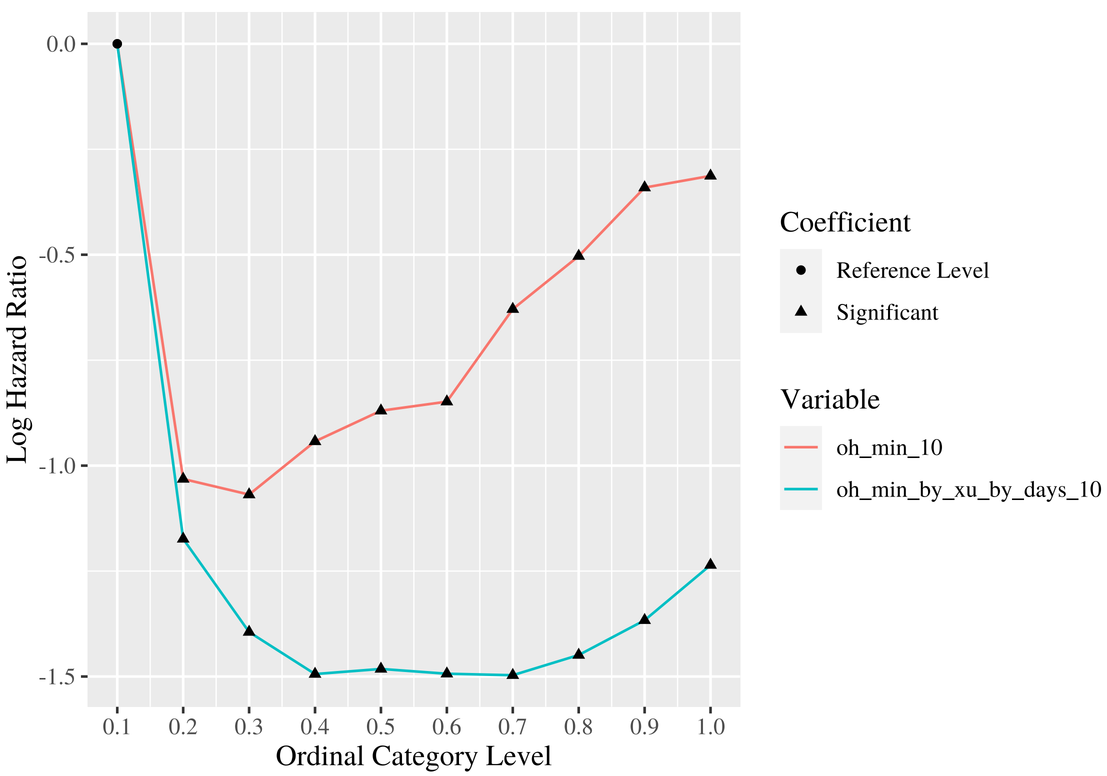

Figure 25. Coefficients of ten category minimum price based overhang variables

Table 10. Kendall's Tau Values and Corresponding P-Values for Coefficients of Ten Category Maximum Price Based Overhang Variables

Variables	Tau	P-value
oh_max_by_xu_by_days_10	-1.00	0.00
oh_max_10	0.33	0.22

Table 11. Ordinal Levels, Coefficients and P-Values of Ordinal Variables in Univariate Models With Ten Category Minimum Price Based Overhang Variables

Variable	Ordinal Level	Coefficient	P-Value	Significance
oh_min_10	0.2	-1.03	0	Significant
oh_min_10	0.3	-1.07	0	Significant
oh_min_10	0.4	-0.94	0	Significant
oh_min_10	0.5	-0.87	0	Significant
oh_min_10	0.6	-0.85	0	Significant
oh_min_10	0.7	-0.63	0	Significant
oh_min_10	0.8	-0.50	0	Significant
oh_min_10	0.9	-0.34	0	Significant
oh_min_10	1.0	-0.31	0	Significant
oh_min_by_xu_by_days_10	0.2	-1.17	0	Significant
oh_min_by_xu_by_days_10	0.3	-1.39	0	Significant
oh_min_by_xu_by_days_10	0.4	-1.49	0	Significant
oh_min_by_xu_by_days_10	0.5	-1.48	0	Significant
oh_min_by_xu_by_days_10	0.6	-1.49	0	Significant
oh_min_by_xu_by_days_10	0.7	-1.50	0	Significant
oh_min_by_xu_by_days_10	0.8	-1.45	0	Significant
oh_min_by_xu_by_days_10	0.9	-1.37	0	Significant
oh_min_by_xu_by_days_10	1.0	-1.24	0	Significant

For summary, in simple models with only a single reference based overhang measure, only maximum reference based variables show a significant decreasing monotonicity in line with reverse disposition effect.

Table 12. Kendall's Tau Values and Corresponding P-Values for Coefficients of Ten Category Minimum Price Based Overhang Variables

Variables	Tau	P-value
oh_min_by_xu_by_days_10	-0.16	0.60
oh_min_10	0.56	0.03

7.2 Multivariate models

In multivariate models, multiple ordinal variables representing the overhang measures with respect to three different reference points, namely cost, maximum price, minimum price, are included. Most models are tested in four versions:

- Only main effects model with five category ordinal variables
- Only triple-way interactions model with five category ordinal variables
- Only main effects model with ten category ordinal variables
- Only triple-way interactions model with ten category ordinal variables.

The model results are presented in groups including all these versions. Since the log likelihood test, score test and Wald test p-values for all Cox models are already reported in Table 6 (p. 139), they are not repeated here. Model level p-values for all models are significant. However, the focus here is to test the monotonicity of coefficients to indicate disposition or reverse disposition effects with respect to separate reference points. In the main effects models, the ordinal variables of overhang measures for multiple reference points are included as is, without interactions. There are three different reference points (cost, maximum and minimum price attained) and each ordinal level is a distinct dichotomous variable. And for each group of ordinal variables, one level is left out from at model as the reference level, as explained earlier. Hence, there are 15 dichotomous variables (12 excluding the three reference levels) in the five ordinal category models and 30 dichotomous variables (27 excluding the three reference levels) in the ten ordinal category models. In each group, the following methods and tests are included for the main effects models:

- Cox proportional hazards model. The coefficients and their p-values are reported. The coefficients of reference ordinal levels are included as 0.

- Kendall's Tau for the univariate monotonicity of the coefficients of each ordinal variable
- Odean's ratios[19] (percent of records closed where each record is unique for a position and day combination) where records are grouped across the unique values of ordinal variables, including also the ordinal level of days as a fourth variable. Since there are 625 combinations of the four ordinal variables with five categories each, separate ratios are reported in Appendix B (p. 219). They are included in a Bayesian regression model with monotonic effects. Instead of the original ratios, a random sample from binomial distribution is generated using the ratios and the size of samples. The direction and significance of the monotonicity of the ratios are reported and visualized.
- KM curves for the survival rate of the positions across days for each combination of the three ordinal variables. This method provides a visual supplement to confirm the findings from other methods.

In the triple-way interactions model, the grid of the unique combinations of each level of three ordinal variables are included. In this model, the interaction variable for a certain combination of ordinal values takes a value of 1 when all three measure's corresponding dichotomous variable for the selected level are 1. Otherwise, the variable takes a value of 0. The original dichotomous variables are not included as

[19] Odean's ratios and enhanced Odean's measures are used interchangeably.

main effects. In the five category version there are 124 separate dichotomous variables and in the ten category version there are 998[20] separate dichotomous variables. In each group, the following methods and tests are included for the triple-way interactions models:

- Cox proportional hazards model. Due to the very high number of coefficients, the percent of coefficients that are significant are reported in main text. Whole set of coefficients are reported in Appendix C (p. 223), Appendix D (p. 224) and Appendix E (p. 230).
- Bayesian regression model with monotonic effects where the dependent variable consists of the coefficients from the Cox model, and the dependent variables are the levels of the three original ordinal variables that make up the triple-way interactions. In Bayesian regression models, instead of the point estimates of the coefficients, a random sample is generated using the point estimate and the standard error of each coefficient. In this model, for each of the ordinal variables, the direction and the difference between the dependent variable's values corresponding to the highest and lowest ordinal categories are tested for significance. The importance of this model is that, since the monotonicity is tested, the steps between consecutive ordinal levels are not fixed but are in the same direction.
- A second version of the Bayesian regression model where the interactions between the levels of ordinal variables are also added. There are three possible two-way interactions among three ordinal variables. However, only the interactions between the ordinal levels of cost reference based overhang and that of the maximum

[20] The number of all triple-way combinations for three variables of ten categories each is 1000. Taking out one of the dummies to prevent linear dependence results in 999 variables. However the model results report 998 coefficients, suggesting that no cases appear for one of the remaining 999 combinations and the fixed 0 valued dummy variable for that combination is also automatically excluded.

reference based overhang, and between the cost reference based overhang and minimum reference based overhang are included. This model tries to test whether approaching the maximum or minimum reference prices changes the strength of the cost reference price's relationship with the propensity to sell.

7.2.1 Overhang controlling for market return and holding period

In this main model set, all three ordinal variables of overhang measures with respect to cost, maximum and minimum prices are calculated using the MUHBOD method developed in this thesis. Hence, the overhang variables are controlled for holding period and market return in a nested manner:

- Holding periods since date of the according reference price are separately discretized into equal sized bins. For the cost reference, the date is the start of the position. For the maximum or minimum price references, it is the date last maximum or minimum price is attained.
- Market returns since the dates of the reference prices are separately discretized into equal sized bins for each category of the corresponding holding period ordinal variables.
- Overhang with respect to the reference prices are separately discretized into equal sized bins for each category combination of the corresponding holding period ordinal variables and market return ordinal variables.

Hence, the ordinal category of overhang measures are controlled for both holding return and market return separately for cost, maximum and minimum reference prices. The details of MUHBOD method are given in Subsection 5.9.2 (p. 105).

The coefficient of reference level for each ordinal variable, automatically omitted from the model to prevent linear dependence is added to the coefficients table with 0 value representing the baseline case.

In the five category simple case with no triple interactions (cost_max_min_5 model), there are four tested categories for each of the three variables, leaving one category as the reference and baseline level, and all coefficients are significant at 0.05 level as shown in Table 13. When the coefficients of the categorical levels of three variables are plotted together in Figure 26, we see the following:

- The coefficients with respect to the levels of overhang on cost reference are almost monotonically increasing, in line with disposition effect.
- The coefficients with respect to the levels of overhang on maximum reference are perfectly monotonically decreasing, in line with reverse disposition effect.
- The coefficients with respect to the levels of overhang on minimum reference are almost monotonically decreasing, in line with reverse disposition effect.
- The reference level coefficients that are automatically added are in line with the monotonicity of coefficients of respective variables.

Table 13. Ordinal Levels, Coefficients and P-Values (in Parentheses) of Ordinal Variables in cost_max_min_5 Model

Level	Variables		
	oh_cost_by_xu_by_days_05	oh_max_by_xu_by_days_05	oh_min_by_xu_by_days_05
0.2	-0.08 (0)	1.91 (0)	0 (-)
0.4	-0.19 (0)	1.55 (0)	-1.07 (0)
0.6	0 (-)	1.32 (0)	-1.21 (0)
0.8	0.45 (0)	0.98 (0)	-1.31 (0)
1.0	0.88 (0)	0 (-)	-1.22 (0)

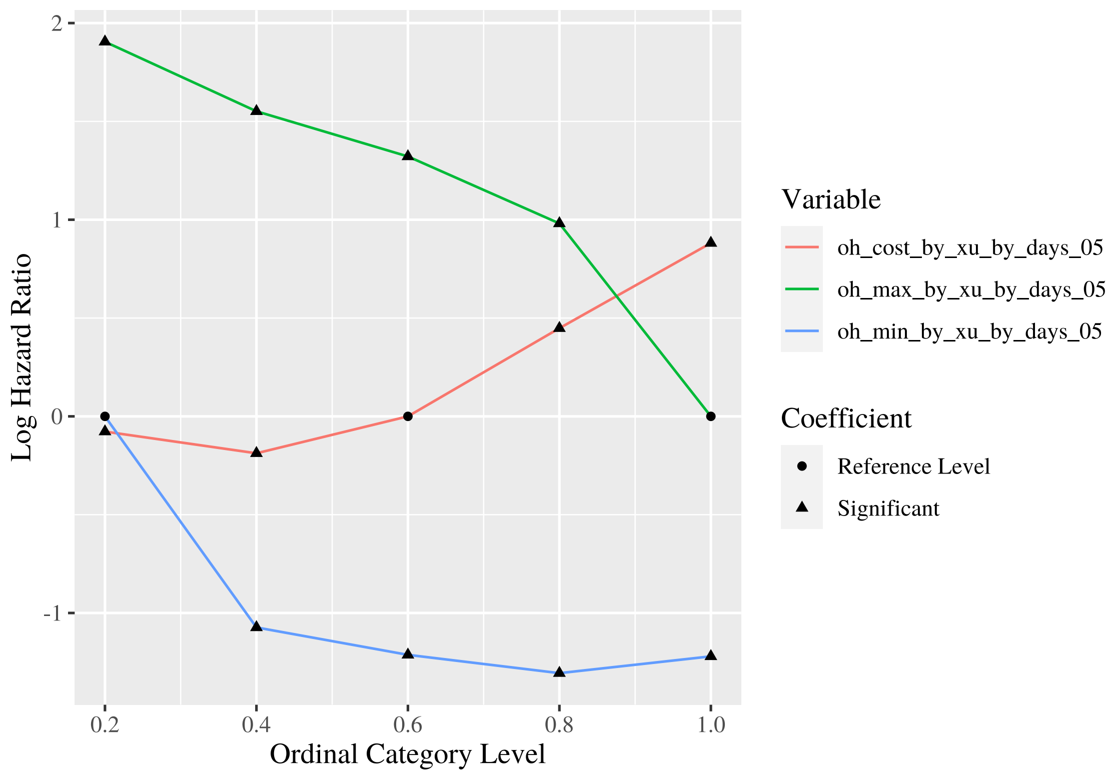

Figure 26. Coefficients of cost_max_min_5 model with corresponding significance status

The level of monotonicity of the coefficients of each variable are tested with Kendall's Tau as presented in Table 14. The table reveals the following:

- Coefficients of overhang on cost reference has a Tau value of 0.8, meaning near perfect positive rank correlation. And Tau value is significant at 0.1 level.
- Coefficients of overhang on maximum reference has a Tau value of -1, meaning perfect negative rank correlation. And Tau value is significant at 0.05 level.

Table 14. Kendall's Tau Values and Corresponding P-Values for Coefficients of cost_max_min_5 Model

Variables	Tau	P-value
oh_max_by_xu_by_days_05	-1.0	0.02
oh_min_by_xu_by_days_05	-0.8	0.08
oh_cost_by_xu_by_days_05	0.8	0.08

• Coefficients of overhang on minimum reference has a Tau value of -0.8, meaning near perfect negative rank correlation. And Tau value is significant at 0.1 level.

The same analyses are repeated for the model with ten categories (cost_max_min_10) and the results are confirmed with higher significance. As shown in Table 15, all the coefficients are significant at 0.05 level. Figure 27 reveals that maximum and minimum based coefficients are perfectly monotonically decreasing while cost based coefficients are almost perfectly monotonically increasing.

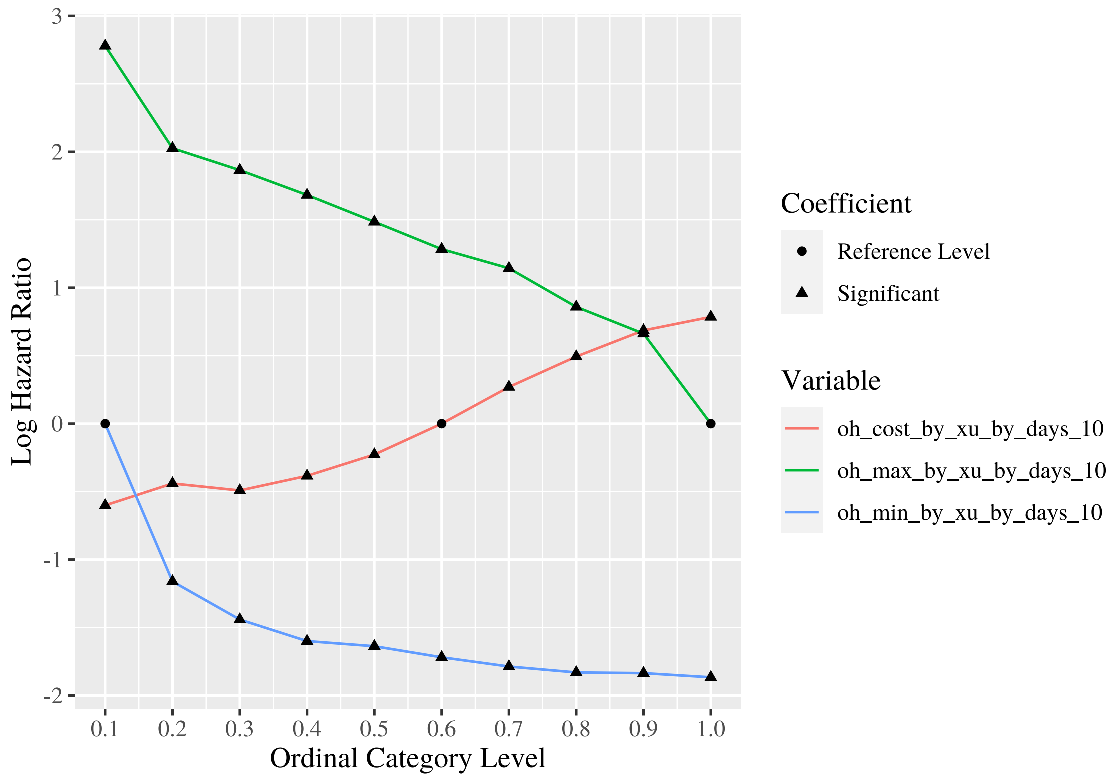

Figure 27. Coefficients of cost_max_min_10 model with corresponding significance status

Kendall's Tau values of the coefficients also show significant monotonicity according to Table 16, and all three Tau values are significant at 0.05 level. When the

Table 15. Ordinal Levels, Coefficients and P-Values (in Parentheses) of Ordinal Variables in cost_max_min_10 Model

Level	Variables		
	oh_cost_by_xu_by_days_10	oh_max_by_xu_by_days_10	oh_min_by_xu_by_days_10
0.1	-0.6 (0)	2.78 (0)	0 (-)
0.2	-0.44 (0)	2.03 (0)	-1.16 (0)
0.3	-0.49 (0)	1.87 (0)	-1.44 (0)
0.4	-0.38 (0)	1.68 (0)	-1.6 (0)
0.5	-0.23 (0)	1.48 (0)	-1.64 (0)
0.6	0 (-)	1.28 (0)	-1.72 (0)
0.7	0.27 (0)	1.14 (0)	-1.79 (0)
0.8	0.49 (0)	0.86 (0)	-1.83 (0)
0.9	0.69 (0)	0.66 (0)	-1.84 (0)
1.0	0.79 (0)	0 (-)	-1.87 (0)

Table 16. Kendall's Tau Values and Corresponding P-Values for Coefficients of cost_max_min_10 Model

Variables	Tau	P-value
oh_max_by_xu_by_days_10	-1.00	0
oh_min_by_xu_by_days_10	-1.00	0
oh_cost_by_xu_by_days_10	0.96	0

three overhang variables are included as separate features in Cox models, the results reveal that investors exhibit the following:

- Disposition effect with respect to overhang on cost reference price. Hence, investors have monotonically increasing propensity to close a position with higher overhang categories above the reference level. And they have monotonically decreasing propensity to close a position with lower overhang categories below the reference level.
- Reverse disposition effect with respect to overhang on maximum reference price. Hence, when the highest category (closest to the maximum price) is taken as the reference level, investors have monotonically increasing propensity to close the position with lower overhang categories below the reference level.
- Reverse disposition effect with respect to overhang on minimum reference price. Hence, when the lowest category (closest to the minimum price) is taken as the reference level, investors have monotonically decreasing propensity to close the position with higher overhang categories above the reference level.

However, these effects should also be tested with models and methods incorporating all three variables in interaction. Before the model results, below figures confirm the above findings in a triple-way interaction, setting while also revealing more complex dynamics beneath.

Figure 28 shows the KM curves of the positions across days since the positions are opened, stratified across three categorical overhang variables. The stratification is done using the variables from the cost_max_min_5 model with five categories and with no triple-interactions among variables. However, the process of stratification across three variables can be interpreted as a triple-way-interaction in itself.

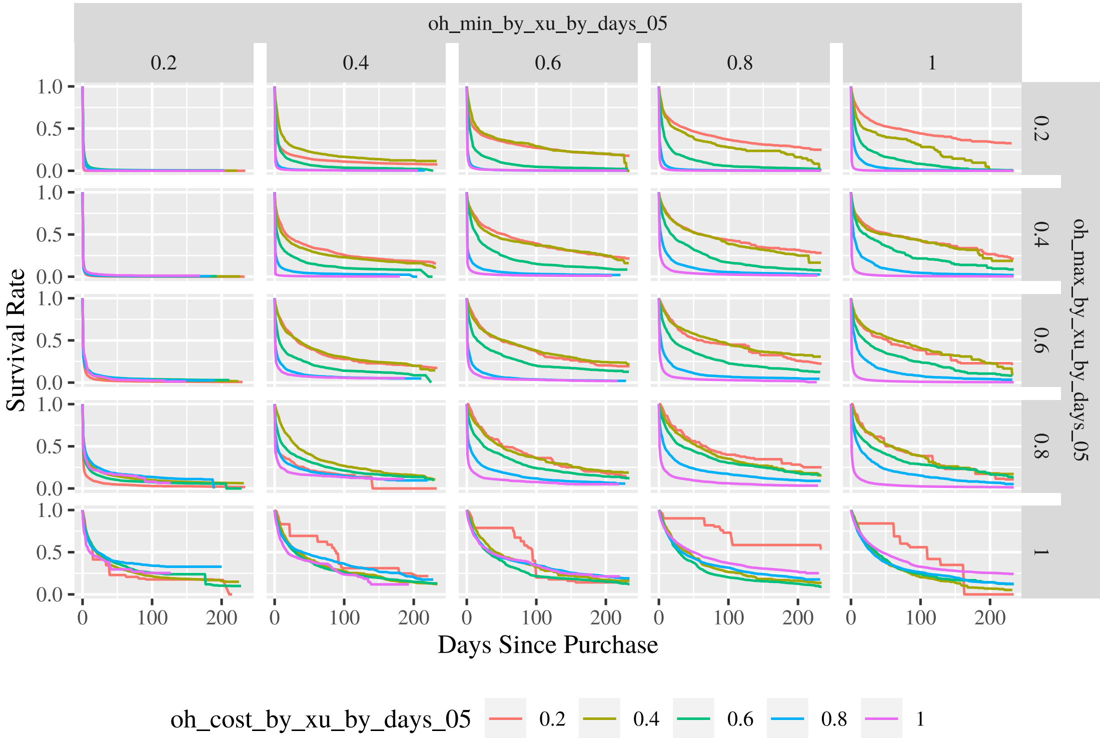

Figure 28. KM curves of cost_max_min_5 model

The layout of the whole figure is as follows:

- Each curve shows the survival rate of the positions across days. They start with 1, meaning all positions are still open by the end of the first day after they are first opened, and they approach 0, meaning none of the positions are left open as days unfold.
- In each facet, each stacked colored line corresponds to a categorical level of the overhang on cost reference price
- Each column of facets, from left to right, shows increasing categories of the overhang on minimum reference price. The leftmost column corresponds to the reference level. Hence, this is the case when current price is closest to the minimum price. The rightmost column shows the case when current price is furthest away from the minimum price.

- Each row of facets, from bottom to top, shows decreasing categories of the overhang on maximum reference price. The lowest row corresponds to the reference level. Hence, this is the case when current price is closest to the maximum price. The topmost row shows the case when current price is furthest away from the maximum price.

The top-right facet is the case when the current price is furthest away from both the maximum and minimum reference prices and is taken as the starting point in the interpretation of this complex figure. When a survival curve is higher, the hazard rate is lower, meaning a smaller portion of the position have been closed, so more positions have survived and vice versa. The findings are as follows:

- When the current price is furthest away from both the maximum and minimum reference prices, the survival curve for the lowest category of overhang on cost reference is the highest one. Hence, propensity to close a position is lower. The curves are monotonically and parallelly lower for higher overhang categories. Hence, propensity to close a position gets stronger. This is in line with the disposition effect.
- Ceteris paribus, going from the top (furthest from maximum reference price) to the bottom (closest to maximum reference price) rows, the survival curves are positioned higher. Hence, as the current price approaches above to maximum reference price, propensity to close a position is weaker, in line with reverse disposition effect.

- Ceteris paribus, going from the rightmost (furthest from minimum reference price) rows to the leftmost (closest to minimum reference price) columns, the survival curves are lower. Hence, as the current price approaches below to minimum reference price, propensity to close a position is stronger, in line with reverse disposition effect.

These findings confirm the those from the coefficients of the simple models with only main effects.

There are also secondary relationships as interactions of overhang on cost reference with the overhang on maximum and minimum references:

- In each row, in facets from right to left, as current prices get closer below to the minimum price, the curves for each overhang on cost category get closer and are even partially reversed. Hence, as price approaches the minimum price, the dominance of the disposition effect coming from the overhang on cost reference price weakens, and reverse disposition effect coming from the overhang on minimum reference price starts to dominate.
- In each column, in facets from top to bottom, as current prices get closer above to the maximum price, the curves for each overhang on cost category get closer and are even partially reversed. Hence, as price approaches the maximum price, the dominance of the disposition effect coming from the overhang on cost reference price weakens, and reverse disposition effect coming from the overhang on maximum reference price starts to dominate.

When the top-right and bottom-left facets are compared, these secondary effects can be understood better: In the bottom-left facet, the curves are closer to each other and the order is mostly reversed as compared to the top-right facet.

For KM curves, there are no further tests for the significance of the direction of these effects.

Further visual methods are also explored. The distribution of the p-values of the coefficients from the five and ten category Cox models with triple-way interactions according to their significance levels are presented in Table 17. In the five category model, 89% of the 124 coefficients are significant at 0.05 or 0.1 levels while in the ten category model, 79% of the 998 coefficients are significant. The coefficients of triple-way interaction variables are presented in detail in Appendix C (p. 223) and Appendix D (p. 224). The coefficients are also organized into multi-faceted plots with a layout similar to the KM plots in Figure 29 and Figure 30. For the ten category version, only the coefficients that correspond to categories with at least 1,000 records in the data are included, since coefficients for smaller data sizes, mostly for the extreme categories, may not be representative for the monotonicity of the effects and may blur the overall picture.

Table 17. Total Numbers and Significance Level Distributions of Triple-Way Interaction Coefficients in cost_max_min_ia_only_5 and cost_max_min_ia_only_10 Models

Model Name	Number of Coefficients	Percent of Total Variables		
		< 0.05	0.05 - 0.1	Insignificant
cost_max_min_ia_only_5	124	0.84	0.05	0.11
cost_max_min_ia_only_10	998	0.74	0.05	0.21

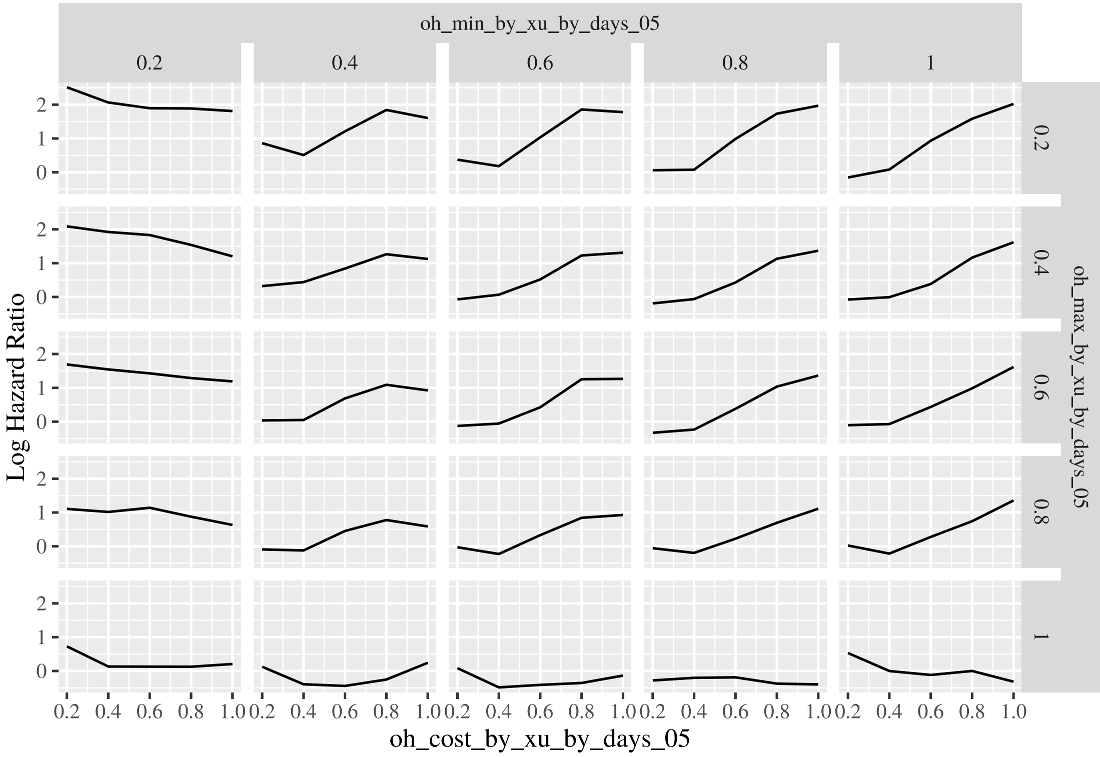

Figure 29. Triple-way interaction coefficients of cost_max_min_ia_only_5 model

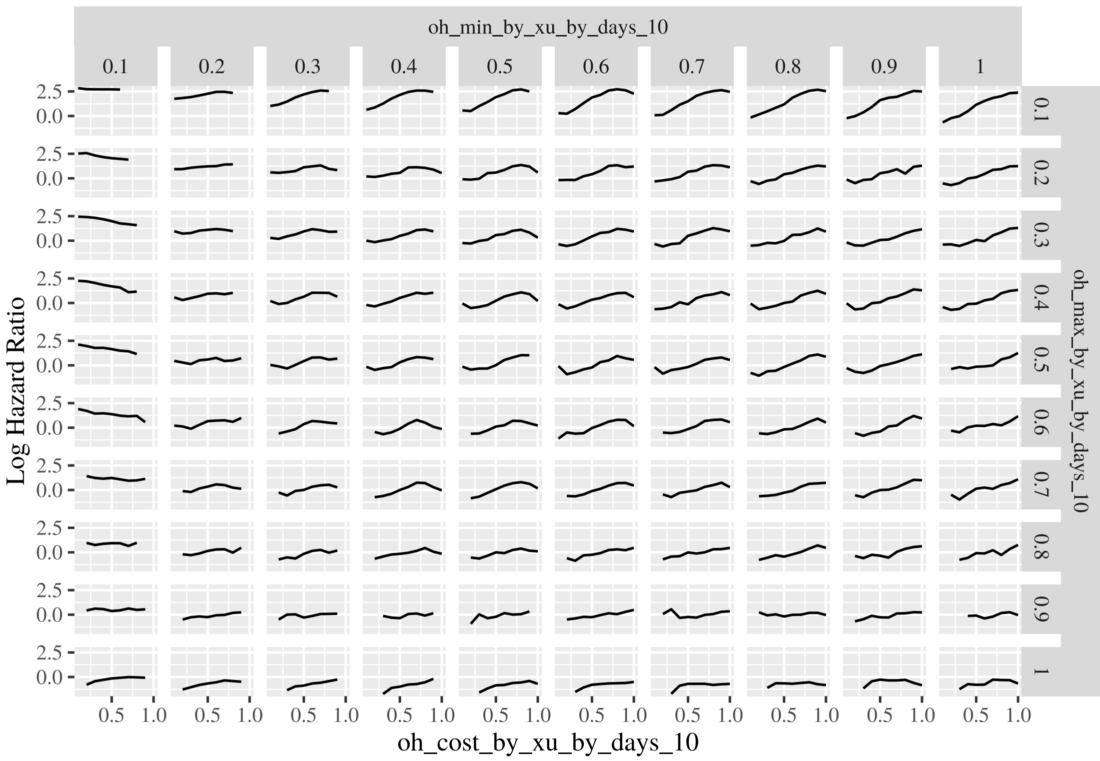

Figure 30. Triple-way interaction coefficients of cost_max_min_ia_only_10 model

The layout of the whole figures are as follows:

- Each curve shows the triple-way interaction coefficients across ordinal levels of the overhang on cost reference price.
- Each column of facets, from left to right, shows increasing categories of the overhang on minimum reference price. The leftmost column corresponds to the reference level. Hence, this is the case when current price is closest to the minimum price. The rightmost column shows the case when current price is furthest away from the minimum price.
- Each row of facets, from bottom to top, shows decreasing categories of the overhang on maximum reference price. The lowest row corresponds to the reference level. Hence, this is the case when current price is closest to the maximum price. The topmost row shows the case when current price is furthest away from the maximum price.

As in the case with the KM figure, the top-right facet is the case when the current price is furthest away from both the maximum and minimum reference prices and is taken as the starting point in the interpretation. A larger coefficient means a higher propensity to close a position. The findings are as follows:

- When the current price is furthest away from both the maximum and minimum reference prices, the curve has the highest positive slope, meaning disposition effect with respect to overhang on cost reference price is at its strongest.
- Ceteris paribus, going from the top (furthest from maximum reference price) to the bottom (closest to maximum reference price) rows, the curves are lower. Hence, as the current price approaches above to maximum reference price, propensity to close a position is weaker, in line with reverse disposition effect.

- Ceteris paribus, going from the rightmost (furthest from minimum reference price) to the leftmost (closest to minimum reference price) columns, the curves are higher. Hence, as the current price approaches below to minimum reference price, propensity to close a position is stronger, in line with reverse disposition effect.

These main effects are in agreement with those derived from the KM curves. The secondary effects also confirm the previous findings:

- In each row, in facets from right to left, as current prices get closer below to the minimum price, the curves become flatter and even negative sloping at the end in some cases. Hence, as price approaches the minimum price, the dominance of the disposition effect coming from the overhang on cost reference price weakens and reverse disposition effect coming from the overhang on minimum reference price starts to dominate.
- In each column, in facets from top to bottom, as current prices get closer below to the maximum price, the curves become flatter and even negative sloping at the end in some cases. Hence, as price approaches the maximum price, the dominance of the disposition effect coming from the overhang on cost reference price weakens and reverse disposition effect coming from the overhang on maximum reference price starts to dominate.

Using these triple-way interaction coefficients from the Cox models, these main and secondary effects are further tested for significant monotonicity across three dimensions with the help of Bayesian regression with monotonic effects below.

A third type of visualization demonstrates an enhanced version of Odean's ratios. In the original version, the proportion of positions realized while at a profit (current price above cost reference) is called as PGR and proportion of positions realized while at a loss (current price below cost reference) is called as PLR. Odean's ratio is the ratio of PGR to PLR. When this ratio is above 1, propensity to sell at profit is higher than propensity to sell at loss, signaling the existence of disposition effect.

In the enhanced version introduced here, apart from the three ordinal variables for cost, maximum and minimum reference prices, the days since the start of a position as an ordinal variable used for controlling the time horizon of other ordinal variables is also included as a fourth dimension. In the KM and Cox methods, the days as the time horizon is an integral part of the analysis, since hazard rate may vary across time. In this enhanced Odean's ratios methods, for each of the collection of records (each record showing the status of a position at a certain day) corresponding to a tuple of categories of the four variables (cost, maximum and minimum reference based overhang categories and the days category), the percent of records in which positions are realized is calculated. The ratios are presented in detail in Appendix B (p. 219) and also as a multi-faceted plot similar to the previous ones in Figure 31.

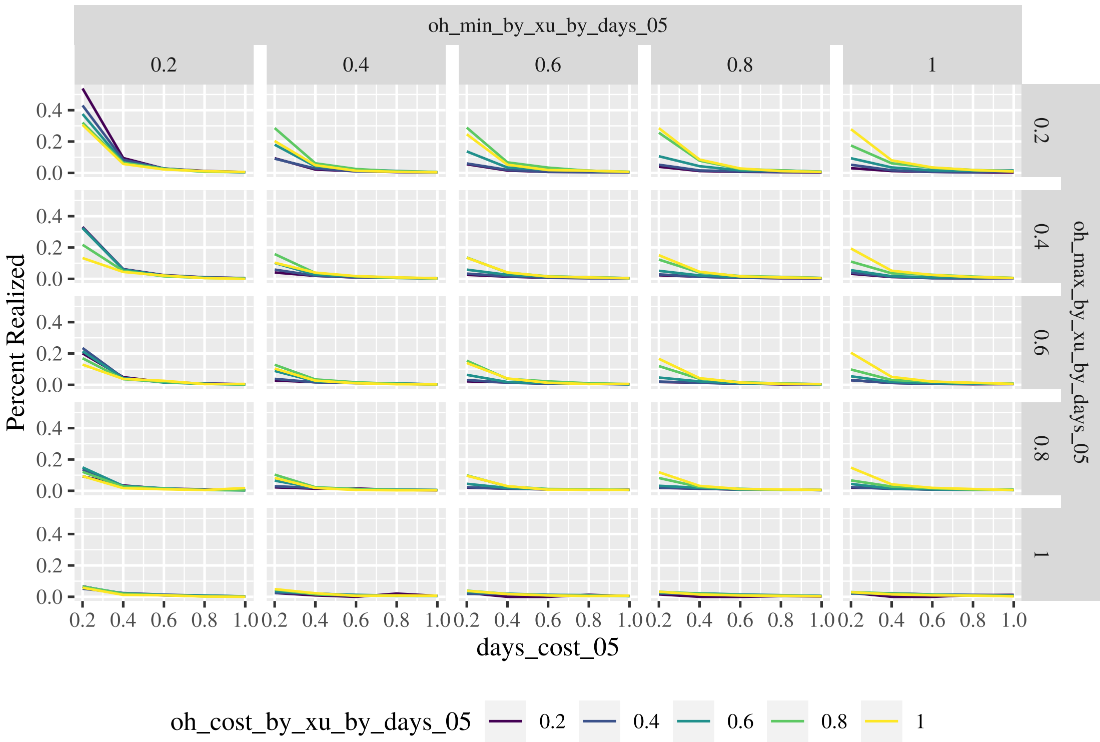

Figure 31. Odean's ratios for cost_max_min_5 model

The layout of the figure is as follows:

- Each curve shows the Odean's ratio or percent of records closed across day categories from shorter to longer horizons.
- In each facet, each stacked colored line corresponds to a categorical level of the overhang on cost reference price.
- Each column of facets, from left to right, shows increasing categories of the overhang on minimum reference price. The leftmost column corresponds to the reference level. Hence, this is the case when current price is closest to the minimum price. The rightmost column shows the case when current price is furthest away from the minimum price.

- Each row of facets, from bottom to top, shows decreasing categories of the overhang on maximum reference price. The lowest row corresponds to the reference level. Hence, this is the case when current price is closest to the maximum price. The topmost row shows the case when current price is furthest away from the maximum price.

As in the case with previous plots, the top-right facet is the case when the current price is furthest away from both the maximum and minimum reference prices and is taken as the starting point in the interpretation. A larger ratio means, a higher propensity to close a position. The findings are as follows:

- When the current price is furthest away from both the maximum and minimum reference prices, the curve for the lowest category of overhang on cost reference is the lowest one. Hence, propensity to close a position is lower. The curves are monotonically and parallelly higher for higher overhang categories. Hence, propensity to close a position gets stronger. This is in line with the disposition effect.
- Ceteris paribus, going from the top (furthest from maximum reference price) to the bottom (closest to maximum reference price) rows, the curves are lower. Hence, as the current price approaches above to maximum reference price, propensity to close a position is weaker, in line with reverse disposition effect.
- Ceteris paribus, going from the rightmost (furthest from minimum reference price) rows to the leftmost (closest to minimum reference price) columns, the curves are higher. Hence, as the current price approaches below to minimum reference price, propensity to close a position is stronger, in line with reverse disposition effect.

These main effects confirm the ones found in KM curves and Cox coefficients.

The secondary effects as interactions of overhang on cost reference with the overhang on maximum and minimum references also exist in the plot:

- In each row, in facets from right to left, as current prices get closer below to the minimum price, the curves for each overhang on cost category get closer and are even partially reversed in some cases. Hence, as price approaches the minimum price, the dominance of the disposition effect coming from the overhang on cost reference price weakens and reverse disposition effect coming from the overhang on minimum reference price starts to dominate.
- In each column, in facets from top to bottom, as current prices get closer below to the maximum price, the curves for each overhang on cost category get closer and are even partially reversed in some cases. Hence, as price approaches the maximum price, the dominance of the disposition effect coming from the overhang on cost reference price weakens and reverse disposition effect coming from the overhang on maximum reference price starts to dominate.

So the secondary effects are also in line with the ones derived from the previous methods.

As in the case of triple-way interaction coefficients from the Cox models, using these Odean's rations, the main effects are further tested for significant monotonicity across three dimensions with the help of Bayesian regression with monotonic effects below.

After the visual methods, the monotonicity of the propensity to close a position across ordinal categories using the coefficients of triple-way interactions among cost, maximum and minimum reference based overhang variables is tested using Bayesian regression methods with monotonic effects.

For this purpose, a new synthetic dataset is formed by sampling from the distribution of each coefficient using the estimate and its standard error. For each coefficient, a random sample of 50 is generated, and this sample size is kept flat across all coefficients. In the original dataset, the number of records that correspond to a certain categorical level of the triple-way interaction variable varies extensively. For the five category model, excluding the records corresponding to the baseline (where all variables have a value of zero), a total of 8,143,826 records are distributed across 124 categorical levels. Within the levels, the smallest number of records is 1,441 and the largest number of records is 425,285. For the ten category model, excluding the baseline records, a total of 8,274,823 records are distributed across 998 categorical levels. Within the levels, the smallest number of records is 8 and the largest number of records is 182,888. By creating a balanced synthetic dataset, the dominance of the coefficients that corresponds to the subsets with the largest number of records is prevented.

At the end, a dataset is created by sampling from the distributions of coefficients using their point estimates and standard errors. In this dataset, the dependent variable consists of the sampled values of the coefficients. Independent variables are three ordinal variables corresponding to the same overhang variables used in the Cox models. In the Cox models, each level of the variable (or the interaction terms if exists) are converted to dichotomous variables. However, in the Bayesian regression with monotonic effects, each monotonic variable is taken as a single variable and is not converted to dummies. In an ordinary linear regression model (including Cox models), a single coefficient is calculated and a linear relationship with the variable is assumed. However, in the monotonic effects model, for each variable two things are calculated that constitute a coefficient:

- The average difference of the effect between subsequent categories, either positive or negative
- A simplex, a set of positive decimal values between 0 and 1 and summing up to 1, that corresponds to the fraction of the total difference of the effect between the highest and lowest categories. Since all simplex values are positive, the steps between subsequent levels are in the same direction, ensuring monotonicity.

So basically, a non-linear but monotonically increasing or decreasing step function is converged for each monotonic variable, in a multivariate setting. When the estimate is positive, the relationship between the dependent variable and the monotonic variable is monotonically increasing and vice versa. As with the case of an ordinary regression, a confidence interval is also reported for each estimate. A monotonic relationship is significant when the lower and upper bounds of the 95% confidence interval are at the same side of 0.

Apart from the significance of relationships, a separate Rhat value is also reported for each estimate. In Bayesian methods, chains of estimates are sampled in which values from the previous iteration acts as the priors to the next iteration. Rhat value is at least 1 and a value closer to 1 shows how well these sampling chains converged. It is recommended that estimates with a Rhat value less than 1.05 should be used (Guo et al., n.d.).

For the coefficients from five-category Cox model with triple-way interactions, the estimates of monotonic effects are presented in Table 18. The chains for all three monotonic variables have converged sufficiently with Rhat values at 1, and the estimates of all three variables are significant, with very narrow confidence intervals. Estimate for the effect of overhang on cost reference price is positive. So the relationship between the coefficients and this variable's levels is monotonically increasing, confirming the disposition effect found in previously mentioned methods. Estimate for the effect of overhang on maximum and minimum reference prices are negative. So the relationship between the coefficients and these variables' levels are monotonically decreasing, confirming the reverse disposition effect found in previously mentioned methods.

The monotonic distribution of the overall effect of the variables across subsequent levels are presented in Figure 32.

Table 18. Estimates From the Bayesian Monotonic Effects Model of the Relationships Between Triple-Way Interaction Coefficients and Ordinal Variables of cost_max_min_ia_only_5 Model

Variable	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
Intercept	1.54	0.02	1.50	1.57	1
mooh_cost_by_xu_by_days_05	0.19	0.00	0.18	0.20	1
mooh_max_by_xu_by_days_05	-0.32	0.00	-0.33	-0.31	1
mooh_min_by_xu_by_days_05	-0.21	0.00	-0.22	-0.20	1

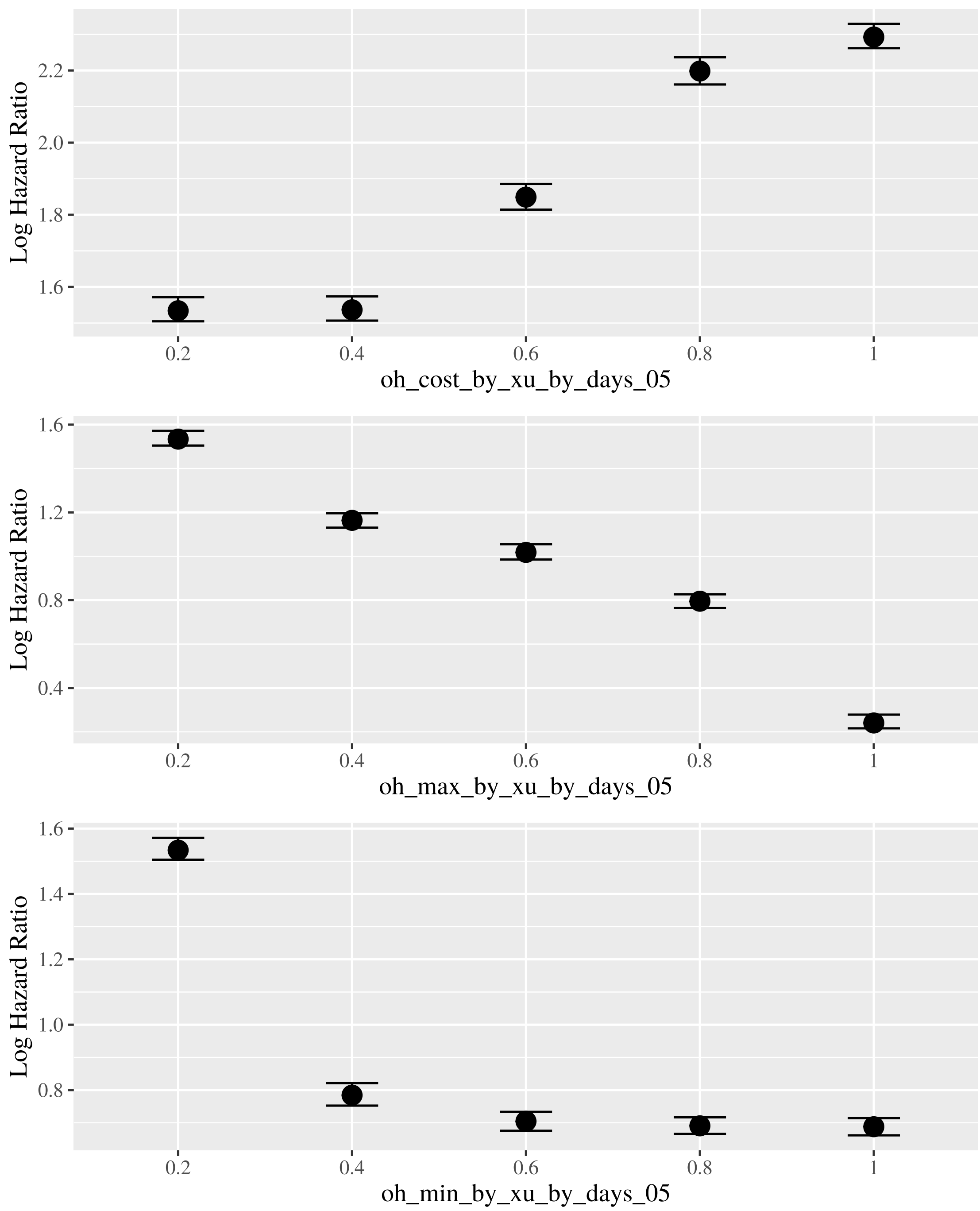

Figure 32. Bayesian conditional monotonic effects for Cox regression coefficients from cost_max_min_5_ia_only model

On one hand, estimates for the effects of subsequent levels of overhang on cost reference price variable are monotonically increasing, confirming disposition effect. On the other hand, estimates for the effects of subsequent levels of overhang on maximum and minimum reference price variables are monotonically decreasing, confirming reverse disposition effect.

The same analysis is repeated for the coefficients from the ten category model, again sampling 50 values for each coefficient. For the coefficients from ten-category Cox model with triple-way interactions, the estimates of monotonic effects are presented in Table 19. All Rhat values are at 1, so the estimates are meaningful since they are from converged chains. The findings confirm those derived from the five-category version above, with the exception that estimate for the effect of overhang on minimum reference price is not significant at 0.05 level. Secondary effects for the interaction between overhang on cost reference price and other two variables will be investigated further, making the main effects of overhang on minimum reference price significant.

The monotonic distribution of the overall effect of the variables across subsequent levels are presented in Figure 33.

Table 19. Estimates From the Bayesian Monotonic Effects Model of the Relationships Between Triple-Way Interaction Coefficients and Ordinal Variables of cost_max_min_ia_only_10 Model

Variable	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
Intercept	0.53	0.55	-0.50	1.64	1.00
mooh_cost_by_xu_by_days_10	0.23	0.05	0.14	0.34	1.00
mooh_max_by_xu_by_days_10	-0.43	0.06	-0.55	-0.32	1.01
mooh_min_by_xu_by_days_10	-0.05	0.05	-0.14	0.04	1.00

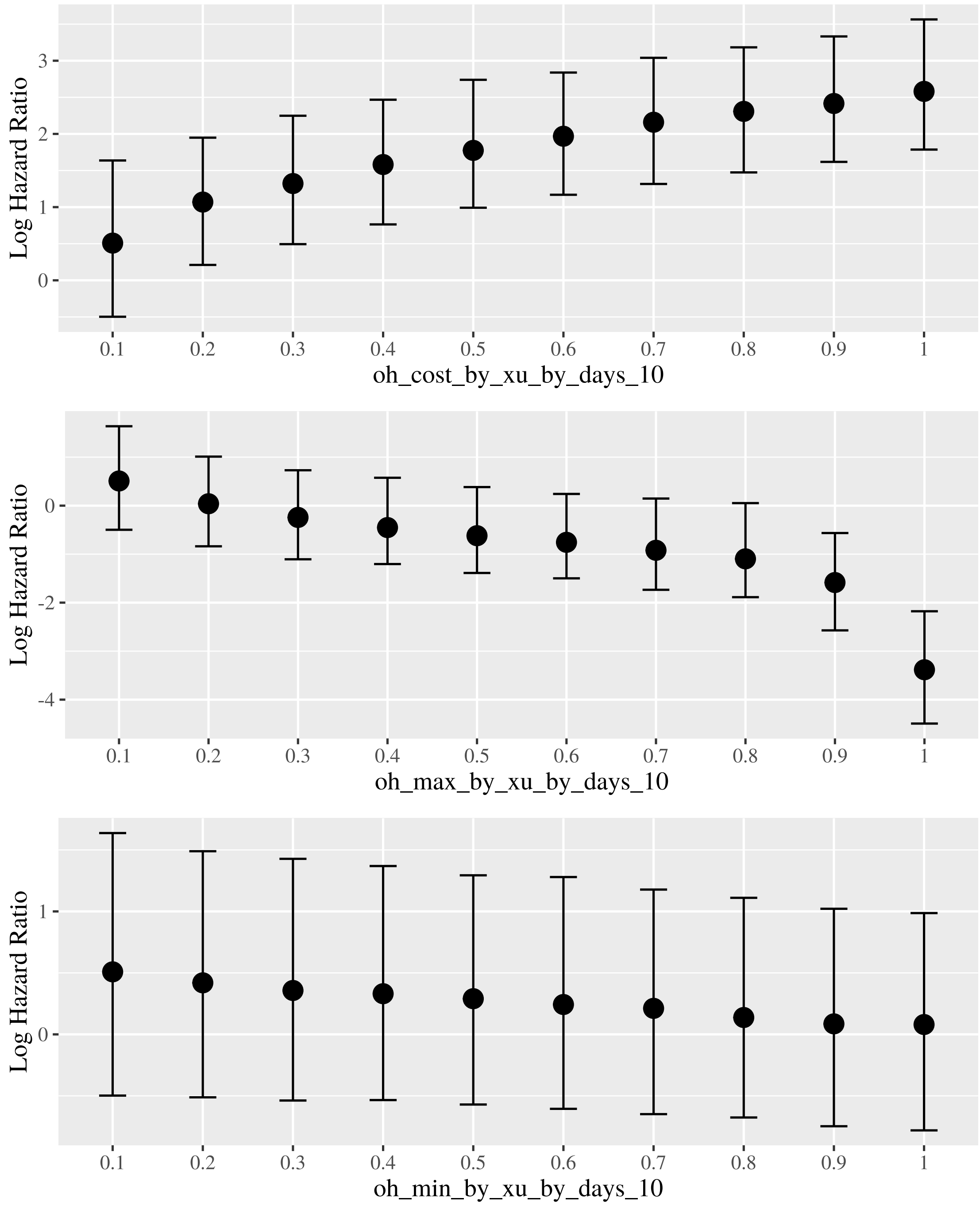

Figure 33. Bayesian conditional monotonic effects for Cox regression coefficients from cost_max_min_10_ia_only model

The direction of the monotonic effects of three overhang variables confirm the above results from the five category model.

The Odean’s ratios from five category cost_max_min_5 model are also processed similarly to run a Bayesian monotonic model to confirm the results from the coefficients of Cox models. Unlike coefficient estimates from Cox models, Odean's ratios are simple percent values without any standard error values. However, these ratios are derived from dividing the number of records in which a position is closed to the total number of records at the intersection of the selected levels of three overhang variables and the categorical days variable. Since each record can have either of two status values, position closed or position not closed, binomial distribution can be used for sampling values from the point ratios. The sampled number of records with positions closed are divided by the total number of records for the combination of category levels to calculate the sampled ratios.

Due to the larger number of category combinations coming from the additional fourth variable of days, the sample per Odean's ratio is kept at 20. And since records are now distributed across four five-category variables, the record numbers in some combinations may be too low to be considered representative of the underlying effects. For this reason, only those Odean's measures for which at least 100 records exist are included in the sampling process and the model.

The estimates of monotonic effects are presented in Table 20. All Rhat values of the overhang variables are at most 1.01, so the estimates for these variables are

Table 20. Estimates From the Bayesian Monotonic Effects Model of the Relationships Between Odean's Ratios and Ordinal Variables of cost_max_min_5 Model

Variable	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
Intercept	0.14	0	0.14	0.15	1.01
mooh_cost_by_xu_by_days_05	0.00	0	0.00	0.01	1.01
mooh_max_by_xu_by_days_05	-0.01	0	-0.01	-0.01	1.00
mooh_min_by_xu_by_days_05	-0.01	0	-0.01	-0.01	1.01
modays_cost_05	-0.03	0	-0.03	-0.02	1.06

meaningful since they are from converged chains. The Rhat value for the days variable is 1.06, so it is disregarded. The estimates for the overhang variables are all significant and their directions confirm the previous findings on disposition and reverse-disposition effects.

The monotonic distribution of the overall effect of the variables across subsequent levels are presented in Figure 34.

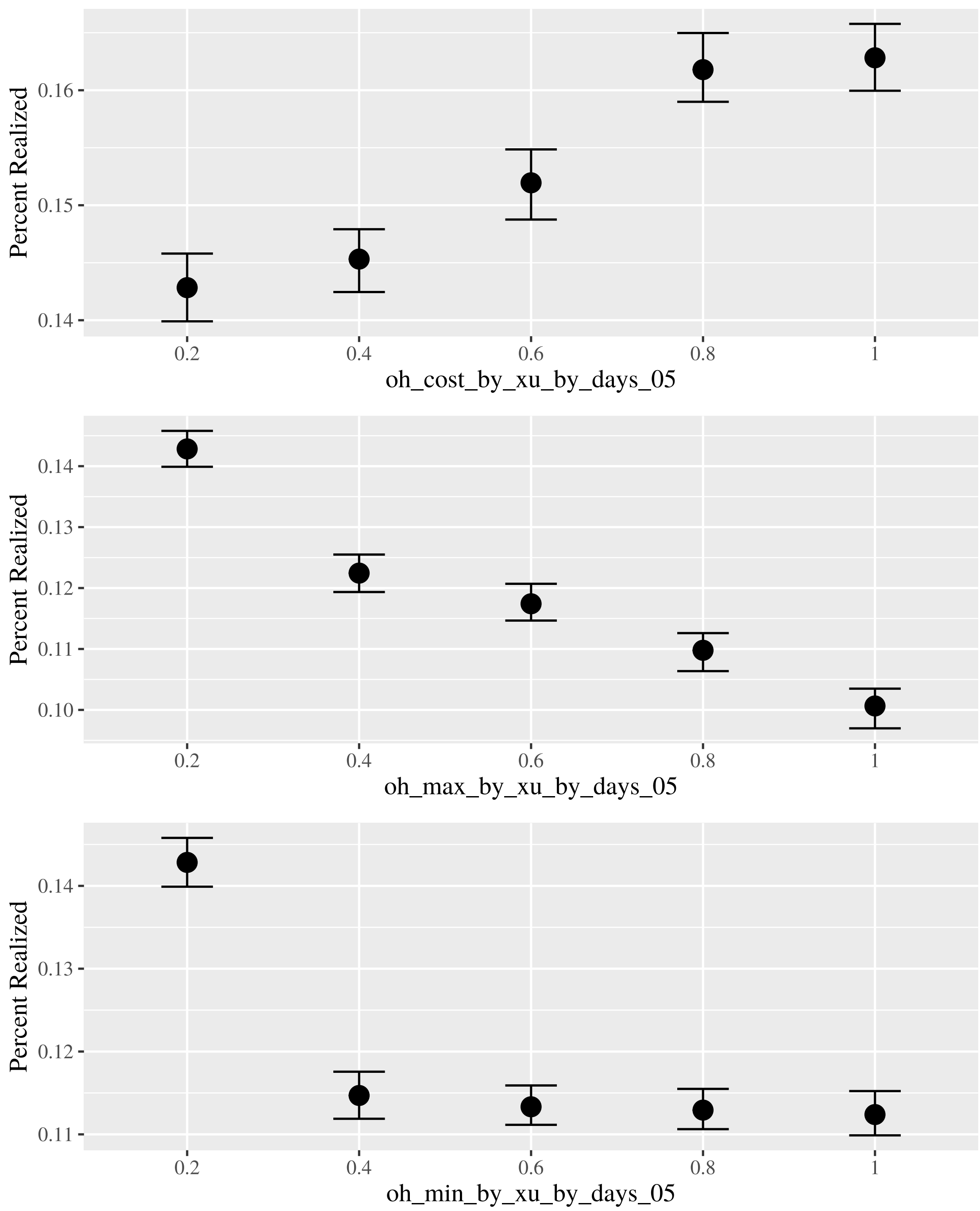

Figure 34. Bayesian conditional monotonic effects for Odean's ratios from cost_max_min_5 model

The direction of the monotonic effects of three overhang variables confirm the above results from the similar models using Cox regression coefficients.

Two secondary effects are discovered in visual methods:

- With higher categories of overhang on minimum price reference, as the current price is further away from the minimum price reference, the disposition effect on cost reference gets stronger, and vice versa.
- With higher categories of overhang on maximum price reference, as the current price is closer to the maximum price reference, the disposition effect on cost reference weakens, and vice versa.

So when interaction terms are added to the Bayesian monotonic effects models:

- A positive estimate is expected for the interaction between overhang on minimum price reference and overhang on cost reference.
- A negative estimate is expected for the interaction between overhang on maximum price reference and overhang on cost reference.

For the coefficients from five-category Cox model with triple-way interactions, the estimates of the Bayesian monotonic effects model with interaction terms for secondary effects are presented in Table 21. Each coefficient is sampled 50 times. All

Table 21. Estimates From the Bayesian Monotonic Effects Model (With Secondary Effects) of the Relationships Between Triple-Way Interaction Coefficients and Ordinal Variables of cost_max_min_ia_only_5 Model

Variable	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
Intercept	1.83	0.01	1.81	1.86	1.01
mooh_cost_by_xu_by_days_05	0.02	0.00	0.01	0.03	1.01
mooh_max_by_xu_by_days_05	-0.11	0.00	-0.11	-0.10	1.01
mooh_min_by_xu_by_days_05	-0.40	0.00	-0.41	-0.39	1.00
mooh_cost_by_xu_by_days_05: mooh_max_by_xu_by_days_05	-0.10	0.00	-0.10	-0.10	1.01
mooh_cost_by_xu_by_days_05: mooh_min_by_xu_by_days_05	0.10	0.00	0.10	0.11	1.01

Rhat values are below 1.05, so the estimates are meaningful since they are from converged chains. Estimates of all variables are significant. The main effects of the three overhang variables are in line with the previous findings:

- Positive estimate of overhang on cost reference confirms disposition effect.
- Negative estimates of overhang on maximum and minimum price references confirm reverse disposition effect.

The estimates for the secondary effects are also in line with the expectations:

- Estimate for the interaction term between overhang on maximum price and overhang on cost is negative. So as the current price approaches maximum price reference with higher categories of the overhang on maximum price reference, disposition effect from overhang on cost reference weakens and vice versa.
- Estimate for the interaction term between overhang on minimum price and overhang on cost is positive. So as the current price moves away from minimum price reference with higher categories of the overhang on minimum price reference, disposition effect from overhang on cost reference gets stronger.

The monotonic distribution of the overall effect of the variables across subsequent levels are presented in Figure 35 and Figure 36.

The directions of the main effects are similar to the ones found in previous monotonic effects models without interaction terms for secondary effects.

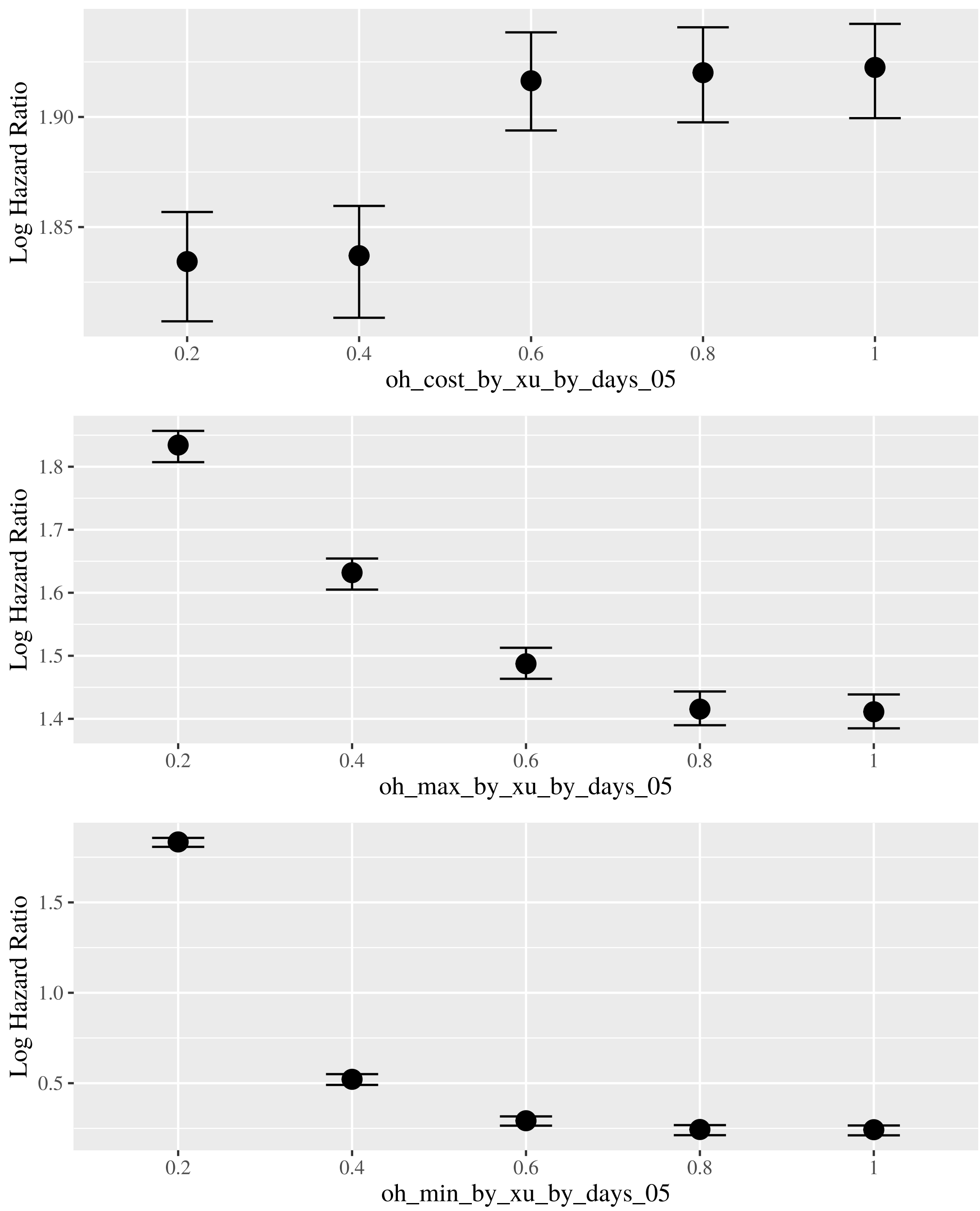

Figure 35. Bayesian conditional monotonic effects (with secondary effects) for Cox regression coefficients from cost_max_min_5_ia_only model, main effects

Figure 36 summarizes the secondary effects from interactions between monotonic effects visually. In the upper plot, the secondary effect of overhang on cost reference is upward trending. Hence, it exhibits a slight disposition effect for the lowest category of overhang on maximum price reference when the current price is furthest away from the maximum price reference. With higher categories of overhang on maximum price as the current price approaches the maximum price, the secondary effect of overhang on cost reference is reversed and exhibits stronger downward trend in line with reverse disposition effect.

In the lower plot, the secondary effect of overhang on cost reference is upward trending. Hence, it exhibits a slight disposition effect for the lowest category of overhang on minimum price reference when the current price is closest to the minimum price reference. With higher categories of overhang on minimum price as the current price moves away from the minimum price, the secondary effect of overhang on cost reference exhibits a steeper upward trend, in line with stronger disposition effect.

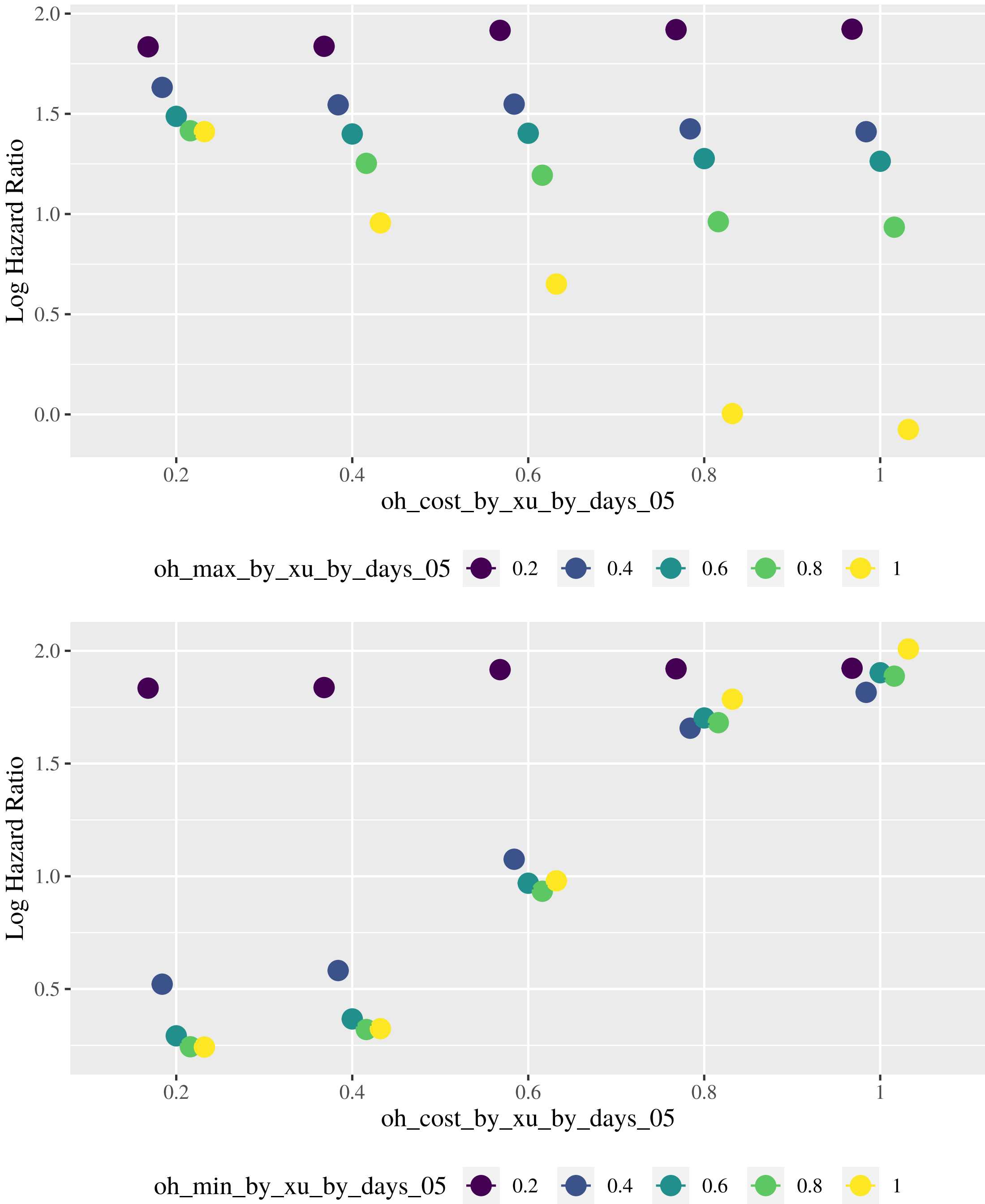

Figure 36. Bayesian conditional monotonic effects (with secondary effects) for Cox regression coefficients from cost_max_min_5_ia_only model, secondary effects

Bayesian monotonic effects model with interaction terms for secondary effects is repeated for the Cox regression coefficients from ten category model. A high number of levels arise from working with ten category variables coupled with interaction terms in this model. So, in order to get more reliable results, the sampling is done for only those Cox regression coefficients that are significant at 0.05 level and that have at least 500 corresponding records. Each coefficient is sampled 50 times. Estimates are presented in Table 22. All the Rhat values except one are below 1.05. Rhat of the interaction term between overhang on maximum price reference and overhang on cost reference is at the margin of 1.05. While this estimate should be disregarded because of not being converged sufficiently, the estimate is significant, and its direction confirms the above findings. Estimates of all other variables are also significant, and their directions are in line with the ones found in the five category case above.

The monotonic distribution of the overall effect of the variables across subsequent levels are presented in Figure 37 and Figure 38. The directions of the main effects confirm the previous case with five categories. Secondary effects also exhibit the same trends with the previous case.

Table 22. Estimates From the Bayesian Monotonic Effects Model (With Secondary Effects) of the Relationships Between Triple-Way Interaction Coefficients and Ordinal Variables of cost_max_min_ia_only_10 Model

Variable	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
Intercept	2.78	0.01	2.77	2.80	1.01
mooh_cost_by_xu_by_days_10	0.01	0.00	0.01	0.02	1.00
mooh_max_by_xu_by_days_10	-0.18	0.00	-0.18	-0.17	1.04
mooh_min_by_xu_by_days_10	-0.32	0.00	-0.32	-0.31	1.00
mooh_cost_by_xu_by_days_10: mooh_max_by_xu_by_days_10	-0.02	0.00	-0.02	-0.02	1.05
mooh_cost_by_xu_by_days_10: mooh_min_by_xu_by_days_10	0.03	0.00	0.03	0.03	1.00

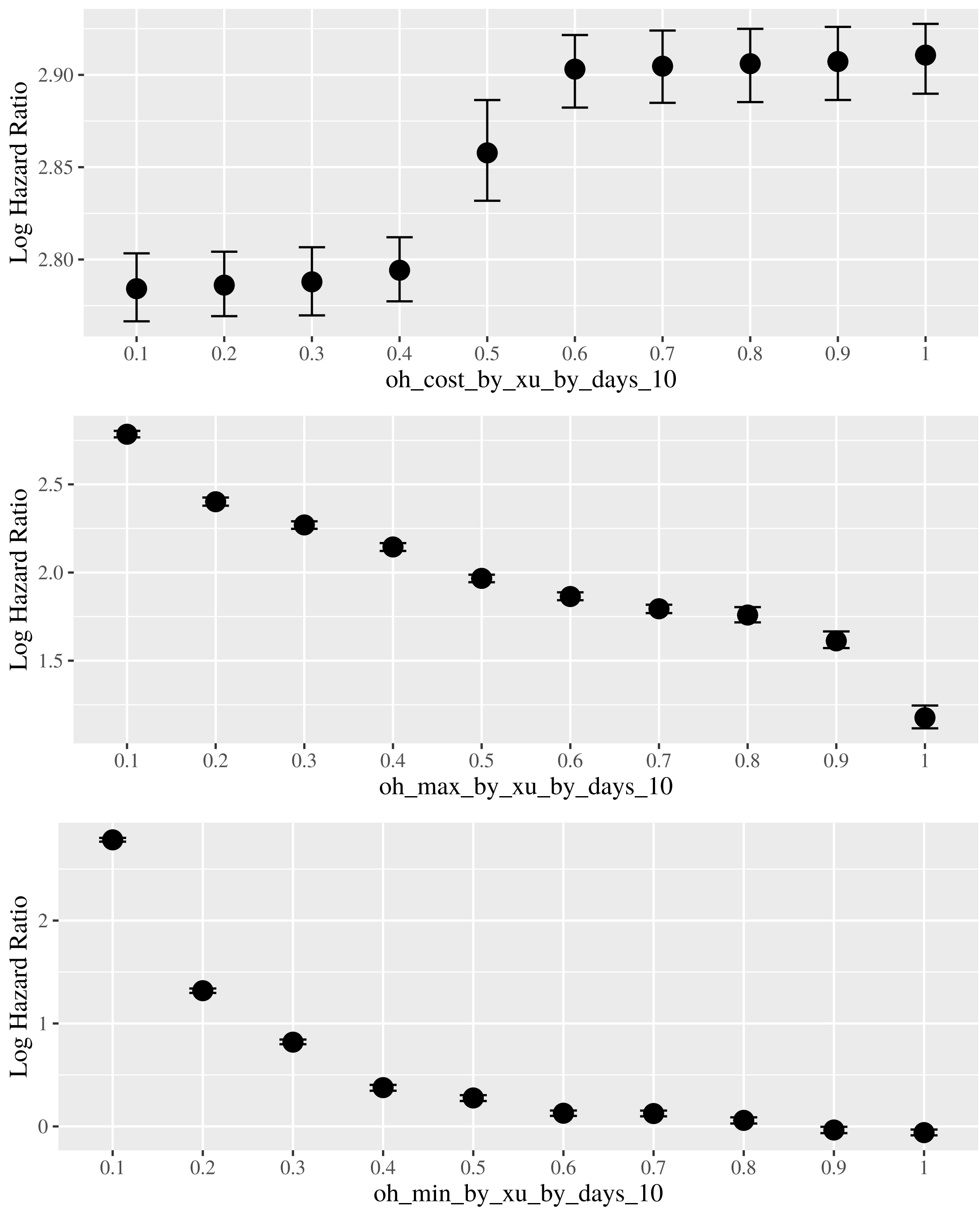

Figure 37. Bayesian conditional monotonic effects (with secondary effects) for Cox regression coefficients from cost_max_min_10_ia_only model, main effects

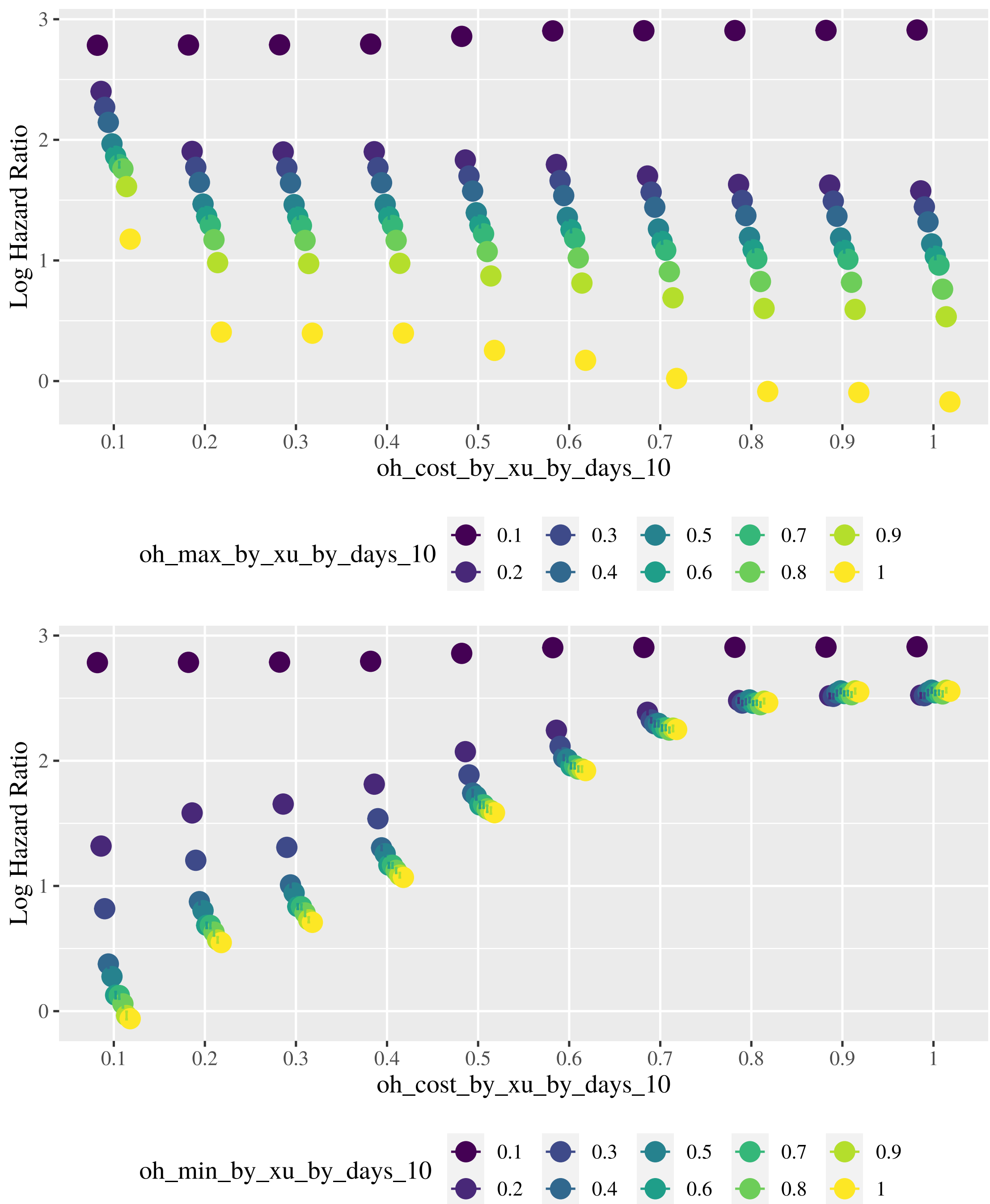

Figure 38. Bayesian conditional monotonic effects (with secondary effects) for Cox regression coefficients from cost_max_min_10_ia_only model, secondary effects

To sum up, in these group of models all overhang calculations are done using maximum intraday price for days positions are open, while actual transaction prices are used for days positions are closed. Ordinal overhang on cost, maximum and minimum price reference variables in this model are controlled for the following:

- Ordinal days variable
- Ordinal market return variable controlled for ordinal days variable

The methods utilized are as follows:

- Cox model regressions including ordinal overhang variables separately (five category and ten category versions)
- Kendall's Tau for the coefficients of all levels of each ordinal overhang variable from the above model (five category and ten category versions)
- KM survival curves across three ordinal overhang variables (five category version)
- Odean's ratios of records where positions are closed across three ordinal overhang variables and also ordinal days variable (five category version)
- Bayesian regression with monotonic effects using random samples from Odean's ratios (five category version)
- Cox model regressions including triple-way interactions of ordinal overhang variables (five and ten category versions)
- Bayesian regression with monotonic effects using random samples from estimates and standard errors of coefficients in the Cox models (five category and ten category versions)
- Bayesian regression with monotonic effects including interaction terms between overhang ordinals using random samples from estimates and standard errors of coefficients in Cox models (five category and ten category versions).

The findings can be summarized as follows:

- The propensity to close a position increases as current price moves above the cost reference and the propensity decreases as current price moves below the cost reference, in accordance with disposition effect.

- The propensity to close a position decreases as current price is closer above to the maximum price reference, in accordance with reverse disposition effect.
- The propensity to close a position increases as current price is closer below to the minimum price reference, in accordance with reverse disposition effect.
- These three effects appear and alter the investor's propensity to close a position together, so investors consider the three reference prices together in their behaviors. Maximum and minimum reference prices can be updated as each investor experiences new highs and lows for their position. And the level of overhang on these reference prices are evaluated within the contexts of the time horizon since the position is opened (or new maximum or minimum price is attained by the investor) and the market return since the position is opened (or new maximum or minimum price is attained by the investor).
- The strength of the disposition effect from the cost reference is also altered by the reverse-disposition effects from the maximum and minimum references. When the current price is away from maximum and minimum prices, the disposition effect from the cost reference is at its strongest. As the current price is closer to maximum and/or minimum prices, this disposition effect gets weaker and is even overtaken by the reverse disposition effect from maximum and minimum price references.

In numerous studies, reverse disposition effect is found alongside disposition effect using overhang on cost reference, in different regions of profit or loss. The simultaneous testing of three reference prices brings an explanation to this inconsistent or hard to explain phenomenon.

7.2.2 Overhang without controlling for holding period or market return

In the main model, the overhang variables are controlled for holding period and market return in a nested manner using MUHBOD as explained in detail in Subsection 7.2.1 (p. 149). The advantage of that model is, the ordinal categorization is made within contexts, namely holding period and market return in this example, which may alter the perception of the extent of the overhang. Another advantage is the balancing of the records across categories. However, without MUHBOD, number of records especially in the combination of extreme categories is too low and even zero.

An alternative set of models are also run to check whether these effects can be observed when the variables are calculated using simpler methods. The failure to observe monotonic effects between propensity to close a position and ordinal overhangs over reference prices calculated by different methods may be inspiring to understand that simultaneous and interacting disposition and reverse disposition effects are elusive and not easy to observe with data using traditional data wrangling and modeling methods. While alternative reference prices have been explored in the literature, the simultaneous and interacting effect of three reference prices have not been incorporated before. The reason may be that consistent and significant linear or monotonic effects are hard or impossible to observe, especially when advanced data wrangling and modeling methods as done in this thesis are not utilized.

In the set of simple models reported here, the overhang variables are discretized without controlling for other ordinal variables in a nested manner. Without MUHBOD, a certain numeric level of overhang will be deemed to have the same ordinal level regardless of the time horizon of the position or the accompanying market return during the same period.

The coefficients and p-values of the five category simple case with no triple interactions (cost_max_min_simple_5 model) are shown in Table 23. All coefficients are significantly different from zero. When the coefficients of the categorical levels of the three variables are plotted together in Figure 39, it is clear that there is no apparent perfect or almost monotonic relationship between the coefficients and the ordinal levels of overhang variables.

Table 23. Ordinal Levels, Coefficients and P-Values (in Parentheses) of Ordinal Variables in cost_max_min_5 Model

	Variables		
Level	oh_cost_05	oh_max_05	oh_min_05
0.2	0.21 (0)	0.91 (0)	0 (-)
0.4	0.1 (0)	0.99 (0)	-0.72 (0)
0.6	0 (-)	1.09 (0)	-0.56 (0)
0.8	0.55 (0)	1.06 (0)	-0.27 (0)
1.0	0.59 (0)	0 (-)	-0.06 (0)

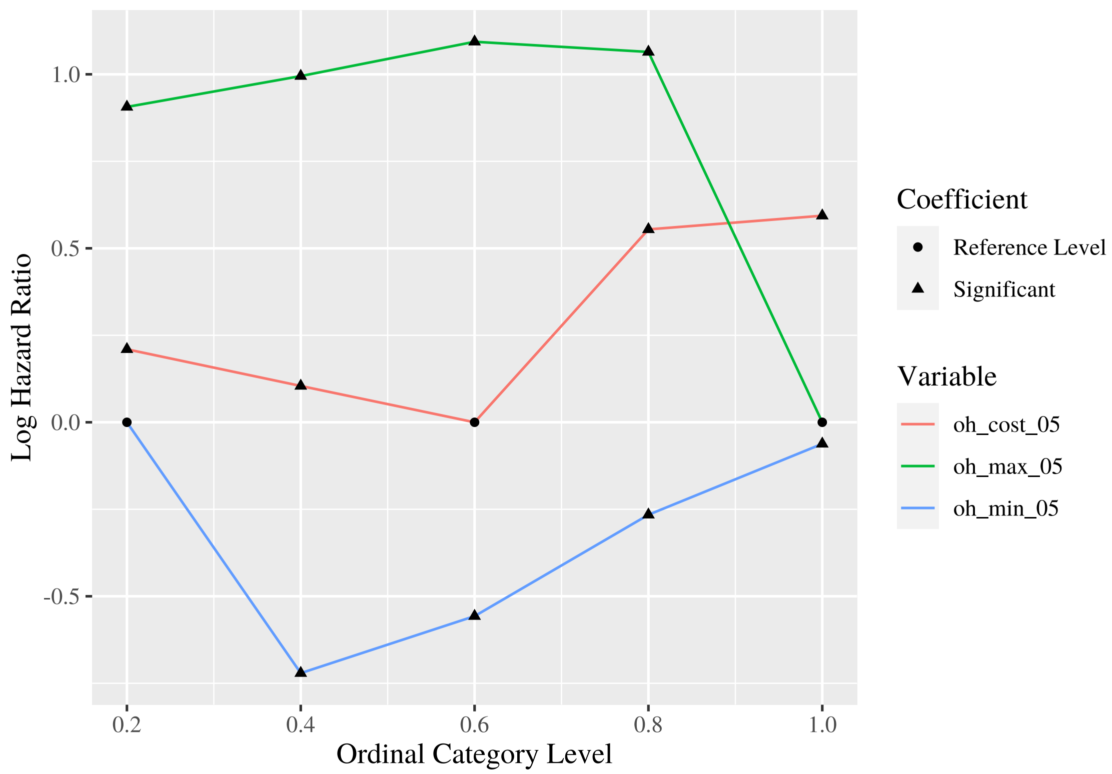

Figure 39. Coefficients of cost_max_min_simple_5 model with corresponding significance status

The level of monotonicity of the coefficients of each variable are tested with Kendall's Tau as presented in Table 24. The Tau values are close to zero and are all insignificant. The results of the ten category version of the model are also concordant with those of the five category version. The coefficients and p-values of the ten category simple case with no triple-way interactions (cost_max_min_simple_10 model) are shown in Table 25. All coefficients are significantly different from zero.

Table 24. Kendall's Tau Values and Corresponding P-Values for Coefficients of cost_max_min_simple_5 Model

Variables	Tau	P-value
oh_max_05	0.0	1.00
oh_min_05	0.2	0.82
oh_cost_05	0.4	0.48

Table 25. Ordinal Levels, Coefficients and P-Values (in Parentheses) of Ordinal Variables in cost_max_min_10 Model

	Variables		
Level	oh_cost_10	oh_max_10	oh_min_10
0.1	0.72 (0)	2.03 (0)	0 (-)
0.2	0.58 (0)	2.03 (0)	-1.09 (0)
0.3	0.47 (0)	2.11 (0)	-1.25 (0)
0.4	0.31 (0)	2.18 (0)	-1.2 (0)
0.5	0.02 (0.02)	2.22 (0)	-1.14 (0)
0.6	0 (-)	2.25 (0)	-1.09 (0)
0.7	0.9 (0)	2.24 (0)	-0.89 (0)
0.8	1.42 (0)	2.1 (0)	-0.71 (0)
0.9	1.25 (0)	1.91 (0)	-0.48 (0)
1.0	0.83 (0)	0 (-)	-0.45 (0)

The plot of ordinal variable coefficients in Figure 40 confirms the lack of apparent monotonicity.

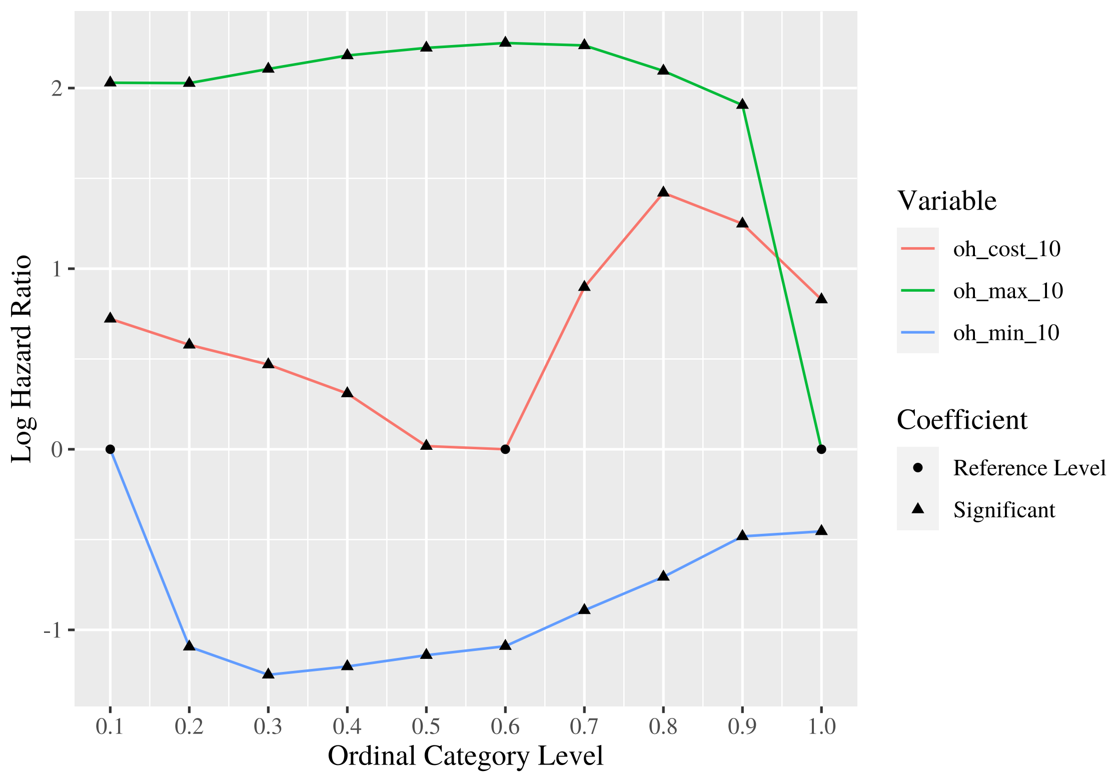

Figure 40. Coefficients of cost_max_min_simple_10 model with corresponding significance status

The level of monotonicity of the coefficients of each variable are tested with Kendall's Tau as presented in Table 26. The Tau values are close to zero. Only overhang on minimum reference is significant at 0.1. However, the weak positive relationship is reversed at higher categories as revealed in Figure 40. Five category Cox model with triple-way interactions is also run for this set of variables. The distribution of the p-values of the coefficients from this model according to their significance levels are presented in Table 27. 96% of the 99 coefficients are significant at 0.05 level. However, the total number of coefficients is far lower than that of the main model (cost_max_min_5_ia_only). The reason is that, without nested controlling of categorical variables in MUHBOD, the distribution of records across categories may be imbalanced, so that in some combinations of the categories of three ordinal variables, there are no records at all. The coefficients are presented in detail in Appendix E (p. 230). They are also organized into a multi-faceted plot in Figure 41. Only the coefficients that correspond to categories with at least 500 records in the data are included, to exclude outlying or unreliable coefficients due to small sample size.

Table 26. Kendall's Tau Values and Corresponding P-Values for Coefficients of cost_max_min_simple_10 Model

Variables	Tau	P-value
oh_max_10	-0.07	0.86
oh_cost_10	0.16	0.60
oh_min_10	0.47	0.07

Table 27. Total Number and Significance Level Distribution of Triple-Way Interaction Coefficients in cost_max_min_simple_ia_only_5 Model

Model Name	Number of Coefficients	Percent of Total Variables		
		< 0.05	0.05 - 0.1	Insignificant
cost_max_min_simple_ia_only_5	99	0.96	0	0.04

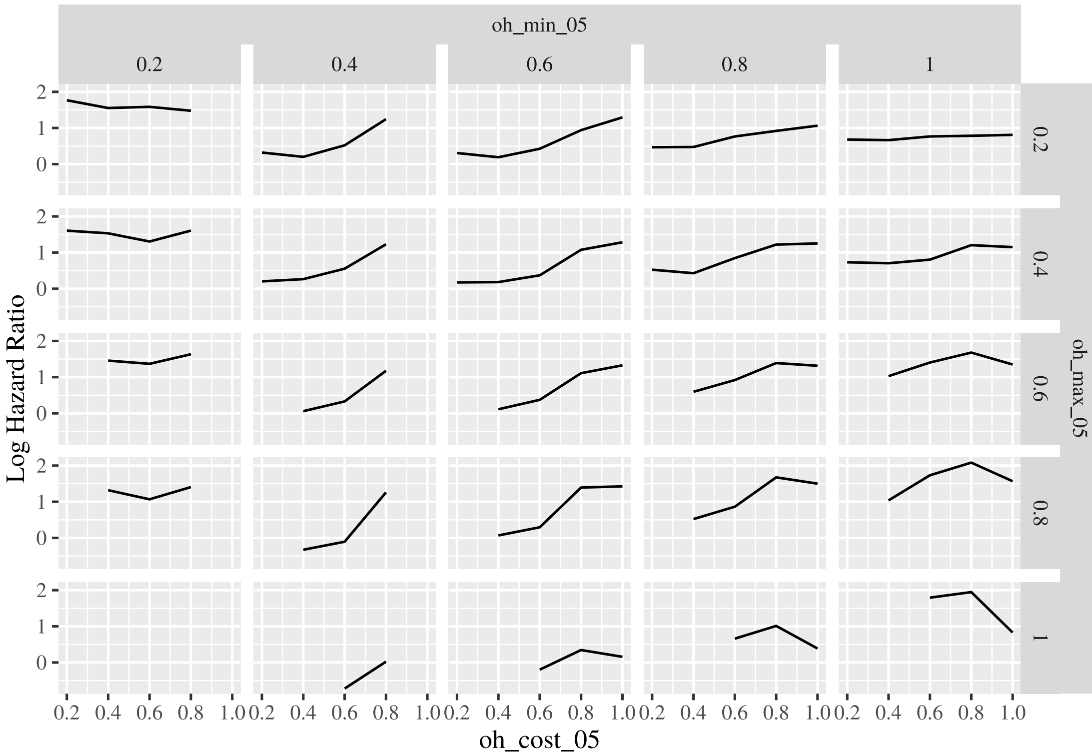

Figure 41. Triple-way interaction coefficients of cost_max_min_simple_ia_only_5 model

Unlike the plots generated from coefficients of the main model, there is no apparent trend that covers any of the three dimensions across whole plot.

A Bayesian regression model with monotonic effects is again run for the triple-way interaction coefficients and the estimates of monotonic effects are presented in Table 28. The Rhat values of the estimates are all below 1.05, so the

Table 28. Estimates From the Bayesian Monotonic Effects Model of the Relationships Between Triple-Way Interaction Coefficients and Ordinal Variables of cost_max_min_simple_ia_only_5 Model

Variable	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
Intercept	0.52	0.29	-0.01	1.21	1.00
mooh_cost_05	0.09	0.10	-0.09	0.28	1.00
mooh_max_05	0.00	0.09	-0.21	0.17	1.01
mooh_min_05	0.05	0.09	-0.18	0.21	1.00

chain converged sufficiently and estimates are usable. However, none of the estimates for the ordinal variables are significantly different from zero, confirming the results from the models without triple-way interactions.

So without controlling for the holding period and market returns for each of the ordinal overhang variables on the three reference prices, the monotonic relationships confirming disposition and reverse disposition effects cannot be observed.

CHAPTER 8

DISCUSSION

The results confirm that all null hypotheses in Chapter 3 (p. 50) are rejected:

- Propensity to sell is significantly and monotonically related to capital gains/losses overhang with respect to different reference points separately and also in interaction. There is a monotonically increasing or isotonic relationship between propensity to sell and overhang with respect to the purchase price, as an evidence of disposition effect. There is a monotonically decreasing or antitonic relationship between propensity to sell and overhang with respect to the maximum price that an investor experiences, as an evidence of reverse disposition effect. There is a monotonically decreasing or antitonic relationship between propensity to sell and overhang with respect to the minimum price that an investor experiences, as an evidence of reverse disposition effect.
- These monotonic relationships are also interrelated. Holding other things equal, as current price is closer down to the minimum reference point, the disposition effect with respect to the cost is less pronounced. Hence, the positive change in propensity to sell across rising overhang levels with respect to cost is to a lesser degree. Holding other things equal, as current price is closer up to the maximum reference point, the disposition effect with respect to the cost is less pronounced. Hence, the positive change in propensity to sell across rising overhang levels with respect to cost is to a lesser degree.
- The consistent monotonic relationships are only revealed when the overhang measures are discretized controlling for holding period and market state that the

investors experience, in a nested manner. So the perception of overhang by the investors is context-sensitive for the holding period and market state for all three reference prices. When the discretization is done in a simple manner without taking into account the holding period and market return of the investors, no consistent and significant monotonicity across different overhang levels is observed.

These findings clearly show that investors incorporate these three reference points into their decisions simultaneously. And their perception of overhang levels are formed within the contexts of holding period and market state. Furthermore, investors demonstrate disposition effect and reverse disposition effect simultaneously with respect to different reference points and these conflicting effects also interact.

As to the author's knowledge, this is the first study to incorporate three reference points simultaneously and find significant monotonic and interacting relationships. Possible failure of many previous studies to detect these effects may be rooted in how data are filtered and wrangled and measures are constructed.

First of all, in this thesis, positions are filtered so that there are no issues that may blur the perception or salience of those prices, such as multiple purchases or sales during the holding period of the position or any cash dividends or stock splits. Literature shows that less salience of prices exacerbates their role of acting as reference points for investors.

Purchase price itself is fixed as a reference point, while maximum and minimum prices for a position may set new high or low records in due course of the price path. However, the perception of the level of capital gains/losses overhang is formed within the hierarchical contexts of holding period and market state. When the overhang measures are constructed without considering this hierarchical context of the holding period and market state in alternative models, the strong findings of monotonic relationships in the main models completely vanish. And in this case, the resulting shapes are similar to the V-shape findings commonly reported in the literature.

Discretization of the variables included in disposition effect studies has a long tradition. This process is preferred due to the elasticity in discovering the relationships between the response and predictors and allows for different shapes to be revealed. However, the inclusion of triple-way interactions has never been tried as to the author's knowledge. One obstacle is the inflated number of covariates, a fact which makes the model computationally much harder to be executed. In this thesis, the obstacle is solved by using state-of-the-art high performance statistical tools that can execute algorithms parallelly and can scale efficiently with a very large number of predictors. Another obstacle is the absence of a standard method to test the multivariate monotonicity of the found coefficients. In this thesis, this issue is solved by creating a novel two-step method in which the coefficients estimated in Cox models are fed into a Bayesian regression model with monotonic effects. The coefficients are not treated as point estimates, but random samples are generated using the estimate and standard error of each coefficient, creating a synthetic dataset representing the confidence intervals of coefficient estimates. This approach is followed in order to be consistent with the philosophy of Bayesian approach which takes into account the whole distribution of parameters.

An important question is raised by the findings as to how and why investors demonstrate simultaneous disposition effect and reverse disposition effect in opposing directions with respect to different reference points.

The answer may be searched for in the literature for emotions in decision making, as covered in Subsection 2.3.6 (p. 39). According to the SP/A theory, individuals make choices according to two separate factors, namely dispositional and situational factors, simultaneously, and these factors can reinforce or oppose each other in terms of their effects on individual's decisions. So both risk-averse and risk-seeking choices coexist interactively in the behaviors (Lopes, 1987).

According to affect account theory, there are two emotional dimensions that act simultaneously on decisions. Anticipatory feelings dimension includes hope of earning and fear of losing associated with positive or negative returns respectively. Anticipated feelings of elation and disappointment are associated with decisions to realize gains and losses respectively. The balance of anticipatory feelings (hope and fear) and the balance of anticipated feelings (elation and disappointment) occur simultaneously as price evolves (Gärling et al., 2017).

Two-dimensional affect account can provide a consistent answer to the findings of this thesis. Overhang with respect to the cost is the actual level of gain or loss, realized or unrealized, an investor experiences and evokes feelings across the disappointment and elation continuum. Across the loss region, disappointment is evoked holding the investor from realizing the loss. Across the gain region, elation is evoked so that investor realizes the gain. Hence, an investor demonstrates disposition effect with respect to cost reference.

Overhang with respect to the maximum price that the investor has experienced shows how close the current price is to the investor's own historic high. Current price can be as high as that maximum price and as current price gets closer to that maximum or is at a new maximum level, feeling of hope, or greed, that the price can reach new maximum levels is evoked. That feeling can be likened to the hopeful or greedy feelings of conquerors or explorers having the desire to go beyond existing politic or geographic limits as they reach current limits. So holding other factors constant, the propensity to sell decreases as current price approaches (or is at) the investor's historic maximum price of the position, since the expectations that price may reach new maximum levels intensify. Hence, an investor demonstrates reverse disposition effect with respect to maximum reference.

Overhang with respect to the minimum price that the investor has experienced shows how close the current price is to the investor's own historic low. Current price can be as low as that minimum price and as current price gets closer to that minimum or is at a new minimum level, feeling of fear that the price can reach new minimum levels is evoked. That feeling can be likened to the fearful feelings of a retreating army at a battlefield or a soccer team that is consistently getting behind its opponent, in order not to risk losing the whole ground. So holding other factors constant, the propensity to sell increases as current price approaches (or is at) the investor's historic minimum price of the position, since the expectations that price may reach new minimum levels intensify. Hence, an investor demonstrates reverse disposition effect with respect to minimum reference.

The anticipatory feelings and the resulting reverse disposition effect with respect to maximum and minimum reference prices act in contrast to the anticipated feelings and the resulting disposition effect with respect to cost reference. The resulting complex balance from this contrasting effects at different overhang levels with respect to different reference points through the course of price movements may explain the V-shape evidence found frequently in disposition effect literature.

Various studies in the literature simplify this situation by including past returns within a fixed moving period of time, such as daily, weekly or monthly returns, as additional variables to show the relevance of momentum or contrarian investment strategies. The problem with preferring periodic returns across moving windows to capital gains/losses overhang with respect to reference points that are fixed or only updated on attaining new high or low records is that the results may not be robust to different window lengths. This fact is explained in detail in Section 5.7 (p. 94) and is illustrated with simulated data in Figure 13 (p. 95). The sign and magnitude of a return calculated on a fixed window length basis is highly dependent on the choice of the window length. And even if this measure is calculated by investors, the choice of window length may be different across investors. Furthermore, for most unsophisticated investors, tracking the moving reference price for the return calculation is mentally and technically hard or even impossible. However, taking initial price and maximum or minimum prices attained so far by the investor in a position is easier, since these prices are salient and can easily be remembered by the investor and are easier to maintain, since continuous updating is not always required. In the case of attaining new highs or lows continually in up or down rallies, maximum and minimum prices are frequently updated. As a result, the overhang between the current price and the reference is always close to or at zero, which is perceived as the highest overhang level for the maximum reference price and the lowest overhang level for the minimum reference price. So the direction and speed of price movement may not be consistent across different time horizons in most cases. However, the distance to the reference points are more robust and consistent in order to calculate and perceive. What matters for investors is how far they are away from the purchase price and highest and lowest marks that they experience personally in their stock positions.

The treatment of the overhang variables by controlling for the respective holding period and market state for each reference point also answers the question of how reference points are adapted. In many studies, an arbitrary chosen time period is used as a time frame to update the reference prices. The findings in this thesis show that maximum and minimum prices are updated only when an investor experiences new record high or low in a certain position, while the cost reference is not updated with new stock prices. The overhang levels are calculated accordingly. However, the perception of the level of overhangs is updated with respect to the evolving holding period and market returns, as explained in detail in Subsection 5.9.2 (p. 105) and illustrated in Figure 18 (p. 106). So for example, a certain numeric level of positive (negative) overhang can be perceived as relatively high (low) in a shorter time frame and relatively low (high) in a longer time frame. Similarly, a numeric level of overhang can be perceived as relatively high (low) when lower (higher) market return is experienced.

The future direction to build upon these findings might be to conduct experimental studies incorporating simulated data, real person subjects and various price paths that can reflect different overhang level combinations across three reference points for each investor. To test whether fear and hope or greed drive the reverse disposition effect when investor is closer to the extreme reference points, expected price levels may be required to be input frequently. While expected returns have been included in several studies, they are either naively calculated returns or are not tested for simultaneous disposition and reverse-disposition effects against multiple reference points. Hypothesized anticipatory feelings driving reverse-disposition effect may increase the probability that future price levels are expected to go beyond minimum and maximum prices experienced by an investor, when the investor is closer to those extreme reference points, controlling for the overhang with respect to the purchase price.

Another possible venue is to include investor level variables such as the total portfolio sizes of investors, diversification levels of the portfolios, shares of the positions in the investor portfolios and the experience levels of the investors in terms of number of transactions. Two major challenges in this expanded setting will be the access to data and adapting the Cox model and overhang calculations to partial sales, subsequent purchases and stock splits and cash dividends.

In this thesis, only the transaction data are used, as the case with most of the disposition effect studies. This causes a natural left truncation since the initial level of holdings of the investors are unknown. Furthermore, shares may be transferred between accounts of the same investor or different investors bypassing the order book and transactions. The inclusion of balance information along with the transaction data may be necessary to calculate the investor portfolio size related measures. Due to the highly confidential nature and the size of the data, an exhaustive dataset for all investors and a long period may be technically and bureaucratically hard to get.

In this thesis, only simple positions in which investors make a single purchase and either they do not sell any shares at all or make a single sale are filtered for. Intraday positions that are opened and closed on the same day, and positions that are opened in the last date of the time range of the dataset are also excluded, since dynamics of trade positions with an intraday round trip duration may not conform with disposition effect. Furthermore, stocks are selected so that no stock splits or cash dividends occur throughout the year. These filters are implemented in order to include only those positions in which purchase price is a simple, easily trackable, unambiguous and salient reference point. Excluding partial liquidations, the filters also allow the hold or sell status to be a dichotomous variable suitable for survival analysis. When complex trade positions involving intertwined multiple purchase and sale transactions are included in the analysis, the cost reference has to be updated continually after each transaction and a scheme such as first-in-first-out, last-in-first-out or average costing must be selected to guide the cost calculation. And also, in stock splits and cash dividends, reference prices have to be adjusted. That adjustment may be a quite complicated issue, since the date of split or dividend payout may be across a time span, or the date data and actual dates may be different, or individuals may not exert their preemptive rights at all. Apart from these calculation issues, since this continual update and adjustment task is too complicated for an average investor, they may still take the unadjusted prices as reference points, a fact which have to be controlled with additional variables. Widely accepted methods in disposition effect studies, such as Cox model, logistic regression or Odean's measure are suitable for dichotomous response variables which mark whether a position is sold or held. In complex trade positions, partial sales have to be included in the response variable, a fact which may create a methodological challenge.

Another possible venue for further research may be extending the methodology to futures market and incorporating short positions along with the long side of each contract. However, the margin requirements and leverage undertaken by the investors must also be included in the analysis in order to arrive at a more realistic calculation of overhang measures.

Disposition effect is a field of study in behavioral finance that attracts continuing interest over time. The result of the exhaustive literature review conducted for this thesis through Google Scholar, OpenCitations and Crossref (“Crossref”, n.d.; “Google Scholar”, n.d.; “Open Citation”, n.d.) and including working papers, presentations and theses along with published articles shows that number of studies in the field of disposition effect increased from three in 1998, to ten in 2001, 13 in 2005 and peaked at 77 in 2020[21]. The research activity decreased afterward, probably due to the effects of COVID-19 pandemics. The access to larger datasets, development of new statistical methodologies, progress in computing power and computational algorithms and incorporation of ideas and themes from the field of psychology drive the interest, innovation and complexity in disposition effect research. This thesis is intended to be a research study to bring innovation to the disposition effect field in terms of simultaneous treatment of multiple reference points, the context-sensitive hierarchical discretization approach, execution of very large models involving nearly 1000 covariates of triple-way interactions and the two-step method to test the multivariate monotonicity of coefficients across multiple ordinal variables. The proposed behavioral explanation of two-dimensional emotions needs further research to be confirmed.

[21] The numbers are corrected as much as possible for multiple versions of the same study with slightly differing names, or working papers that are later published as articles.

CHAPTER 9

CONCLUSION

Disposition effect is the propensity of investors to sell winning positions earlier and hold on to losing position longer. Reverse disposition effect, which is observed in later studies, is the propensity of investors to hold on to winning positions longer and sell losing positions earlier. While the disposition effect is first observed and mentioned by Schlarbaum et al. (1978), Shefrin and Statman (1985) are the first to define, test and bring a prospect theoretic explanation to the effect. From Kahneman and Tversky (1979), they borrowed the idea of S-shaped valuation function which is concave in the gain region, so that economic actors are risk-averse, and convex in the loss region, so that economic actors are risk-loving. The gain and loss is evaluated against a reference point. From Thaler (1985), they borrowed the idea of mental accounts in which gains and losses in different games may be segregated or integrated. The idea laid mostly dormant until 1998 due to the lack of transaction level stock market data and a lack of a methodology to test the existence of the effect. Odean (1998) ignited interest in the field, providing answers to these two issues by using a large transaction level dataset from a brokerage house and developing a measure comparing propensity of investors to sell winning and losing positions in a simple way. Weber and Camerer (1998) introduced experimental design, Grinblatt and Keloharju (2001) applied logistic regression and Genesove and Mayer (2001) brought survival analysis and Cox model into the research in the field. Feng and Seasholes (2005) and Shumway and Wu (2005) expanded the opportunities that parametric survival methods and Cox model can provide and also started to discretize numeric variables into ordered categories in order not to impose a certain shape for the relationships.

Recurring themes emerged in disposition effect studies, creating new questions to be answered. In some studies, the propensity to sell scheme that is higher towards the gain region and lower towards the loss region is reversed so that investors more readily close losing positions than they do winning positions. Hence, they demonstrate a reverse disposition effect. V-shaped relationships increasing in some regions and decreasing in others have been found between propensity to sell and certain explanatory variables, especially capital gains/losses overhang with respect to a reference point. The idea that reference points other than the purchase price may be used as anchors in investor decisions and how these reference points are formed and updated are continually explored in the literature. Especially the highest and lowest prices in an investor's experience or in a time frame are suggested as alternative reference points. Although existence of multiple reference points simultaneously is first mentioned by Kahneman (1992), the idea is still underexplored. While using reference points to evaluate the current situations in order to perceive gains and losses is one of the pillars of the genesis study of Shefrin and Statman (1985), how that reference point is visible or salient in order to modify the level of disposition effect is later visited in several studies. Shefrin and Statman (1985) mentioned the role of regret aversion for the disposition of investors to hold on to losing positions and thus incorporated emotions in the explanation. Summers and Duxbury (2007) focused more on the role of emotions in disposition effect. Borrowing ideas from the field of psychology, this is a promising area of research in the disposition effect field. In due course, frameworks alternative to prospect theory were proposed for explaining disposition effect, such as realization utility by Barberis and Xiong (2012), rational disposition effect through rebalancing by Dorn and Strobl (2010) and affect account by Gärling et al. (2017).

Discretization of numeric data into nominal or ordinal features is usually employed as a methodological and practical measure to allow for an elastic shaped relationship between the response and covariates. However, it may also be linked to a theoretical ground in psychology. The idea of categorical perception suggests that observed continuous stimuli perceived to be from different categories are distinguished better than stimuli perceived to be in the same category (Goldstone et al., 1996; Harnad, 1987).

This thesis mainly proposes and explores the idea that frequently observed seemingly non-monotonic relationship between the propensity to sell and overhang with respect to the selected reference point, which may flatten or reverse in the extremes or has a V-shape, is possibly due to omission of multiple reference points acting simultaneously and in interaction on investor decisions and also possibly due to inappropriate handling of overhang variables, which does not take into account the time horizon and market state contexts. So the main hypothesis is that multiple reference points simultaneously have monotonic effects on the propensity to sell. The three reference points included are the original purchase price and maximum and minimum prices that investors experience throughout the holding period. These effects can be in the positive direction (isotonic) or negative direction (antitonic). An isotonic relationship suggests the existence of disposition effect and an antitonic relationship suggests the existence of reverse disposition effect. Furthermore, the intensity of those monotonic relationships may be altered by interactions between them.

The main dataset analyzed in the study is the whole bilateral transactions realized on BIST between 1 February 2005 and 30 December 2005. Some ancillary datasets are also used, including data on capital increases and dividend payments and BIST index values in ten second intervals. Main unit of data is a position which describes an investor's transactions in a certain stock share. So a total of 39,749,152 bilateral transactions made in 235 transaction dates between 1 February 2005 and 30 December 2005 on 313 separate stock codes are first split into buyer and seller sides to create a unilateral transaction set of 79,498,304. And they are later organized into 4,761,393 positions held by 552,389 unique investor accounts.

Major challenges are addressed and solved in this thesis. The first challenge is filtering the dataset for positions that can be handled better with the statistical models and for positions which investors can have a clear and salient view of and track purchase prices as the cost of their positions and reference points. This filtering is done in two steps. In one step, only those stocks which never had any stock splits or dividend payouts in 2005 are selected. So the purchase price does not need to be adjusted for these events. Such an adjustment may create a confusion between the historic and adjusted purchase prices for the investors. In the other step, only those positions in which a single purchase exists and if any sale occurs, it occurs at one time are filtered. Each buy and sell order is executed either as a single transaction or multiple transactions that occur on the same second and at the same price. Intraday positions are also ruled out.

In line with disposition effect studies, for each filtered position, a separate record is generated for each day between and including the purchase and sale dates for closed positions and between and including the purchase data and the last date of the time range of the dataset (or the last transaction date for a stock) for unclosed positions. Each record is unique for the combination of a position and date. For each record, the response variable to be predicted is the status of whether the position is closed, when the status is 1, or it is held, when the status is 0. Total number of records in the inflated dataset that forms the basis for calculation of the covariates to be modeled is 8,608,331.

Covariates are calculated in three steps for each record: First, three separate reference points and market indices corresponding to the reference points are tracked. Then overhang values, realized or unrealized gains or losses with respect to reference points, and respective market returns are calculated. And at the last step, the numeric overhang values are discretized into ordinal variables incorporating the time horizon and market return contexts.

Calculation and updating of precise reference points and their corresponding market index values also pose a challenge. While the cost reference is straightforward since there is a single purchase at a single price and no stock splits or dividend payouts occur, the issue is more complicated for maximum and minimum reference prices. The daily high and low prices for each stock are summarized from the transactions dataset, and the closest market index values corresponding to the time in seconds the high and low prices are first attained are also recorded. For all days after the initial purchase, and before the date of sale if any, the maximum and minimum prices are updated simply if any new extreme price is encountered for each position separately. However, for the dates of purchases and sales, seconds level precision is considered to determine the maximum and minimum prices that the investor encounters on that day, since within a single day each position that is started or ended at different time points, different maximum and minimum prices may be experienced as prices fluctuate. The transactions dataset is summarized at the seconds level for daily cumulative maximum and minimum prices in two separate runs, once for determining the maximum/minimum after each second until the end of the day for the purchase dates and then for determining the maximum/minimum before each second from the beginning of the day for the sale dates. The corresponding closest market index values for the time maximum and minimum prices are attained are also recorded. Final maximum and minimum reference prices to be used are calculated by getting the cumulative maximum and minimum values for each position. An algorithm for that level of precision and detail is developed for the first time as to the author's knowledge.

Calculation of overhang values are done using the actual transaction prices of sales. For days with no transactions in the position, overhang values are calculated using the maximum prices of the day, since for long positions, the realization opportunity foregone by the investors can be best evaluated at the highest price attained during the day. The market returns between the closest second that any of the reference prices is formed or updated and the closest second daily maximum price is attained or the sale transaction is realized are also calculated. Raw market index values are in ten seconds detail, while transactions are in seconds detail.

Another challenge is how numeric overhang values and market returns are discretized into context-sensitive ordinal variables. Any certain level of positive or negative overhang may be perceived as relatively high or low considering the time horizon the overhang occurs and the perceived level of market return in the same time horizon. According to the MUHBOD method proposed here, first holding periods in days across all records are discretized into categories following the equal frequency method. At the next level, market returns are discretized with the same method for each subset of the records with a holding period category. So the cutting points of discretized market returns may be different for each holding period category. At the last level, overhang measures are discretized for each subset of the records with a combination of holding period and market return category. So the cutting points of discretized overhangs may be different for the intersection of each holding period and market return category. For example, a certain level of positive return may be perceived as a member of a higher category in the short term, while it may be perceived as a member of a lower category in the long term. Similarly, the same return may be perceived as a member of a higher category in a lower market return, while it may be perceived as a member of a lower category in a higher market return.

This method transforms returns and overhangs into context-sensitive categories in a multidimensional and hierarchical approach. Another advantage is the balanced distribution of overhang cases across the intersection of time horizon and market state contexts.

Five methodologies are used in tandem for testing the monotonicity of the relationship of propensity of sale to overhang levels with respect to multiple reference points. In the first step, extended Cox regression model and enhanced Odean's measure, a method that is first proposed here, are conducted in order to extract the coefficients and ratios for measuring the level of the propensity of sale across the ordinal levels of overhang with respect to three variables. Cox models are conducted in two versions, one by incorporating overhang variables as main effects, and the other by incorporating the triple-way interactions of those ordinal overhang variables. KM survival curves method is conducted in order to confirm Cox results visually. KM survival curves summarize the survival rates starting from 1 and converging to 0 as step functions over time horizons. The curves are calculated for each triple-way combination of the ordinal levels of overhang with respect to three reference points. In the second step, monotonicity of the coefficients or ratios across ordinal overhang levels are tested either in univariate or multivariate settings. Kendall's Tau is utilized in order to get the rank correlation between the coefficients of each variable in the main effects Cox models and the ordinal levels of those variables. Another challenge is to test the multivariate monotonicity of the coefficients of triple-way interactions in Cox models and ratios in enhanced Odean's measure method. The requirement here is to test the existence of a monotonic step function with respect to ordinal measures. Most methods treat ordinal variables as dichotomous dummy variables, as in the case with the Cox model, or numeric variables with equal intervals. While isotonic regression methods exist for fitting stepped relationship between numeric responses and numeric features, the only option to model monotonic stepped relationship between numeric responses and ordinal features is Bayesian regression model with monotonic effects as proposed and detailed in Bürkner and Charpentier (2020). As opposed to standard linear models which report proportional relationship between the response and covariates as summarized by a single coefficient, this model allows for unequal effects across categories provided that these effects are in the same direction. Another important feature of the model is to allow for interactions between monotonic effects of ordinal variables, similar to the interactions between numeric or dichotomous variables in standard linear models.

Monotonic effects models are first conducted for the Cox regression coefficients of triple-way interactions and ratios representing enhanced Odean's measure. In line with the general principle of Bayesian analysis which emphasizes modeling the whole distribution instead of creating point estimates, random samples are created using the coefficient estimates and standard errors from normal distribution for the Cox model coefficients. Samples are also created using the number of cases for calculating each ratio from binomial distribution for the enhanced Odean's measures. Furthermore, additional monotonic effects models are conducted to incorporate the two interactions, one between the ordinal overhang with respect to the cost reference and the ordinal overhang with respect to the maximum reference, the other between the ordinal overhang with respect to the cost reference and the ordinal overhang with respect to the minimum reference. Inclusion of these secondary effects allows for understanding more complex relationships regarding disposition and reverse disposition effects.

Cox models are in four versions for each set, involving five or ten category ordinal variables and incorporating only the main variables or only triple-way interactions. All models are significant in terms of log likelihood test, score test and Wald test.

For each ordinal overhang variable, there are two versions, one is the simple discretized value using equal frequency, the other one is formed by the MUHBOD method explained above by controlling for the time horizon and market state contexts. Using the ten category variables, a separate univariate model is run for six variables, two versions of ordinal variables for each reference point. While most of the coefficients are statistically significant, rank correlations do not exhibit significant or strong monotonicity for the coefficients of most models, except for the model with the time horizon and market state controlled MUHBOD version of ordinal overhang variable with respect to maximum reference.

First set of multivariate Cox models incorporate the five and ten category versions of the ordinal overhang variables that are controlled for time horizon and market state. Each model includes a triplet of five or ten category ordinal overhang variables for each of the reference points. For each set of variables, first a simple model including only main effects is run and then a larger model including only triple-way interactions of the ordinal variables is conducted. There are 12 covariates in the five category version and 27 covariates in the ten category version of the main effects only models without interactions. There are 124 covariates in the five category version and 998 covariates in the ten category version of the triple-way interaction models. All coefficients of both five and ten category main effect Cox models are significant, and the coefficients of each ordinal variable exhibit strong monotonicity across levels. Variables which do not exhibit monotonicity when included in a univariate model, are in a strong monotonic relationship with the propensity to sell when they are combined in multivariate models. The results of the main effects multivariate Cox models show that there is a monotonically increasing, or isotonic, relationship between the propensity to sell and cost based ordinal variables, documenting disposition effect. The relationships between the propensity to sell and ordinal variables based on maximum and minimum references are monotonically decreasing, or antitonic. Hence, they document the existence of the reverse disposition effect.

These results are further confirmed with supporting methods and more complex models. Five and ten category triple-way interaction models allow for controlling all three variables for each other simultaneously. A great portion of the large number of coefficients in these models are significant. From each of the five and ten category large Cox models, random samples are constructed for each coefficient using estimates and standard errors. These samples are fed into a Bayesian regression model with monotonic effects in which ordinal variables are independent variables keeping their ordinal nature and not being converted to numeric or dummy variables. The size of the effect is not summarized as a single coefficient estimate as in the case of standard regression models. Instead, an average positive or negative effect for each ordinal feature is estimated. Also normalized positive differences between sequential categories of the ordinal variable, that sum up to 1 are modeled as ratios. Monotonic effects models demonstrate that cost based ordinal overhang variable has a significant positive monotonic effect, while the maximum and minimum ordinal overhang variables have significant negative monotonic effects, confirming the results of simpler Cox models. Enhanced Odean's measure ratios across combinations of three five category ordinal overhang variables and the five category discretized holding period as a fourth dimension are also sampled using binomial distribution for a separate confirmatory Bayesian regression with monotonic effects. The results confirm those of the previous monotonic effects models with significant positive monotonic effect for the cost based ordinal overhang variable and significant negative monotonic effects for the maximum and minimum based ordinal overhang variables.

Visual inspection of the coefficients from five and ten category triple-way interaction Cox models, enhanced Odean's measures calculated across five category ordinal variables and KM survival curves of five category ordinal variables are used as complementary methods. They also confirm the co-existence of disposition effect with respect to the cost based overhang variable and reverse disposition effect with respect to the maximum and minimum based overhang variables.

However, a secondary effect is also apparent in the visual methods: When current price is furthest from both the minimum and maximum reference points, the disposition effect is at its strongest. In that case, the propensity to sell increases more with higher cost based overhang categories. As the price is closer to the minimum or maximum reference points, or closer to both which is possible when prices fluctuate in a narrower band, the strength of the disposition effect diminishes and the increase (decrease) over higher (lower) cost based overhang categories flattens. A more complex version of the Bayesian regression models are run in order to test these secondary effects. In addition to the main effects of ordinal variables, two interaction terms are also included. First interaction term is between cost and maximum based ordinal variables while the second interaction term is between cost and minimum based ordinal variables. These models are conducted for the coefficients from five and ten category Cox models. The results confirm the visual findings. The category when current price is closest to the minimum reference is the smallest one, and higher categories are further away from this reference. The category when current price is closest to the maximum reference is the largest one, so higher categories are closer to this reference. The main effects show the same direction of monotonicity: Positive for cost based overhang ordinal, negative for maximum and minimum based overhang ordinals. The effect of interaction between the cost and maximum based ordinals is negative: So for larger categories closer to the maximum reference, the strength of disposition effect diminishes. The effect of interaction between the cost and minimum based ordinals is negative: So for smaller categories closer to the minimum reference, the strength of disposition effect diminishes.

Alternative models which include the versions of ordinal variables that are discretized from numeric features in a straightforward way using equal frequencies and without taking into account the hierarchical nature of time horizon and market state contexts are also conducted. However, they exhibit no significant monotonic relationship across overhang levels. So the failure to document significant and strong unidirectional effects arises from the lack of incorporating the time horizon and market return contexts in creating the features for overhang or failure to test the effects of multiple reference points simultaneously.

To summarize, investors incorporate three reference points simultaneously in their decisions. Cost reference is fixed, while maximum and minimum references can be updated as new extremes are experienced by the investor. Furthermore, the level of the overhang with respect to these reference points are perceived within the context of the time horizon and the market state. Investors exhibit simultaneous and interacting disposition effect with respect to cost reference and reverse disposition effect with respect to maximum and minimum references separately. The disposition effect is stronger as the current price is perceived to be further away from maximum and/or minimum references and diminishes as the current price is perceived to be closer to maximum and/or minimum references. The co-existence of multiple effects that contrast each other may be explained by affect account theory by Gärling et al. (2017). According to affect account theory, there are two simultaneous emotional dimensions. Anticipatory feelings dimension includes hope of earning and fear of losing with future returns, while anticipated feelings of elation and disappointment are associated with decisions to realize the current overhang. Anticipated feelings dimension can be related to the disposition effect with respect to the cost reference. Investors feel elation in gains and disappointment in losses. Anticipatory feelings dimension can be related to the reverse disposition effect with respect to the maximum and minimum references. Investors feel the fear of losing more when prices are closer down to the minimum references. And they feel the hope of earning more when prices are closer up to the maximum reference.

For a future direction, these findings and the proposed explanation can be confirmed with experimental studies. While the realized or unrealized gains/losses are straightforward to observe for the anticipated feelings dimension, investors' self reported expected prices can be collected in order to justify the existence of fear and hope near minimum or maximum references for the anticipatory feelings dimension to explain reverse disposition effect, in case that effect is observed experimentally.

This thesis is intended to address several challenges through innovations in data wrangling routines, a discretization method and a two-step modeling solution to test for multivariate monotonic relationships between numeric response and ordinal covariates.

APPENDIX A

NON-ORDINAL VARIABLES

Variable Name	Data Class	Definition
pos_no2	numeric	Unique position number, investor stock combination
new_min_day	integer	Day a new minimum was attained. 0 if not
xu_min	numeric	BIST 100 return index at the day of last minimum
close_min	numeric	Closing price at the day of last minimum
days	integer	Days since purchase. 0 at the purchase day
new_max_day	integer	Day a new maximum was attained. 0 if not
xu_max	numeric	BIST 100 index at the day of last maximum
close_max	numeric	Closing price at the day of last maximum
date1	Date	Date of the day
xu_close	numeric	BIST 100 index closing level
stock	character	Stock ticker
min	numeric	Minimum price of the day
max	numeric	Maximum price of the day
close	numeric	Closing price of the day
xu_trans_min	numeric	BIST 100 index corresponding to the time minimum price of the day is attained
xu_trans_max	numeric	BIST 100 index corresponding to the time maximum price of the day is attained
qty_sum	integer	Total trade quantity of stock at the day
proceeds_sum	numeric	Total trade value of stock at the day
dateno	integer	Number of trading days since the beginning of dataset
member	factor	Securities firm for the transaction
buy_qty	numeric	Total purchase quantity at the day
buy_proceeds	numeric	Total purchase value at the day
buy_xu	numeric	BIST 100 index corresponding to purchases, weighted with quantities
sell_qty	numeric	Total sales quantity at the day
sell_proceeds	numeric	Total sales proceeds at the day
sell_xu	numeric	BIST 100 index corresponding to sales, weighted with quantities
N	numeric	Number of total transactions in position. 1 or 2
cum_qty	numeric	Cumulative net quantity until the day (after sales of the day if any)
cum_proceeds	numeric	Cumulative net proceeds until the day (after sales of the day if any)
cum_xu	numeric	Cumulative net BIST 100 index corresponding to purchases and sales, weighted with quantities
cum_proceeds_bs	numeric	Cumulative net quantity until the day (before sales of the day if any)

(continued)

Variable Name	Data Class	Definition
cum_xu_bs	numeric	Cumulative net BIST 100 index corresponding to purchases and sales, weighted with quantities (before sales of the day if any)
cum_qty_bs	numeric	Cumulative net proceeds until the day (before sales of the day if any)
ref_cost	numeric	Purchase cost as reference price
xu_cost	numeric	BIST 100 index corresponding to purchase cost
av_sell_price	numeric	Average selling price of the day if any
av_sell_xu	numeric	BIST 100 index corresponding to average selling price of the day if any
ref_min	numeric	Last minimum price since start of transaction as reference price
ref_max	numeric	Last maximum price since start of transaction as reference price
current_price	numeric	If sales occured, average sales price, else maximum price of the day
xu	numeric	BIST 100 index corresponding to current price
oh_cost	numeric	Logarithmic overhang over cost reference
oh_min	numeric	Logarithmic overhang over minimum reference
oh_max	numeric	Logarithmic overhang over maximum reference
status	integer	If any sales on day, 1, else 0
days_max	integer	Days since last maximum
days_min	integer	Days since last minimum
oh_xu_cost	numeric	Logarithmic BIST 100 overhang on xu_cost
oh_xu_max	numeric	Logarithmic BIST 100 overhang on xu_max
oh_xu_min	numeric	Logarithmic BIST 100 overhang on xu_min
pos_current	numeric	Current size of position
profit	logical	TRUE when profit on cost
loss	logical	TRUE when loss on cost
days_start	integer	Number of days since the beginning of position decremented with one, for counting process in Cox models

APPENDIX B

COST_MAX_MIN_5 MODEL ODEAN'S MEASURES

Cost	Max	Min	Days	Ratio	Cost	Max	Min	Days	Ratio	Cost	Max	Min	Days	Ratio
0.2	0.2	0.2	0.2	0.54	0.2	0.2	0.2	0.4	0.09	0.2	0.2	0.2	0.6	0.03
0.2	0.2	0.2	0.8	0.01	0.2	0.2	0.2	1	0.01	0.2	0.2	0.4	0.2	0.09
0.2	0.2	0.4	0.4	0.02	0.2	0.2	0.4	0.6	0.01	0.2	0.2	0.4	0.8	0.01
0.2	0.2	0.4	1	0.00	0.2	0.2	0.6	0.2	0.06	0.2	0.2	0.6	0.4	0.01
0.2	0.2	0.6	0.6	0.01	0.2	0.2	0.6	0.8	0.01	0.2	0.2	0.6	1	0.00
0.2	0.2	0.8	0.2	0.04	0.2	0.2	0.8	0.4	0.01	0.2	0.2	0.8	0.6	0.01
0.2	0.2	0.8	0.8	0.00	0.2	0.2	0.8	1	0.00	0.2	0.2	1	0.2	0.03
0.2	0.2	1	0.4	0.01	0.2	0.2	1	0.6	0.01	0.2	0.2	1	0.8	0.00
0.2	0.2	1	1	0.00	0.2	0.4	0.2	0.2	0.33	0.2	0.4	0.2	0.4	0.06
0.2	0.4	0.2	0.6	0.02	0.2	0.4	0.2	0.8	0.01	0.2	0.4	0.2	1	0.00
0.2	0.4	0.4	0.2	0.04	0.2	0.4	0.4	0.4	0.02	0.2	0.4	0.4	0.6	0.01
0.2	0.4	0.4	0.8	0.01	0.2	0.4	0.4	1	0.00	0.2	0.4	0.6	0.2	0.03
0.2	0.4	0.6	0.4	0.01	0.2	0.4	0.6	0.6	0.01	0.2	0.4	0.6	0.8	0.00
0.2	0.4	0.6	1	0.00	0.2	0.4	0.8	0.2	0.02	0.2	0.4	0.8	0.4	0.01
0.2	0.4	0.8	0.6	0.01	0.2	0.4	0.8	0.8	0.00	0.2	0.4	0.8	1	0.00
0.2	0.4	1	0.2	0.03	0.2	0.4	1	0.4	0.01	0.2	0.4	1	0.6	0.01
0.2	0.4	1	0.8	0.00	0.2	0.4	1	1	0.00	0.2	0.6	0.2	0.2	0.20
0.2	0.6	0.2	0.4	0.05	0.2	0.6	0.2	0.6	0.02	0.2	0.6	0.2	0.8	0.01
0.2	0.6	0.2	1	0.00	0.2	0.6	0.4	0.2	0.03	0.2	0.6	0.4	0.4	0.02
0.2	0.6	0.4	0.6	0.01	0.2	0.6	0.4	0.8	0.01	0.2	0.6	0.4	1	0.00
0.2	0.6	0.6	0.2	0.02	0.2	0.6	0.6	0.4	0.02	0.2	0.6	0.6	0.6	0.01
0.2	0.6	0.6	0.8	0.01	0.2	0.6	0.6	1	0.00	0.2	0.6	0.8	0.2	0.02
0.2	0.6	0.8	0.4	0.01	0.2	0.6	0.8	0.6	0.01	0.2	0.6	0.8	0.8	0.00
0.2	0.6	0.8	1	0.00	0.2	0.6	1	0.2	0.03	0.2	0.6	1	0.4	0.01
0.2	0.6	1	0.6	0.01	0.2	0.6	1	0.8	0.00	0.2	0.6	1	1	0.00
0.2	0.8	0.2	0.2	0.09	0.2	0.8	0.2	0.4	0.03	0.2	0.8	0.2	0.6	0.01
0.2	0.8	0.2	0.8	0.01	0.2	0.8	0.2	1	0.00	0.2	0.8	0.4	0.2	0.02
0.2	0.8	0.4	0.4	0.01	0.2	0.8	0.4	0.6	0.01	0.2	0.8	0.4	0.8	0.01
0.2	0.8	0.4	1	0.00	0.2	0.8	0.6	0.2	0.02	0.2	0.8	0.6	0.4	0.01
0.2	0.8	0.6	0.6	0.01	0.2	0.8	0.6	0.8	0.01	0.2	0.8	0.6	1	0.01
0.2	0.8	0.8	0.2	0.02	0.2	0.8	0.8	0.4	0.01	0.2	0.8	0.8	0.6	0.01
0.2	0.8	0.8	0.8	0.01	0.2	0.8	0.8	1	0.00	0.2	0.8	1	0.2	0.02
0.2	0.8	1	0.4	0.02	0.2	0.8	1	0.6	0.01	0.2	0.8	1	0.8	0.01
0.2	0.8	1	1	0.00	0.2	1	0.2	0.2	0.07	0.2	1	0.2	0.4	0.02
0.2	1	0.2	0.6	0.01	0.2	1	0.2	0.8	0.01	0.2	1	0.2	1	0.00
0.2	1	0.4	0.2	0.02	0.2	1	0.4	0.4	0.01	0.2	1	0.4	0.6	0.00
0.2	1	0.4	0.8	0.02	0.2	1	0.4	1	0.01	0.2	1	0.6	0.2	0.03
0.2	1	0.6	0.4	0.00	0.2	1	0.6	0.6	0.00	0.2	1	0.6	0.8	0.01
0.2	1	0.6	1	0.00	0.2	1	0.8	0.2	0.02	0.2	1	0.8	0.4	0.00
0.2	1	0.8	0.6	0.00	0.2	1	0.8	0.8	0.00	0.2	1	0.8	1	0.00
0.2	1	1	0.2	0.03	0.2	1	1	0.4	0.00	0.2	1	1	0.6	0.00
0.2	1	1	0.8	0.01	0.2	1	1	1	0.01	0.4	0.2	0.2	0.2	0.43
0.4	0.2	0.2	0.4	0.08	0.4	0.2	0.2	0.6	0.03	0.4	0.2	0.2	0.8	0.01
0.4	0.2	0.2	1	0.00	0.4	0.2	0.4	0.2	0.09	0.4	0.2	0.4	0.4	0.03
0.4	0.2	0.4	0.6	0.01	0.4	0.2	0.4	0.8	0.01	0.4	0.2	0.4	1	0.00
0.4	0.2	0.6	0.2	0.06	0.4	0.2	0.6	0.4	0.02	0.4	0.2	0.6	0.6	0.01
0.4	0.2	0.6	0.8	0.00	0.4	0.2	0.6	1	0.00	0.4	0.2	0.8	0.2	0.05
0.4	0.2	0.8	0.4	0.02	0.4	0.2	0.8	0.6	0.01	0.4	0.2	0.8	0.8	0.01
0.4	0.2	0.8	1	0.01	0.4	0.2	1	0.2	0.05	0.4	0.2	1	0.4	0.02
0.4	0.2	1	0.6	0.01	0.4	0.2	1	0.8	0.00	0.4	0.2	1	1	0.01
0.4	0.4	0.2	0.2	0.33	0.4	0.4	0.2	0.4	0.06	0.4	0.4	0.2	0.6	0.02
0.4	0.4	0.2	0.8	0.01	0.4	0.4	0.2	1	0.00	0.4	0.4	0.4	0.2	0.06
0.4	0.4	0.4	0.4	0.02	0.4	0.4	0.4	0.6	0.01	0.4	0.4	0.4	0.8	0.01
0.4	0.4	0.4	1	0.00	0.4	0.4	0.6	0.2	0.03	0.4	0.4	0.6	0.4	0.02

(continued)

Cost	Max	Min	Days	Ratio	Cost	Max	Min	Days	Ratio	Cost	Max	Min	Days	Ratio
0.4	0.4	0.6	0.6	0.01	0.4	0.4	0.6	0.8	0.00	0.4	0.4	0.6	1	0.00
0.4	0.4	0.8	0.2	0.03	0.4	0.4	0.8	0.4	0.01	0.4	0.4	0.8	0.6	0.01
0.4	0.4	0.8	0.8	0.00	0.4	0.4	0.8	1	0.00	0.4	0.4	1	0.2	0.04
0.4	0.4	1	0.4	0.01	0.4	0.4	1	0.6	0.00	0.4	0.4	1	0.8	0.00
0.4	0.4	1	1	0.00	0.4	0.6	0.2	0.2	0.23	0.4	0.6	0.2	0.4	0.04
0.4	0.6	0.2	0.6	0.02	0.4	0.6	0.2	0.8	0.01	0.4	0.6	0.2	1	0.00
0.4	0.6	0.4	0.2	0.04	0.4	0.6	0.4	0.4	0.02	0.4	0.6	0.4	0.6	0.01
0.4	0.6	0.4	0.8	0.01	0.4	0.6	0.4	1	0.00	0.4	0.6	0.6	0.2	0.03
0.4	0.6	0.6	0.4	0.01	0.4	0.6	0.6	0.6	0.01	0.4	0.6	0.6	0.8	0.01
0.4	0.6	0.6	1	0.00	0.4	0.6	0.8	0.2	0.02	0.4	0.6	0.8	0.4	0.01
0.4	0.6	0.8	0.6	0.01	0.4	0.6	0.8	0.8	0.01	0.4	0.6	0.8	1	0.00
0.4	0.6	1	0.2	0.03	0.4	0.6	1	0.4	0.01	0.4	0.6	1	0.6	0.01
0.4	0.6	1	0.8	0.01	0.4	0.6	1	1	0.01	0.4	0.8	0.2	0.2	0.14
0.4	0.8	0.2	0.4	0.03	0.4	0.8	0.2	0.6	0.01	0.4	0.8	0.2	0.8	0.01
0.4	0.8	0.2	1	0.00	0.4	0.8	0.4	0.2	0.03	0.4	0.8	0.4	0.4	0.02
0.4	0.8	0.4	0.6	0.01	0.4	0.8	0.4	0.8	0.01	0.4	0.8	0.4	1	0.01
0.4	0.8	0.6	0.2	0.02	0.4	0.8	0.6	0.4	0.01	0.4	0.8	0.6	0.6	0.01
0.4	0.8	0.6	0.8	0.01	0.4	0.8	0.6	1	0.00	0.4	0.8	0.8	0.2	0.02
0.4	0.8	0.8	0.4	0.01	0.4	0.8	0.8	0.6	0.01	0.4	0.8	0.8	0.8	0.01
0.4	0.8	0.8	1	0.01	0.4	0.8	1	0.2	0.02	0.4	0.8	1	0.4	0.01
0.4	0.8	1	0.6	0.01	0.4	0.8	1	0.8	0.01	0.4	0.8	1	1	0.01
0.4	1	0.2	0.2	0.05	0.4	1	0.2	0.4	0.02	0.4	1	0.2	0.6	0.01
0.4	1	0.2	0.8	0.01	0.4	1	0.2	1	0.00	0.4	1	0.4	0.2	0.02
0.4	1	0.4	0.4	0.02	0.4	1	0.4	0.6	0.01	0.4	1	0.4	0.8	0.01
0.4	1	0.4	1	0.01	0.4	1	0.6	0.2	0.02	0.4	1	0.6	0.4	0.02
0.4	1	0.6	0.6	0.01	0.4	1	0.6	0.8	0.01	0.4	1	0.6	1	0.01
0.4	1	0.8	0.2	0.02	0.4	1	0.8	0.4	0.02	0.4	1	0.8	0.6	0.01
0.4	1	0.8	0.8	0.01	0.4	1	0.8	1	0.01	0.4	1	1	0.2	0.02
0.4	1	1	0.4	0.02	0.4	1	1	0.6	0.01	0.4	1	1	0.8	0.01
0.4	1	1	1	0.01	0.6	0.2	0.2	0.2	0.38	0.6	0.2	0.2	0.4	0.07
0.6	0.2	0.2	0.6	0.03	0.6	0.2	0.2	0.8	0.01	0.6	0.2	0.2	1	0.00
0.6	0.2	0.4	0.2	0.18	0.6	0.2	0.4	0.4	0.04	0.6	0.2	0.4	0.6	0.02
0.6	0.2	0.4	0.8	0.01	0.6	0.2	0.4	1	0.00	0.6	0.2	0.6	0.2	0.14
0.6	0.2	0.6	0.4	0.03	0.6	0.2	0.6	0.6	0.02	0.6	0.2	0.6	0.8	0.01
0.6	0.2	0.6	1	0.01	0.6	0.2	0.8	0.2	0.11	0.6	0.2	0.8	0.4	0.04
0.6	0.2	0.8	0.6	0.02	0.6	0.2	0.8	0.8	0.02	0.6	0.2	0.8	1	0.01
0.6	0.2	1	0.2	0.09	0.6	0.2	1	0.4	0.03	0.6	0.2	1	0.6	0.02
0.6	0.2	1	0.8	0.01	0.6	0.2	1	1	0.02	0.6	0.4	0.2	0.2	0.32
0.6	0.4	0.2	0.4	0.06	0.6	0.4	0.2	0.6	0.02	0.6	0.4	0.2	0.8	0.01
0.6	0.4	0.2	1	0.00	0.6	0.4	0.4	0.2	0.10	0.6	0.4	0.4	0.4	0.03
0.6	0.4	0.4	0.6	0.02	0.6	0.4	0.4	0.8	0.01	0.6	0.4	0.4	1	0.00
0.6	0.4	0.6	0.2	0.06	0.6	0.4	0.6	0.4	0.03	0.6	0.4	0.6	0.6	0.01
0.6	0.4	0.6	0.8	0.01	0.6	0.4	0.6	1	0.00	0.6	0.4	0.8	0.2	0.05
0.6	0.4	0.8	0.4	0.02	0.6	0.4	0.8	0.6	0.01	0.6	0.4	0.8	0.8	0.01
0.6	0.4	0.8	1	0.01	0.6	0.4	1	0.2	0.06	0.6	0.4	1	0.4	0.02
0.6	0.4	1	0.6	0.01	0.6	0.4	1	0.8	0.01	0.6	0.4	1	1	0.01
0.6	0.6	0.2	0.2	0.21	0.6	0.6	0.2	0.4	0.04	0.6	0.6	0.2	0.6	0.01
0.6	0.6	0.2	0.8	0.01	0.6	0.6	0.2	1	0.00	0.6	0.6	0.4	0.2	0.09
0.6	0.6	0.4	0.4	0.02	0.6	0.6	0.4	0.6	0.01	0.6	0.6	0.4	0.8	0.01
0.6	0.6	0.4	1	0.00	0.6	0.6	0.6	0.2	0.06	0.6	0.6	0.6	0.4	0.02
0.6	0.6	0.6	0.6	0.01	0.6	0.6	0.6	0.8	0.01	0.6	0.6	0.6	1	0.00
0.6	0.6	0.8	0.2	0.05	0.6	0.6	0.8	0.4	0.02	0.6	0.6	0.8	0.6	0.01
0.6	0.6	0.8	0.8	0.01	0.6	0.6	0.8	1	0.00	0.6	0.6	1	0.2	0.05
0.6	0.6	1	0.4	0.02	0.6	0.6	1	0.6	0.01	0.6	0.6	1	0.8	0.01
0.6	0.6	1	1	0.01	0.6	0.8	0.2	0.2	0.15	0.6	0.8	0.2	0.4	0.03
0.6	0.8	0.2	0.6	0.02	0.6	0.8	0.2	0.8	0.01	0.6	0.8	0.2	1	0.00
0.6	0.8	0.4	0.2	0.06	0.6	0.8	0.4	0.4	0.02	0.6	0.8	0.4	0.6	0.01
0.6	0.8	0.4	0.8	0.01	0.6	0.8	0.4	1	0.00	0.6	0.8	0.6	0.2	0.04
0.6	0.8	0.6	0.4	0.02	0.6	0.8	0.6	0.6	0.01	0.6	0.8	0.6	0.8	0.01
0.6	0.8	0.6	1	0.00	0.6	0.8	0.8	0.2	0.03	0.6	0.8	0.8	0.4	0.02

(continued)

Cost	Max	Min	Days	Ratio	Cost	Max	Min	Days	Ratio	Cost	Max	Min	Days	Ratio
0.6	0.8	0.8	0.6	0.01	0.6	0.8	0.8	0.8	0.01	0.6	0.8	0.8	1	0.00
0.6	0.8	1	0.2	0.04	0.6	0.8	1	0.4	0.01	0.6	0.8	1	0.6	0.01
0.6	0.8	1	0.8	0.01	0.6	0.8	1	1	0.00	0.6	1	0.2	0.2	0.07
0.6	1	0.2	0.4	0.02	0.6	1	0.2	0.6	0.01	0.6	1	0.2	0.8	0.01
0.6	1	0.2	1	0.00	0.6	1	0.4	0.2	0.04	0.6	1	0.4	0.4	0.02
0.6	1	0.4	0.6	0.01	0.6	1	0.4	0.8	0.01	0.6	1	0.4	1	0.01
0.6	1	0.6	0.2	0.03	0.6	1	0.6	0.4	0.02	0.6	1	0.6	0.6	0.01
0.6	1	0.6	0.8	0.01	0.6	1	0.6	1	0.00	0.6	1	0.8	0.2	0.03
0.6	1	0.8	0.4	0.02	0.6	1	0.8	0.6	0.01	0.6	1	0.8	0.8	0.01
0.6	1	0.8	1	0.01	0.6	1	1	0.2	0.02	0.6	1	1	0.4	0.02
0.6	1	1	0.6	0.01	0.6	1	1	0.8	0.01	0.6	1	1	1	0.01
0.8	0.2	0.2	0.2	0.32	0.8	0.2	0.2	0.4	0.06	0.8	0.2	0.2	0.6	0.03
0.8	0.2	0.2	0.8	0.01	0.8	0.2	0.2	1	0.01	0.8	0.2	0.4	0.2	0.29
0.8	0.2	0.4	0.4	0.06	0.8	0.2	0.4	0.6	0.02	0.8	0.2	0.4	0.8	0.01
0.8	0.2	0.4	1	0.01	0.8	0.2	0.6	0.2	0.29	0.8	0.2	0.6	0.4	0.07
0.8	0.2	0.6	0.6	0.03	0.8	0.2	0.6	0.8	0.01	0.8	0.2	0.6	1	0.01
0.8	0.2	0.8	0.2	0.26	0.8	0.2	0.8	0.4	0.08	0.8	0.2	0.8	0.6	0.02
0.8	0.2	0.8	0.8	0.02	0.8	0.2	0.8	1	0.01	0.8	0.2	1	0.2	0.17
0.8	0.2	1	0.4	0.06	0.8	0.2	1	0.6	0.03	0.8	0.2	1	0.8	0.02
0.8	0.2	1	1	0.01	0.8	0.4	0.2	0.2	0.22	0.8	0.4	0.2	0.4	0.05
0.8	0.4	0.2	0.6	0.02	0.8	0.4	0.2	0.8	0.01	0.8	0.4	0.2	1	0.00
0.8	0.4	0.4	0.2	0.16	0.8	0.4	0.4	0.4	0.04	0.8	0.4	0.4	0.6	0.02
0.8	0.4	0.4	0.8	0.01	0.8	0.4	0.4	1	0.00	0.8	0.4	0.6	0.2	0.14
0.8	0.4	0.6	0.4	0.04	0.8	0.4	0.6	0.6	0.02	0.8	0.4	0.6	0.8	0.01
0.8	0.4	0.6	1	0.00	0.8	0.4	0.8	0.2	0.12	0.8	0.4	0.8	0.4	0.04
0.8	0.4	0.8	0.6	0.02	0.8	0.4	0.8	0.8	0.01	0.8	0.4	0.8	1	0.01
0.8	0.4	1	0.2	0.11	0.8	0.4	1	0.4	0.04	0.8	0.4	1	0.6	0.03
0.8	0.4	1	0.8	0.01	0.8	0.4	1	1	0.01	0.8	0.6	0.2	0.2	0.17
0.8	0.6	0.2	0.4	0.04	0.8	0.6	0.2	0.6	0.02	0.8	0.6	0.2	0.8	0.01
0.8	0.6	0.2	1	0.00	0.8	0.6	0.4	0.2	0.13	0.8	0.6	0.4	0.4	0.03
0.8	0.6	0.4	0.6	0.02	0.8	0.6	0.4	0.8	0.01	0.8	0.6	0.4	1	0.00
0.8	0.6	0.6	0.2	0.15	0.8	0.6	0.6	0.4	0.04	0.8	0.6	0.6	0.6	0.02
0.8	0.6	0.6	0.8	0.01	0.8	0.6	0.6	1	0.00	0.8	0.6	0.8	0.2	0.12
0.8	0.6	0.8	0.4	0.04	0.8	0.6	0.8	0.6	0.02	0.8	0.6	0.8	0.8	0.01
0.8	0.6	0.8	1	0.00	0.8	0.6	1	0.2	0.10	0.8	0.6	1	0.4	0.03
0.8	0.6	1	0.6	0.02	0.8	0.6	1	0.8	0.01	0.8	0.6	1	1	0.01
0.8	0.8	0.2	0.2	0.12	0.8	0.8	0.2	0.4	0.03	0.8	0.8	0.2	0.6	0.01
0.8	0.8	0.2	0.8	0.00	0.8	0.8	0.2	1	0.00	0.8	0.8	0.4	0.2	0.10
0.8	0.8	0.4	0.4	0.02	0.8	0.8	0.4	0.6	0.01	0.8	0.8	0.4	0.8	0.01
0.8	0.8	0.4	1	0.01	0.8	0.8	0.6	0.2	0.10	0.8	0.8	0.6	0.4	0.03
0.8	0.8	0.6	0.6	0.01	0.8	0.8	0.6	0.8	0.01	0.8	0.8	0.6	1	0.01
0.8	0.8	0.8	0.2	0.08	0.8	0.8	0.8	0.4	0.02	0.8	0.8	0.8	0.6	0.01
0.8	0.8	0.8	0.8	0.01	0.8	0.8	0.8	1	0.01	0.8	0.8	1	0.2	0.07
0.8	0.8	1	0.4	0.03	0.8	0.8	1	0.6	0.02	0.8	0.8	1	0.8	0.01
0.8	0.8	1	1	0.01	0.8	1	0.2	0.2	0.07	0.8	1	0.2	0.4	0.02
0.8	1	0.2	0.6	0.01	0.8	1	0.2	0.8	0.00	0.8	1	0.2	1	0.00
0.8	1	0.4	0.2	0.05	0.8	1	0.4	0.4	0.02	0.8	1	0.4	0.6	0.01
0.8	1	0.4	0.8	0.01	0.8	1	0.4	1	0.01	0.8	1	0.6	0.2	0.04
0.8	1	0.6	0.4	0.02	0.8	1	0.6	0.6	0.01	0.8	1	0.6	0.8	0.00
0.8	1	0.6	1	0.01	0.8	1	0.8	0.2	0.03	0.8	1	0.8	0.4	0.02
0.8	1	0.8	0.6	0.01	0.8	1	0.8	0.8	0.01	0.8	1	0.8	1	0.00
0.8	1	1	0.2	0.03	0.8	1	1	0.4	0.02	0.8	1	1	0.6	0.01
0.8	1	1	0.8	0.01	0.8	1	1	1	0.01	1	0.2	0.2	0.2	0.31
1	0.2	0.2	0.4	0.06	1	0.2	0.2	0.6	0.02	1	0.2	0.2	0.8	0.01
1	0.2	0.2	1	0.01	1	0.2	0.4	0.2	0.20	1	0.2	0.4	0.4	0.05
1	0.2	0.4	0.6	0.01	1	0.2	0.4	0.8	0.01	1	0.2	0.4	1	0.00
1	0.2	0.6	0.2	0.25	1	0.2	0.6	0.4	0.05	1	0.2	0.6	0.6	0.02
1	0.2	0.6	0.8	0.01	1	0.2	0.6	1	0.01	1	0.2	0.8	0.2	0.29
1	0.2	0.8	0.4	0.08	1	0.2	0.8	0.6	0.03	1	0.2	0.8	0.8	0.01
1	0.2	0.8	1	0.01	1	0.2	1	0.2	0.28	1	0.2	1	0.4	0.08

(continued)

Cost	Max	Min	Days	Ratio	Cost	Max	Min	Days	Ratio	Cost	Max	Min	Days	Ratio
1	0.2	1	0.6	0.03	1	0.2	1	0.8	0.02	1	0.2	1	1	0.01
1	0.4	0.2	0.2	0.13	1	0.4	0.2	0.4	0.04	1	0.4	0.2	0.6	0.02
1	0.4	0.2	0.8	0.00	1	0.4	0.2	1	0.00	1	0.4	0.4	0.2	0.10
1	0.4	0.4	0.4	0.04	1	0.4	0.4	0.6	0.02	1	0.4	0.4	0.8	0.01
1	0.4	0.4	1	0.00	1	0.4	0.6	0.2	0.13	1	0.4	0.6	0.4	0.04
1	0.4	0.6	0.6	0.01	1	0.4	0.6	0.8	0.01	1	0.4	0.6	1	0.00
1	0.4	0.8	0.2	0.15	1	0.4	0.8	0.4	0.05	1	0.4	0.8	0.6	0.02
1	0.4	0.8	0.8	0.01	1	0.4	0.8	1	0.01	1	0.4	1	0.2	0.19
1	0.4	1	0.4	0.05	1	0.4	1	0.6	0.02	1	0.4	1	0.8	0.01
1	0.4	1	1	0.01	1	0.6	0.2	0.2	0.13	1	0.6	0.2	0.4	0.04
1	0.6	0.2	0.6	0.03	1	0.6	0.2	0.8	0.00	1	0.6	0.2	1	0.01
1	0.6	0.4	0.2	0.10	1	0.6	0.4	0.4	0.03	1	0.6	0.4	0.6	0.01
1	0.6	0.4	0.8	0.00	1	0.6	0.4	1	0.00	1	0.6	0.6	0.2	0.14
1	0.6	0.6	0.4	0.04	1	0.6	0.6	0.6	0.01	1	0.6	0.6	0.8	0.01
1	0.6	0.6	1	0.01	1	0.6	0.8	0.2	0.17	1	0.6	0.8	0.4	0.04
1	0.6	0.8	0.6	0.01	1	0.6	0.8	0.8	0.01	1	0.6	0.8	1	0.01
1	0.6	1	0.2	0.20	1	0.6	1	0.4	0.05	1	0.6	1	0.6	0.02
1	0.6	1	0.8	0.01	1	0.6	1	1	0.01	1	0.8	0.2	0.2	0.09
1	0.8	0.2	0.4	0.02	1	0.8	0.2	0.6	0.01	1	0.8	0.2	0.8	0.01
1	0.8	0.2	1	0.02	1	0.8	0.4	0.2	0.08	1	0.8	0.4	0.4	0.02
1	0.8	0.4	0.6	0.01	1	0.8	0.4	0.8	0.00	1	0.8	0.4	1	0.00
1	0.8	0.6	0.2	0.10	1	0.8	0.6	0.4	0.03	1	0.8	0.6	0.6	0.01
1	0.8	0.6	0.8	0.01	1	0.8	0.6	1	0.01	1	0.8	0.8	0.2	0.12
1	0.8	0.8	0.4	0.03	1	0.8	0.8	0.6	0.01	1	0.8	0.8	0.8	0.01
1	0.8	0.8	1	0.01	1	0.8	1	0.2	0.15	1	0.8	1	0.4	0.04
1	0.8	1	0.6	0.02	1	0.8	1	0.8	0.01	1	0.8	1	1	0.00
1	1	0.2	0.2	0.06	1	1	0.2	0.4	0.01	1	1	0.2	0.6	0.01
1	1	0.2	0.8	0.00	1	1	0.2	1	0.00	1	1	0.4	0.2	0.05
1	1	0.4	0.4	0.02	1	1	0.4	0.6	0.01	1	1	0.4	0.8	0.01
1	1	0.4	1	0.01	1	1	0.6	0.2	0.04	1	1	0.6	0.4	0.02
1	1	0.6	0.6	0.01	1	1	0.6	0.8	0.00	1	1	0.6	1	0.01
1	1	0.8	0.2	0.03	1	1	0.8	0.4	0.01	1	1	0.8	0.6	0.01
1	1	0.8	0.8	0.01	1	1	0.8	1	0.00	1	1	1	0.2	0.03
1	1	1	0.4	0.02	1	1	1	0.6	0.01	1	1	1	0.8	0.01
1	1	1	1	0.00										

Note: Cost, Max and Min are shorthands for the names of the ordinal overhang variables associated with cost, maximum and minimum reference prices. Days is the shorthand for the name of the ordinal variable of days since purchase.

APPENDIX C

COST_MAX_MIN_IA_ONLY_5 MODEL OUTPUT

Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min
2.52	0.00	0.2	0.2	0.2	0.86	0.00	0.2	0.2	0.4	0.37	0.00	0.2	0.2	0.6
0.06	0.02	0.2	0.2	0.8	-0.15	0.00	0.2	0.2	1.0	2.09	0.00	0.2	0.4	0.2
0.32	0.00	0.2	0.4	0.4	-0.07	0.02	0.2	0.4	0.6	-0.19	0.00	0.2	0.4	0.8
-0.08	0.07	0.2	0.4	1.0	1.69	0.00	0.2	0.6	0.2	0.04	0.32	0.2	0.6	0.4
-0.13	0.00	0.2	0.6	0.6	-0.33	0.00	0.2	0.6	0.8	-0.10	0.10	0.2	0.6	1.0
1.11	0.00	0.2	0.8	0.2	-0.09	0.14	0.2	0.8	0.4	-0.03	0.70	0.2	0.8	0.6
-0.06	0.50	0.2	0.8	0.8	0.02	0.82	0.2	0.8	1.0	0.73	0.00	0.2	1.0	0.2
0.13	0.42	0.2	1.0	0.4	0.08	0.66	0.2	1.0	0.6	-0.28	0.30	0.2	1.0	0.8
0.53	0.02	0.2	1.0	1.0	2.07	0.00	0.4	0.2	0.2	0.51	0.00	0.4	0.2	0.4
0.18	0.00	0.4	0.2	0.6	0.08	0.07	0.4	0.2	0.8	0.08	0.09	0.4	0.2	1.0
1.92	0.00	0.4	0.4	0.2	0.44	0.00	0.4	0.4	0.4	0.07	0.07	0.4	0.4	0.6
-0.06	0.15	0.4	0.4	0.8	-0.01	0.90	0.4	0.4	1.0	1.54	0.00	0.4	0.6	0.2
0.05	0.10	0.4	0.6	0.4	-0.06	0.10	0.4	0.6	0.6	-0.23	0.00	0.4	0.6	0.8
-0.07	0.16	0.4	0.6	1.0	1.02	0.00	0.4	0.8	0.2	-0.12	0.00	0.4	0.8	0.4
-0.23	0.00	0.4	0.8	0.6	-0.20	0.00	0.4	0.8	0.8	-0.22	0.00	0.4	0.8	1.0
0.13	0.00	0.4	1.0	0.2	-0.39	0.00	0.4	1.0	0.4	-0.48	0.00	0.4	1.0	0.6
-0.20	0.00	0.4	1.0	0.8	0.00	0.99	0.4	1.0	1.0	1.90	0.00	0.6	0.2	0.2
1.21	0.00	0.6	0.2	0.4	1.03	0.00	0.6	0.2	0.6	0.99	0.00	0.6	0.2	0.8
0.93	0.00	0.6	0.2	1.0	1.83	0.00	0.6	0.4	0.2	0.84	0.00	0.6	0.4	0.4
0.52	0.00	0.6	0.4	0.6	0.43	0.00	0.6	0.4	0.8	0.38	0.00	0.6	0.4	1.0
1.43	0.00	0.6	0.6	0.2	0.69	0.00	0.6	0.6	0.4	0.42	0.00	0.6	0.6	0.6
0.38	0.00	0.6	0.6	0.8	0.44	0.00	0.6	0.6	1.0	1.14	0.00	0.6	0.8	0.2
0.45	0.00	0.6	0.8	0.4	0.33	0.00	0.6	0.8	0.6	0.23	0.00	0.6	0.8	0.8
0.28	0.00	0.6	0.8	1.0	-0.44	0.00	0.6	1.0	0.4	-0.41	0.00	0.6	1.0	0.6
-0.19	0.00	0.6	1.0	0.8	-0.12	0.00	0.6	1.0	1.0	1.89	0.00	0.8	0.2	0.2
1.84	0.00	0.8	0.2	0.4	1.86	0.00	0.8	0.2	0.6	1.73	0.00	0.8	0.2	0.8
1.58	0.00	0.8	0.2	1.0	1.54	0.00	0.8	0.4	0.2	1.26	0.00	0.8	0.4	0.4
1.23	0.00	0.8	0.4	0.6	1.13	0.00	0.8	0.4	0.8	1.16	0.00	0.8	0.4	1.0
1.29	0.00	0.8	0.6	0.2	1.09	0.00	0.8	0.6	0.4	1.25	0.00	0.8	0.6	0.6
1.04	0.00	0.8	0.6	0.8	0.98	0.00	0.8	0.6	1.0	0.88	0.00	0.8	0.8	0.2
0.78	0.00	0.8	0.8	0.4	0.84	0.00	0.8	0.8	0.6	0.70	0.00	0.8	0.8	0.8
0.74	0.00	0.8	0.8	1.0	0.13	0.00	0.8	1.0	0.2	-0.25	0.00	0.8	1.0	0.4
-0.35	0.00	0.8	1.0	0.6	-0.37	0.00	0.8	1.0	0.8	0.00	0.97	0.8	1.0	1.0
1.81	0.00	1.0	0.2	0.2	1.61	0.00	1.0	0.2	0.4	1.78	0.00	1.0	0.2	0.6
1.97	0.00	1.0	0.2	0.8	2.02	0.00	1.0	0.2	1.0	1.20	0.00	1.0	0.4	0.2
1.12	0.00	1.0	0.4	0.4	1.31	0.00	1.0	0.4	0.6	1.37	0.00	1.0	0.4	0.8
1.62	0.00	1.0	0.4	1.0	1.19	0.00	1.0	0.6	0.2	0.92	0.00	1.0	0.6	0.4
1.27	0.00	1.0	0.6	0.6	1.36	0.00	1.0	0.6	0.8	1.62	0.00	1.0	0.6	1.0
0.63	0.00	1.0	0.8	0.2	0.59	0.00	1.0	0.8	0.4	0.93	0.00	1.0	0.8	0.6
1.11	0.00	1.0	0.8	0.8	1.36	0.00	1.0	0.8	1.0	0.21	0.01	1.0	1.0	0.2
0.24	0.00	1.0	1.0	0.4	-0.14	0.00	1.0	1.0	0.6	-0.40	0.00	1.0	1.0	0.8
-0.32	0.00	1.0	1.0	1.0										

Note: Cost, Max and Min are shorthands for the names of the ordinal overhang variables associated with cost, maximum and minimum reference prices.

APPENDIX D

COST_MAX_MIN_IA_ONLY_10 MODEL OUTPUT

Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min
2.84	0.00	0.1	0.1	0.1	1.75	0.00	0.1	0.1	0.2	1.02	0.00	0.1	0.1	0.3
0.63	0.00	0.1	0.1	0.4	0.56	0.00	0.1	0.1	0.5	0.29	0.00	0.1	0.1	0.6
0.08	0.18	0.1	0.1	0.7	-0.16	0.00	0.1	0.1	0.8	-0.22	0.00	0.1	0.1	0.9
-0.65	0.00	0.1	0.1	1.0	2.52	0.00	0.1	0.2	0.1	0.93	0.00	0.1	0.2	0.2
0.61	0.00	0.1	0.2	0.3	0.19	0.00	0.1	0.2	0.4	-0.09	0.17	0.1	0.2	0.5
-0.18	0.01	0.1	0.2	0.6	-0.32	0.00	0.1	0.2	0.7	-0.29	0.00	0.1	0.2	0.8
-0.12	0.10	0.1	0.2	0.9	-0.52	0.00	0.1	0.2	1.0	2.44	0.00	0.1	0.3	0.1
0.96	0.00	0.1	0.3	0.2	0.29	0.00	0.1	0.3	0.3	0.01	0.93	0.1	0.3	0.4
-0.24	0.00	0.1	0.3	0.5	-0.37	0.00	0.1	0.3	0.6	-0.33	0.00	0.1	0.3	0.7
-0.52	0.00	0.1	0.3	0.8	-0.18	0.05	0.1	0.3	0.9	-0.40	0.00	0.1	0.3	1.0
2.26	0.00	0.1	0.4	0.1	0.57	0.00	0.1	0.4	0.2	0.20	0.02	0.1	0.4	0.3
-0.20	0.06	0.1	0.4	0.4	-0.05	0.61	0.1	0.4	0.5	-0.12	0.24	0.1	0.4	0.6
-0.63	0.00	0.1	0.4	0.7	-0.06	0.61	0.1	0.4	0.8	-0.03	0.80	0.1	0.4	0.9
-0.42	0.02	0.1	0.4	1.0	2.12	0.00	0.1	0.5	0.1	0.45	0.00	0.1	0.5	0.2
0.04	0.72	0.1	0.5	0.3	-0.15	0.30	0.1	0.5	0.4	-0.14	0.33	0.1	0.5	0.5
-0.07	0.66	0.1	0.5	0.6	-0.18	0.27	0.1	0.5	0.7	-0.77	0.00	0.1	0.5	0.8
-0.28	0.18	0.1	0.5	0.9	-0.80	0.01	0.1	0.5	1.0	1.90	0.00	0.1	0.6	0.1
0.21	0.15	0.1	0.6	0.2	-0.44	0.05	0.1	0.6	0.3	-0.75	0.01	0.1	0.6	0.4
-0.51	0.05	0.1	0.6	0.5	-1.11	0.00	0.1	0.6	0.6	-0.70	0.04	0.1	0.6	0.7
-0.35	0.23	0.1	0.6	0.8	-0.44	0.22	0.1	0.6	0.9	-0.37	0.46	0.1	0.6	1.0
1.25	0.00	0.1	0.7	0.1	0.20	0.47	0.1	0.7	0.2	-0.29	0.39	0.1	0.7	0.3
-1.52	0.13	0.1	0.7	0.4	-0.11	0.82	0.1	0.7	0.5	-0.92	0.19	0.1	0.7	0.6
0.29	0.47	0.1	0.7	0.7	-0.74	0.29	0.1	0.7	0.8	-0.82	0.41	0.1	0.7	0.9
-0.40	0.69	0.1	0.7	1.0	1.39	0.05	0.1	0.8	0.1	1.19	0.23	0.1	0.8	0.2
-7.88	0.93	0.1	0.8	0.3	-7.88	0.95	0.1	0.8	0.4	1.82	0.00	0.1	0.8	0.5
1.91	0.00	0.1	0.8	0.6	-7.87	0.93	0.1	0.8	0.7	1.12	0.11	0.1	0.8	0.8
1.29	0.20	0.1	0.8	0.9	2.31	0.02	0.1	0.8	1.0	0.85	0.14	0.1	0.9	0.1
-7.99	0.94	0.1	0.9	0.2	-7.84	0.97	0.1	0.9	0.3	-7.93	0.99	0.1	0.9	0.4
2.12	0.03	0.1	0.9	0.5	1.68	0.09	0.1	0.9	0.6	-7.90	0.94	0.1	0.9	0.7
1.73	0.08	0.1	0.9	0.8	-7.88	0.94	0.1	0.9	0.9	1.71	0.09	0.1	0.9	1.0
-8.00	0.93	0.1	1.0	0.1	-8.06	0.96	0.1	1.0	0.2	-8.05	0.98	0.1	1.0	0.3
-7.90	0.95	0.1	1.0	0.5	-7.91	0.96	0.1	1.0	0.6	-7.98	0.96	0.1	1.0	0.7
-7.93	0.98	0.1	1.0	0.8	-7.90	0.98	0.1	1.0	0.9	-7.90	0.98	0.1	1.0	1.0
2.73	0.00	0.2	0.1	0.1	1.83	0.00	0.2	0.1	0.2	1.17	0.00	0.2	0.1	0.3
0.87	0.00	0.2	0.1	0.4	0.51	0.00	0.2	0.1	0.5	0.23	0.02	0.2	0.1	0.6
0.12	0.26	0.2	0.1	0.7	0.16	0.10	0.2	0.1	0.8	-0.02	0.88	0.2	0.1	0.9
-0.22	0.06	0.2	0.1	1.0	2.57	0.00	0.2	0.2	0.1	0.94	0.00	0.2	0.2	0.2
0.56	0.00	0.2	0.2	0.3	0.14	0.07	0.2	0.2	0.4	-0.13	0.13	0.2	0.2	0.5
-0.16	0.08	0.2	0.2	0.6	-0.21	0.02	0.2	0.2	0.7	-0.58	0.00	0.2	0.2	0.8
-0.49	0.00	0.2	0.2	0.9	-0.69	0.00	0.2	0.2	1.0	2.41	0.00	0.2	0.3	0.1
0.72	0.00	0.2	0.3	0.2	0.18	0.00	0.2	0.3	0.3	-0.15	0.04	0.2	0.3	0.4
-0.29	0.00	0.2	0.3	0.5	-0.54	0.00	0.2	0.3	0.6	-0.60	0.00	0.2	0.3	0.7
-0.45	0.00	0.2	0.3	0.8	-0.49	0.00	0.2	0.3	0.9	-0.38	0.00	0.2	0.3	1.0
2.21	0.00	0.2	0.4	0.1	0.30	0.00	0.2	0.4	0.2	-0.10	0.12	0.2	0.4	0.3
-0.35	0.00	0.2	0.4	0.4	-0.52	0.00	0.2	0.4	0.5	-0.56	0.00	0.2	0.4	0.6
-0.58	0.00	0.2	0.4	0.7	-0.63	0.00	0.2	0.4	0.8	-0.66	0.00	0.2	0.4	0.9
-0.70	0.00	0.2	0.4	1.0	1.97	0.00	0.2	0.5	0.1	0.29	0.00	0.2	0.5	0.2
-0.11	0.14	0.2	0.5	0.3	-0.47	0.00	0.2	0.5	0.4	-0.43	0.00	0.2	0.5	0.5
-0.91	0.00	0.2	0.5	0.6	-0.84	0.00	0.2	0.5	0.7	-1.06	0.00	0.2	0.5	0.8
-0.65	0.00	0.2	0.5	0.9	-0.38	0.01	0.2	0.5	1.0	1.72	0.00	0.2	0.6	0.1
0.12	0.07	0.2	0.6	0.2	-0.60	0.00	0.2	0.6	0.3	-0.44	0.00	0.2	0.6	0.4
-0.62	0.00	0.2	0.6	0.5	-0.49	0.00	0.2	0.6	0.6	-0.50	0.00	0.2	0.6	0.7
-0.58	0.00	0.2	0.6	0.8	-0.56	0.00	0.2	0.6	0.9	-0.30	0.09	0.2	0.6	1.0
1.42	0.00	0.2	0.7	0.1	-0.11	0.17	0.2	0.7	0.2	-0.26	0.01	0.2	0.7	0.3

(continued)

Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min
-0.74	0.00	0.2	0.7	0.4	-0.84	0.00	0.2	0.7	0.5	-0.61	0.00	0.2	0.7	0.6
-0.45	0.00	0.2	0.7	0.7	-0.63	0.00	0.2	0.7	0.8	-0.53	0.00	0.2	0.7	0.9
-0.48	0.03	0.2	0.7	1.0	0.96	0.00	0.2	0.8	0.1	-0.20	0.05	0.2	0.8	0.2
-0.73	0.00	0.2	0.8	0.3	-0.65	0.00	0.2	0.8	0.4	-0.53	0.00	0.2	0.8	0.5
-0.59	0.00	0.2	0.8	0.6	-0.71	0.00	0.2	0.8	0.7	-0.77	0.00	0.2	0.8	0.8
-0.36	0.14	0.2	0.8	0.9	-0.67	0.02	0.2	0.8	1.0	0.43	0.00	0.2	0.9	0.1
-0.50	0.00	0.2	0.9	0.2	-0.50	0.01	0.2	0.9	0.3	-0.22	0.30	0.2	0.9	0.4
-0.93	0.00	0.2	0.9	0.5	-0.49	0.07	0.2	0.9	0.6	0.06	0.79	0.2	0.9	0.7
0.24	0.32	0.2	0.9	0.8	-0.68	0.07	0.2	0.9	0.9	-0.30	0.40	0.2	0.9	1.0
-0.79	0.00	0.2	1.0	0.1	-1.28	0.00	0.2	1.0	0.2	-0.99	0.00	0.2	1.0	0.3
-2.25	0.00	0.2	1.0	0.4	-1.48	0.00	0.2	1.0	0.5	-2.22	0.00	0.2	1.0	0.6
-2.27	0.00	0.2	1.0	0.7	-0.64	0.02	0.2	1.0	0.8	-0.62	0.08	0.2	1.0	0.9
-1.00	0.05	0.2	1.0	1.0	2.72	0.00	0.3	0.1	0.1	1.93	0.00	0.3	0.1	0.2
1.47	0.00	0.3	0.1	0.3	1.27	0.00	0.3	0.1	0.4	1.01	0.00	0.3	0.1	0.5
0.71	0.00	0.3	0.1	0.6	0.59	0.00	0.3	0.1	0.7	0.48	0.00	0.3	0.1	0.8
0.37	0.00	0.3	0.1	0.9	-0.01	0.93	0.3	0.1	1.0	2.33	0.00	0.3	0.2	0.1
1.08	0.00	0.3	0.2	0.2	0.64	0.00	0.3	0.2	0.3	0.27	0.01	0.3	0.2	0.4
-0.04	0.74	0.3	0.2	0.5	-0.17	0.18	0.3	0.2	0.6	-0.09	0.48	0.3	0.2	0.7
-0.24	0.09	0.3	0.2	0.8	-0.16	0.26	0.3	0.2	0.9	-0.49	0.01	0.3	0.2	1.0
2.33	0.00	0.3	0.3	0.1	0.78	0.00	0.3	0.3	0.2	0.44	0.00	0.3	0.3	0.3
0.02	0.81	0.3	0.3	0.4	-0.03	0.78	0.3	0.3	0.5	-0.37	0.00	0.3	0.3	0.6
-0.34	0.00	0.3	0.3	0.7	-0.22	0.06	0.3	0.3	0.8	-0.50	0.00	0.3	0.3	0.9
-0.54	0.00	0.3	0.3	1.0	2.04	0.00	0.3	0.4	0.1	0.51	0.00	0.3	0.4	0.2
0.00	0.97	0.3	0.4	0.3	-0.08	0.32	0.3	0.4	0.4	-0.40	0.00	0.3	0.4	0.5
-0.34	0.00	0.3	0.4	0.6	-0.41	0.00	0.3	0.4	0.7	-0.48	0.00	0.3	0.4	0.8
-0.56	0.00	0.3	0.4	0.9	-0.60	0.00	0.3	0.4	1.0	1.78	0.00	0.3	0.5	0.1
0.13	0.04	0.3	0.5	0.2	-0.32	0.00	0.3	0.5	0.3	-0.29	0.00	0.3	0.5	0.4
-0.33	0.00	0.3	0.5	0.5	-0.70	0.00	0.3	0.5	0.6	-0.48	0.00	0.3	0.5	0.7
-0.61	0.00	0.3	0.5	0.8	-0.78	0.00	0.3	0.5	0.9	-0.18	0.23	0.3	0.5	1.0
1.43	0.00	0.3	0.6	0.1	-0.11	0.09	0.3	0.6	0.2	-0.39	0.00	0.3	0.6	0.3
-0.66	0.00	0.3	0.6	0.4	-0.59	0.00	0.3	0.6	0.5	-0.61	0.00	0.3	0.6	0.6
-0.55	0.00	0.3	0.6	0.7	-0.65	0.00	0.3	0.6	0.8	-0.82	0.00	0.3	0.6	0.9
-0.48	0.01	0.3	0.6	1.0	1.23	0.00	0.3	0.7	0.1	-0.20	0.00	0.3	0.7	0.2
-0.56	0.00	0.3	0.7	0.3	-0.61	0.00	0.3	0.7	0.4	-0.68	0.00	0.3	0.7	0.5
-0.64	0.00	0.3	0.7	0.6	-0.72	0.00	0.3	0.7	0.7	-0.59	0.00	0.3	0.7	0.8
-0.73	0.00	0.3	0.7	0.9	-0.97	0.00	0.3	0.7	1.0	0.75	0.00	0.3	0.8	0.1
-0.29	0.00	0.3	0.8	0.2	-0.51	0.00	0.3	0.8	0.3	-0.42	0.00	0.3	0.8	0.4
-0.63	0.00	0.3	0.8	0.5	-0.86	0.00	0.3	0.8	0.6	-0.43	0.00	0.3	0.8	0.7
-0.55	0.00	0.3	0.8	0.8	-0.59	0.00	0.3	0.8	0.9	-0.76	0.00	0.3	0.8	1.0
0.62	0.00	0.3	0.9	0.1	-0.26	0.02	0.3	0.9	0.2	0.01	0.95	0.3	0.9	0.3
-0.13	0.38	0.3	0.9	0.4	0.03	0.81	0.3	0.9	0.5	-0.38	0.03	0.3	0.9	0.6
0.55	0.00	0.3	0.9	0.7	-0.07	0.72	0.3	0.9	0.8	-0.46	0.06	0.3	0.9	0.9
-0.23	0.41	0.3	0.9	1.0	-0.43	0.00	0.3	1.0	0.1	-1.04	0.00	0.3	1.0	0.2
-1.34	0.00	0.3	1.0	0.3	-1.72	0.00	0.3	1.0	0.4	-1.58	0.00	0.3	1.0	0.5
-1.52	0.00	0.3	1.0	0.6	-1.71	0.00	0.3	1.0	0.7	-1.11	0.00	0.3	1.0	0.8
-1.17	0.00	0.3	1.0	0.9	-1.24	0.00	0.3	1.0	1.0	2.71	0.00	0.4	0.1	0.1
2.09	0.00	0.4	0.1	0.2	1.90	0.00	0.4	0.1	0.3	1.77	0.00	0.4	0.1	0.4
1.43	0.00	0.4	0.1	0.5	1.31	0.00	0.4	0.1	0.6	1.14	0.00	0.4	0.1	0.7
0.84	0.00	0.4	0.1	0.8	0.91	0.00	0.4	0.1	0.9	0.49	0.00	0.4	0.1	1.0
2.16	0.00	0.4	0.2	0.1	1.16	0.00	0.4	0.2	0.2	0.74	0.00	0.4	0.2	0.3
0.46	0.00	0.4	0.2	0.4	0.54	0.00	0.4	0.2	0.5	0.21	0.08	0.4	0.2	0.6
0.13	0.35	0.4	0.2	0.7	-0.10	0.52	0.4	0.2	0.8	-0.09	0.58	0.4	0.2	0.9
-0.02	0.91	0.4	0.2	1.0	2.18	0.00	0.4	0.3	0.1	1.02	0.00	0.4	0.3	0.2
0.64	0.00	0.4	0.3	0.3	0.16	0.13	0.4	0.3	0.4	0.09	0.38	0.4	0.3	0.5
0.02	0.87	0.4	0.3	0.6	-0.27	0.04	0.4	0.3	0.7	-0.27	0.06	0.4	0.3	0.8
-0.21	0.17	0.4	0.3	0.9	-0.25	0.19	0.4	0.3	1.0	1.85	0.00	0.4	0.4	0.1
0.70	0.00	0.4	0.4	0.2	0.38	0.00	0.4	0.4	0.3	0.15	0.09	0.4	0.4	0.4
-0.21	0.05	0.4	0.4	0.5	0.01	0.91	0.4	0.4	0.6	0.07	0.54	0.4	0.4	0.7
-0.29	0.03	0.4	0.4	0.8	-0.02	0.86	0.4	0.4	0.9	-0.10	0.53	0.4	0.4	1.0
1.78	0.00	0.4	0.5	0.1	0.52	0.00	0.4	0.5	0.2	0.04	0.59	0.4	0.5	0.3

(continued)

Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min
-0.18	0.05	0.4	0.5	0.4	-0.32	0.00	0.4	0.5	0.5	-0.41	0.00	0.4	0.5	0.6
-0.35	0.00	0.4	0.5	0.7	-0.55	0.00	0.4	0.5	0.8	-0.53	0.00	0.4	0.5	0.9
-0.32	0.04	0.4	0.5	1.0	1.46	0.00	0.4	0.6	0.1	0.29	0.00	0.4	0.6	0.2
-0.15	0.05	0.4	0.6	0.3	-0.48	0.00	0.4	0.6	0.4	-0.27	0.00	0.4	0.6	0.5
-0.56	0.00	0.4	0.6	0.6	-0.44	0.00	0.4	0.6	0.7	-0.47	0.00	0.4	0.6	0.8
-0.53	0.00	0.4	0.6	0.9	0.03	0.83	0.4	0.6	1.0	1.15	0.00	0.4	0.7	0.1
0.15	0.01	0.4	0.7	0.2	-0.09	0.22	0.4	0.7	0.3	-0.38	0.00	0.4	0.7	0.4
-0.29	0.00	0.4	0.7	0.5	-0.45	0.00	0.4	0.7	0.6	-0.29	0.01	0.4	0.7	0.7
-0.50	0.00	0.4	0.7	0.8	-0.27	0.04	0.4	0.7	0.9	-0.38	0.04	0.4	0.7	1.0
0.87	0.00	0.4	0.8	0.1	-0.12	0.09	0.4	0.8	0.2	-0.63	0.00	0.4	0.8	0.3
-0.22	0.01	0.4	0.8	0.4	-0.34	0.00	0.4	0.8	0.5	-0.30	0.00	0.4	0.8	0.6
-0.40	0.00	0.4	0.8	0.7	-0.29	0.02	0.4	0.8	0.8	-0.25	0.07	0.4	0.8	0.9
-0.56	0.01	0.4	0.8	1.0	0.57	0.00	0.4	0.9	0.1	-0.18	0.03	0.4	0.9	0.2
0.03	0.72	0.4	0.9	0.3	-0.30	0.01	0.4	0.9	0.4	-0.36	0.00	0.4	0.9	0.5
-0.22	0.07	0.4	0.9	0.6	-0.33	0.02	0.4	0.9	0.7	0.03	0.80	0.4	0.9	0.8
-0.11	0.48	0.4	0.9	0.9	-0.13	0.55	0.4	0.9	1.0	-0.28	0.00	0.4	1.0	0.1
-0.81	0.00	0.4	1.0	0.2	-0.95	0.00	0.4	1.0	0.3	-1.11	0.00	0.4	1.0	0.4
-1.17	0.00	0.4	1.0	0.5	-1.06	0.00	0.4	1.0	0.6	-0.88	0.00	0.4	1.0	0.7
-0.64	0.00	0.4	1.0	0.8	-0.44	0.00	0.4	1.0	0.9	-0.73	0.00	0.4	1.0	1.0
2.72	0.00	0.5	0.1	0.1	2.27	0.00	0.5	0.1	0.2	2.20	0.00	0.5	0.1	0.3
2.15	0.00	0.5	0.1	0.4	1.92	0.00	0.5	0.1	0.5	1.88	0.00	0.5	0.1	0.6
1.50	0.00	0.5	0.1	0.7	1.18	0.00	0.5	0.1	0.8	1.62	0.00	0.5	0.1	0.9
1.15	0.00	0.5	0.1	1.0	2.05	0.00	0.5	0.2	0.1	1.22	0.00	0.5	0.2	0.2
1.13	0.00	0.5	0.2	0.3	0.56	0.00	0.5	0.2	0.4	0.59	0.00	0.5	0.2	0.5
0.42	0.00	0.5	0.2	0.6	0.68	0.00	0.5	0.2	0.7	0.42	0.00	0.5	0.2	0.8
0.52	0.00	0.5	0.2	0.9	0.11	0.58	0.5	0.2	1.0	1.99	0.00	0.5	0.3	0.1
1.11	0.00	0.5	0.3	0.2	0.95	0.00	0.5	0.3	0.3	0.49	0.00	0.5	0.3	0.4
0.57	0.00	0.5	0.3	0.5	0.43	0.00	0.5	0.3	0.6	0.50	0.00	0.5	0.3	0.7
-0.03	0.85	0.5	0.3	0.8	0.07	0.63	0.5	0.3	0.9	0.08	0.65	0.5	0.3	1.0
1.70	0.00	0.5	0.4	0.1	0.95	0.00	0.5	0.4	0.2	0.64	0.00	0.5	0.4	0.3
0.52	0.00	0.5	0.4	0.4	0.23	0.02	0.5	0.4	0.5	0.34	0.00	0.5	0.4	0.6
-0.13	0.30	0.5	0.4	0.7	0.00	0.97	0.5	0.4	0.8	0.07	0.65	0.5	0.4	0.9
-0.07	0.66	0.5	0.4	1.0	1.65	0.00	0.5	0.5	0.1	0.59	0.00	0.5	0.5	0.2
0.43	0.00	0.5	0.5	0.3	0.31	0.00	0.5	0.5	0.4	0.02	0.88	0.5	0.5	0.5
-0.22	0.06	0.5	0.5	0.6	-0.17	0.15	0.5	0.5	0.7	-0.17	0.21	0.5	0.5	0.8
-0.07	0.59	0.5	0.5	0.9	-0.14	0.42	0.5	0.5	1.0	1.38	0.00	0.5	0.6	0.1
0.67	0.00	0.5	0.6	0.2	0.36	0.00	0.5	0.6	0.3	-0.11	0.22	0.5	0.6	0.4
0.11	0.22	0.5	0.6	0.5	-0.03	0.72	0.5	0.6	0.6	-0.15	0.18	0.5	0.6	0.7
-0.17	0.16	0.5	0.6	0.8	-0.39	0.00	0.5	0.6	0.9	0.20	0.18	0.5	0.6	1.0
1.23	0.00	0.5	0.7	0.1	0.34	0.00	0.5	0.7	0.2	0.01	0.92	0.5	0.7	0.3
0.00	0.98	0.5	0.7	0.4	0.08	0.36	0.5	0.7	0.5	-0.10	0.25	0.5	0.7	0.6
-0.16	0.12	0.5	0.7	0.7	-0.27	0.01	0.5	0.7	0.8	0.00	0.98	0.5	0.7	0.9
0.14	0.30	0.5	0.7	1.0	0.93	0.00	0.5	0.8	0.1	0.15	0.02	0.5	0.8	0.2
-0.15	0.06	0.5	0.8	0.3	-0.15	0.08	0.5	0.8	0.4	0.00	0.99	0.5	0.8	0.5
-0.25	0.01	0.5	0.8	0.6	-0.01	0.94	0.5	0.8	0.7	-0.44	0.00	0.5	0.8	0.8
-0.34	0.02	0.5	0.8	0.9	-0.08	0.62	0.5	0.8	1.0	0.37	0.00	0.5	0.9	0.1
-0.25	0.00	0.5	0.9	0.2	-0.29	0.00	0.5	0.9	0.3	-0.36	0.00	0.5	0.9	0.4
-0.21	0.03	0.5	0.9	0.5	-0.26	0.02	0.5	0.9	0.6	-0.23	0.04	0.5	0.9	0.7
-0.18	0.13	0.5	0.9	0.8	-0.27	0.06	0.5	0.9	0.9	-0.09	0.65	0.5	0.9	1.0
-0.15	0.00	0.5	1.0	0.1	-0.66	0.00	0.5	1.0	0.2	-0.88	0.00	0.5	1.0	0.3
-0.98	0.00	0.5	1.0	0.4	-0.85	0.00	0.5	1.0	0.5	-0.79	0.00	0.5	1.0	0.6
-0.69	0.00	0.5	1.0	0.7	-0.65	0.00	0.5	1.0	0.8	-0.27	0.00	0.5	1.0	0.9
-0.79	0.00	0.5	1.0	1.0	2.69	0.00	0.6	0.1	0.1	2.46	0.00	0.6	0.1	0.2
2.44	0.00	0.6	0.1	0.3	2.45	0.00	0.6	0.1	0.4	2.23	0.00	0.6	0.1	0.5
2.15	0.00	0.6	0.1	0.6	2.05	0.00	0.6	0.1	0.7	1.86	0.00	0.6	0.1	0.8
1.86	0.00	0.6	0.1	0.9	1.54	0.00	0.6	0.1	1.0	1.98	0.00	0.6	0.2	0.1
1.24	0.00	0.6	0.2	0.2	1.22	0.00	0.6	0.2	0.3	1.11	0.00	0.6	0.2	0.4
0.85	0.00	0.6	0.2	0.5	0.76	0.00	0.6	0.2	0.6	0.79	0.00	0.6	0.2	0.7
0.56	0.00	0.6	0.2	0.8	0.66	0.00	0.6	0.2	0.9	0.45	0.00	0.6	0.2	1.0
1.76	0.00	0.6	0.3	0.1	1.18	0.00	0.6	0.3	0.2	1.17	0.00	0.6	0.3	0.3

(continued)

Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min
0.72	0.00	0.6	0.3	0.4	0.67	0.00	0.6	0.3	0.5	0.79	0.00	0.6	0.3	0.6
0.74	0.00	0.6	0.3	0.7	0.59	0.00	0.6	0.3	0.8	0.11	0.47	0.6	0.3	0.9
-0.05	0.79	0.6	0.3	1.0	1.58	0.00	0.6	0.4	0.1	0.98	0.00	0.6	0.4	0.2
1.05	0.00	0.6	0.4	0.3	0.77	0.00	0.6	0.4	0.4	0.66	0.00	0.6	0.4	0.5
0.51	0.00	0.6	0.4	0.6	0.50	0.00	0.6	0.4	0.7	0.14	0.28	0.6	0.4	0.8
0.48	0.00	0.6	0.4	0.9	0.27	0.11	0.6	0.4	1.0	1.49	0.00	0.6	0.5	0.1
0.76	0.00	0.6	0.5	0.2	0.79	0.00	0.6	0.5	0.3	0.63	0.00	0.6	0.5	0.4
0.54	0.00	0.6	0.5	0.5	0.33	0.00	0.6	0.5	0.6	0.15	0.19	0.6	0.5	0.7
0.18	0.15	0.6	0.5	0.8	0.11	0.42	0.6	0.5	0.9	-0.11	0.54	0.6	0.5	1.0
1.22	0.00	0.6	0.6	0.1	0.72	0.00	0.6	0.6	0.2	0.68	0.00	0.6	0.6	0.3
0.40	0.00	0.6	0.6	0.4	0.24	0.01	0.6	0.6	0.5	0.27	0.01	0.6	0.6	0.6
0.16	0.14	0.6	0.6	0.7	-0.13	0.31	0.6	0.6	0.8	0.12	0.34	0.6	0.6	0.9
0.18	0.26	0.6	0.6	1.0	1.09	0.00	0.6	0.7	0.1	0.58	0.00	0.6	0.7	0.2
0.33	0.00	0.6	0.7	0.3	0.31	0.00	0.6	0.7	0.4	0.45	0.00	0.6	0.7	0.5
0.12	0.20	0.6	0.7	0.6	-0.03	0.75	0.6	0.7	0.7	-0.09	0.45	0.6	0.7	0.8
0.03	0.79	0.6	0.7	0.9	0.25	0.07	0.6	0.7	1.0	0.93	0.00	0.6	0.8	0.1
0.29	0.00	0.6	0.8	0.2	0.15	0.03	0.6	0.8	0.3	-0.04	0.60	0.6	0.8	0.4
-0.08	0.39	0.6	0.8	0.5	-0.08	0.41	0.6	0.8	0.6	-0.12	0.25	0.6	0.8	0.7
-0.21	0.06	0.6	0.8	0.8	-0.54	0.00	0.6	0.8	0.9	-0.11	0.45	0.6	0.8	1.0
0.44	0.00	0.6	0.9	0.1	-0.06	0.39	0.6	0.9	0.2	-0.13	0.10	0.6	0.9	0.3
0.06	0.43	0.6	0.9	0.4	0.16	0.06	0.6	0.9	0.5	-0.04	0.63	0.6	0.9	0.6
-0.30	0.01	0.6	0.9	0.7	-0.01	0.91	0.6	0.9	0.8	-0.29	0.03	0.6	0.9	0.9
-0.39	0.04	0.6	0.9	1.0	-0.54	0.00	0.6	1.0	0.2	-0.66	0.00	0.6	1.0	0.3
-0.77	0.00	0.6	1.0	0.4	-0.81	0.00	0.6	1.0	0.5	-0.72	0.00	0.6	1.0	0.6
-0.69	0.00	0.6	1.0	0.7	-0.69	0.00	0.6	1.0	0.8	-0.35	0.00	0.6	1.0	0.9
-0.77	0.00	0.6	1.0	1.0	2.75	0.00	0.7	0.1	0.1	2.46	0.00	0.7	0.1	0.2
2.60	0.00	0.7	0.1	0.3	2.59	0.00	0.7	0.1	0.4	2.61	0.00	0.7	0.1	0.5
2.60	0.00	0.7	0.1	0.6	2.35	0.00	0.7	0.1	0.7	2.26	0.00	0.7	0.1	0.8
1.96	0.00	0.7	0.1	0.9	1.84	0.00	0.7	0.1	1.0	1.91	0.00	0.7	0.2	0.1
1.40	0.00	0.7	0.2	0.2	1.32	0.00	0.7	0.2	0.3	1.13	0.00	0.7	0.2	0.4
1.23	0.00	0.7	0.2	0.5	1.27	0.00	0.7	0.2	0.6	1.19	0.00	0.7	0.2	0.7
0.89	0.00	0.7	0.2	0.8	0.93	0.00	0.7	0.2	0.9	0.86	0.00	0.7	0.2	1.0
1.67	0.00	0.7	0.3	0.1	1.11	0.00	0.7	0.3	0.2	1.07	0.00	0.7	0.3	0.3
1.08	0.00	0.7	0.3	0.4	1.01	0.00	0.7	0.3	0.5	0.87	0.00	0.7	0.3	0.6
1.03	0.00	0.7	0.3	0.7	0.62	0.00	0.7	0.3	0.8	0.40	0.00	0.7	0.3	0.9
0.54	0.00	0.7	0.3	1.0	1.10	0.00	0.7	0.4	0.1	0.89	0.00	0.7	0.4	0.2
1.04	0.00	0.7	0.4	0.3	1.05	0.00	0.7	0.4	0.4	0.89	0.00	0.7	0.4	0.5
0.81	0.00	0.7	0.4	0.6	0.76	0.00	0.7	0.4	0.7	0.76	0.00	0.7	0.4	0.8
0.66	0.00	0.7	0.4	0.9	0.43	0.01	0.7	0.4	1.0	1.43	0.00	0.7	0.5	0.1
0.44	0.00	0.7	0.5	0.2	0.80	0.00	0.7	0.5	0.3	0.84	0.00	0.7	0.5	0.4
0.80	0.00	0.7	0.5	0.5	0.50	0.00	0.7	0.5	0.6	0.56	0.00	0.7	0.5	0.7
0.50	0.00	0.7	0.5	0.8	0.34	0.01	0.7	0.5	0.9	-0.01	0.97	0.7	0.5	1.0
1.16	0.00	0.7	0.6	0.1	0.75	0.00	0.7	0.6	0.2	0.61	0.00	0.7	0.6	0.3
0.78	0.00	0.7	0.6	0.4	0.70	0.00	0.7	0.6	0.5	0.62	0.00	0.7	0.6	0.6
0.69	0.00	0.7	0.6	0.7	0.20	0.08	0.7	0.6	0.8	0.22	0.10	0.7	0.6	0.9
0.37	0.02	0.7	0.6	1.0	0.95	0.00	0.7	0.7	0.1	0.50	0.00	0.7	0.7	0.2
0.48	0.00	0.7	0.7	0.3	0.75	0.00	0.7	0.7	0.4	0.69	0.00	0.7	0.7	0.5
0.45	0.00	0.7	0.7	0.6	0.32	0.00	0.7	0.7	0.7	0.34	0.00	0.7	0.7	0.8
0.27	0.03	0.7	0.7	0.9	0.12	0.46	0.7	0.7	1.0	0.65	0.00	0.7	0.8	0.1
0.32	0.00	0.7	0.8	0.2	0.25	0.00	0.7	0.8	0.3	0.15	0.06	0.7	0.8	0.4
0.24	0.00	0.7	0.8	0.5	0.24	0.01	0.7	0.8	0.6	0.00	0.98	0.7	0.8	0.7
0.03	0.79	0.7	0.8	0.8	0.05	0.68	0.7	0.8	0.9	0.21	0.12	0.7	0.8	1.0
0.63	0.00	0.7	0.9	0.1	-0.02	0.80	0.7	0.9	0.2	0.06	0.41	0.7	0.9	0.3
0.12	0.15	0.7	0.9	0.4	0.01	0.92	0.7	0.9	0.5	0.14	0.12	0.7	0.9	0.6
-0.02	0.80	0.7	0.9	0.7	0.00	0.97	0.7	0.9	0.8	0.12	0.25	0.7	0.9	0.9
-0.19	0.19	0.7	0.9	1.0	-0.01	0.88	0.7	1.0	0.1	-0.35	0.00	0.7	1.0	0.2
-0.58	0.00	0.7	1.0	0.3	-0.72	0.00	0.7	1.0	0.4	-0.61	0.00	0.7	1.0	0.5
-0.65	0.00	0.7	1.0	0.6	-0.69	0.00	0.7	1.0	0.7	-0.61	0.00	0.7	1.0	0.8
-0.35	0.00	0.7	1.0	0.9	-0.27	0.00	0.7	1.0	1.0	2.80	0.00	0.8	0.1	0.1
2.35	0.00	0.8	0.1	0.2	2.54	0.00	0.8	0.1	0.3	2.58	0.00	0.8	0.1	0.4

(continued)

Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min
2.71	0.00	0.8	0.1	0.5	2.72	0.00	0.8	0.1	0.6	2.54	0.00	0.8	0.1	0.7
2.57	0.00	0.8	0.1	0.8	2.27	0.00	0.8	0.1	0.9	2.02	0.00	0.8	0.1	1.0
1.97	0.00	0.8	0.2	0.1	1.43	0.00	0.8	0.2	0.2	0.96	0.00	0.8	0.2	0.3
1.07	0.00	0.8	0.2	0.4	1.36	0.00	0.8	0.2	0.5	1.34	0.00	0.8	0.2	0.6
1.34	0.00	0.8	0.2	0.7	1.13	0.00	0.8	0.2	0.8	0.48	0.00	0.8	0.2	0.9
0.94	0.00	0.8	0.2	1.0	1.58	0.00	0.8	0.3	0.1	0.98	0.00	0.8	0.3	0.2
0.90	0.00	0.8	0.3	0.3	1.14	0.00	0.8	0.3	0.4	1.11	0.00	0.8	0.3	0.5
1.20	0.00	0.8	0.3	0.6	1.28	0.00	0.8	0.3	0.7	0.86	0.00	0.8	0.3	0.8
0.76	0.00	0.8	0.3	0.9	0.85	0.00	0.8	0.3	1.0	1.17	0.00	0.8	0.4	0.1
1.04	0.00	0.8	0.4	0.2	1.02	0.00	0.8	0.4	0.3	0.94	0.00	0.8	0.4	0.4
1.09	0.00	0.8	0.4	0.5	1.00	0.00	0.8	0.4	0.6	0.88	0.00	0.8	0.4	0.7
1.03	0.00	0.8	0.4	0.8	1.00	0.00	0.8	0.4	0.9	1.00	0.00	0.8	0.4	1.0
1.16	0.00	0.8	0.5	0.1	0.50	0.00	0.8	0.5	0.2	0.58	0.00	0.8	0.5	0.3
0.78	0.00	0.8	0.5	0.4	1.04	0.00	0.8	0.5	0.5	0.95	0.00	0.8	0.5	0.6
0.70	0.00	0.8	0.5	0.7	0.96	0.00	0.8	0.5	0.8	0.63	0.00	0.8	0.5	0.9
0.61	0.00	0.8	0.5	1.0	1.20	0.00	0.8	0.6	0.1	0.60	0.00	0.8	0.6	0.2
0.51	0.00	0.8	0.6	0.3	0.51	0.00	0.8	0.6	0.4	0.68	0.00	0.8	0.6	0.5
0.80	0.00	0.8	0.6	0.6	0.79	0.00	0.8	0.6	0.7	0.59	0.00	0.8	0.6	0.8
0.79	0.00	0.8	0.6	0.9	0.24	0.12	0.8	0.6	1.0	0.98	0.00	0.8	0.7	0.1
0.24	0.04	0.8	0.7	0.2	0.54	0.00	0.8	0.7	0.3	0.71	0.00	0.8	0.7	0.4
0.80	0.00	0.8	0.7	0.5	0.70	0.00	0.8	0.7	0.6	0.50	0.00	0.8	0.7	0.7
0.63	0.00	0.8	0.7	0.8	0.66	0.00	0.8	0.7	0.9	0.51	0.00	0.8	0.7	1.0
0.98	0.00	0.8	0.8	0.1	-0.03	0.79	0.8	0.8	0.2	-0.03	0.71	0.8	0.8	0.3
0.45	0.00	0.8	0.8	0.4	0.39	0.00	0.8	0.8	0.5	0.31	0.00	0.8	0.8	0.6
0.31	0.00	0.8	0.8	0.7	0.38	0.00	0.8	0.8	0.8	0.36	0.00	0.8	0.8	0.9
-0.27	0.12	0.8	0.8	1.0	0.49	0.00	0.8	0.9	0.1	0.19	0.09	0.8	0.9	0.2
0.08	0.37	0.8	0.9	0.3	-0.09	0.33	0.8	0.9	0.4	0.05	0.57	0.8	0.9	0.5
0.05	0.59	0.8	0.9	0.6	0.06	0.51	0.8	0.9	0.7	0.19	0.06	0.8	0.9	0.8
0.14	0.20	0.8	0.9	0.9	0.17	0.21	0.8	0.9	1.0	-0.03	0.75	0.8	1.0	0.1
-0.41	0.00	0.8	1.0	0.2	-0.42	0.00	0.8	1.0	0.3	-0.54	0.00	0.8	1.0	0.4
-0.56	0.00	0.8	1.0	0.5	-0.63	0.00	0.8	1.0	0.6	-0.81	0.00	0.8	1.0	0.7
-0.53	0.00	0.8	1.0	0.8	-0.29	0.00	0.8	1.0	0.9	-0.31	0.00	0.8	1.0	1.0
2.64	0.00	0.9	0.1	0.1	2.27	0.00	0.9	0.1	0.2	2.25	0.00	0.9	0.1	0.3
2.46	0.00	0.9	0.1	0.4	2.50	0.00	0.9	0.1	0.5	2.62	0.00	0.9	0.1	0.6
2.64	0.00	0.9	0.1	0.7	2.68	0.00	0.9	0.1	0.8	2.55	0.00	0.9	0.1	0.9
2.33	0.00	0.9	0.1	1.0	1.80	0.00	0.9	0.2	0.1	1.47	0.00	0.9	0.2	0.2
0.84	0.00	0.9	0.2	0.3	0.89	0.00	0.9	0.2	0.4	1.20	0.00	0.9	0.2	0.5
1.13	0.00	0.9	0.2	0.6	1.30	0.00	0.9	0.2	0.7	1.29	0.00	0.9	0.2	0.8
1.18	0.00	0.9	0.2	0.9	1.22	0.00	0.9	0.2	1.0	1.84	0.00	0.9	0.3	0.1
0.93	0.00	0.9	0.3	0.2	0.92	0.00	0.9	0.3	0.3	0.95	0.00	0.9	0.3	0.4
0.82	0.00	0.9	0.3	0.5	1.12	0.00	0.9	0.3	0.6	1.14	0.00	0.9	0.3	0.7
1.24	0.00	0.9	0.3	0.8	1.01	0.00	0.9	0.3	0.9	1.23	0.00	0.9	0.3	1.0
1.23	0.00	0.9	0.4	0.1	0.90	0.00	0.9	0.4	0.2	0.64	0.00	0.9	0.4	0.3
1.05	0.00	0.9	0.4	0.4	0.91	0.00	0.9	0.4	0.5	1.03	0.00	0.9	0.4	0.6
1.11	0.00	0.9	0.4	0.7	1.25	0.00	0.9	0.4	0.8	1.38	0.00	0.9	0.4	0.9
1.23	0.00	0.9	0.4	1.0	1.57	0.00	0.9	0.5	0.1	0.72	0.00	0.9	0.5	0.2
0.69	0.00	0.9	0.5	0.3	0.64	0.00	0.9	0.5	0.4	1.03	0.00	0.9	0.5	0.5
0.72	0.00	0.9	0.5	0.6	0.80	0.00	0.9	0.5	0.7	1.10	0.00	0.9	0.5	0.8
0.97	0.00	0.9	0.5	0.9	0.82	0.00	0.9	0.5	1.0	0.56	0.01	0.9	0.6	0.1
0.99	0.00	0.9	0.6	0.2	0.43	0.00	0.9	0.6	0.3	0.09	0.43	0.9	0.6	0.4
0.45	0.00	0.9	0.6	0.5	0.79	0.00	0.9	0.6	0.6	0.83	0.00	0.9	0.6	0.7
0.92	0.00	0.9	0.6	0.8	1.20	0.00	0.9	0.6	0.9	0.63	0.00	0.9	0.6	1.0
1.14	0.00	0.9	0.7	0.1	0.13	0.46	0.9	0.7	0.2	0.27	0.06	0.9	0.7	0.3
0.29	0.01	0.9	0.7	0.4	0.64	0.00	0.9	0.7	0.5	0.72	0.00	0.9	0.7	0.6
0.74	0.00	0.9	0.7	0.7	0.68	0.00	0.9	0.7	0.8	1.04	0.00	0.9	0.7	0.9
0.70	0.00	0.9	0.7	1.0	1.28	0.00	0.9	0.8	0.1	0.47	0.00	0.9	0.8	0.2
0.19	0.18	0.9	0.8	0.3	0.07	0.59	0.9	0.8	0.4	0.17	0.08	0.9	0.8	0.5
0.22	0.01	0.9	0.8	0.6	0.33	0.00	0.9	0.8	0.7	0.71	0.00	0.9	0.8	0.8
0.54	0.00	0.9	0.8	0.9	0.33	0.01	0.9	0.8	1.0	0.54	0.00	0.9	0.9	0.1
0.24	0.14	0.9	0.9	0.2	0.10	0.45	0.9	0.9	0.3	0.14	0.23	0.9	0.9	0.4

(continued)

Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min
0.32	0.00	0.9	0.9	0.5	0.29	0.00	0.9	0.9	0.6	0.29	0.00	0.9	0.9	0.7
0.19	0.07	0.9	0.9	0.8	0.25	0.01	0.9	0.9	0.9	0.25	0.05	0.9	0.9	1.0
-0.07	0.61	0.9	1.0	0.1	-0.48	0.00	0.9	1.0	0.2	-0.26	0.01	0.9	1.0	0.3
-0.18	0.02	0.9	1.0	0.4	-0.40	0.00	0.9	1.0	0.5	-0.61	0.00	0.9	1.0	0.6
-0.74	0.00	0.9	1.0	0.7	-0.73	0.00	0.9	1.0	0.8	-0.60	0.00	0.9	1.0	0.9
-0.32	0.00	0.9	1.0	1.0	1.21	0.00	1.0	0.1	0.1	1.50	0.00	1.0	0.1	0.2
1.84	0.00	1.0	0.1	0.3	2.17	0.00	1.0	0.1	0.4	2.44	0.00	1.0	0.1	0.5
2.27	0.00	1.0	0.1	0.6	2.46	0.00	1.0	0.1	0.7	2.53	0.00	1.0	0.1	0.8
2.50	0.00	1.0	0.1	0.9	2.38	0.00	1.0	0.1	1.0	0.69	0.49	1.0	0.2	0.1
1.16	0.01	1.0	0.2	0.2	1.30	0.00	1.0	0.2	0.3	0.54	0.02	1.0	0.2	0.4
0.59	0.00	1.0	0.2	0.5	1.21	0.00	1.0	0.2	0.6	1.12	0.00	1.0	0.2	0.7
1.21	0.00	1.0	0.2	0.8	1.29	0.00	1.0	0.2	0.9	1.24	0.00	1.0	0.2	1.0
1.81	0.00	1.0	0.3	0.1	0.97	0.02	1.0	0.3	0.2	0.94	0.01	1.0	0.3	0.3
0.50	0.06	1.0	0.3	0.4	0.31	0.12	1.0	0.3	0.5	0.93	0.00	1.0	0.3	0.6
0.96	0.00	1.0	0.3	0.7	0.92	0.00	1.0	0.3	0.8	1.16	0.00	1.0	0.3	0.9
1.30	0.00	1.0	0.3	1.0	-0.12	0.90	1.0	0.4	0.1	1.53	0.00	1.0	0.4	0.2
-0.19	0.64	1.0	0.4	0.3	0.19	0.49	1.0	0.4	0.4	0.21	0.36	1.0	0.4	0.5
0.59	0.00	1.0	0.4	0.6	0.80	0.00	1.0	0.4	0.7	0.93	0.00	1.0	0.4	0.8
1.29	0.00	1.0	0.4	0.9	1.33	0.00	1.0	0.4	1.0	2.31	0.00	1.0	0.5	0.1
1.40	0.01	1.0	0.5	0.2	0.72	0.03	1.0	0.5	0.3	0.59	0.04	1.0	0.5	0.4
0.68	0.00	1.0	0.5	0.5	0.56	0.00	1.0	0.5	0.6	0.54	0.00	1.0	0.5	0.7
0.87	0.00	1.0	0.5	0.8	1.12	0.00	1.0	0.5	0.9	1.26	0.00	1.0	0.5	1.0
1.29	0.07	1.0	0.6	0.1	0.62	0.21	1.0	0.6	0.2	0.42	0.30	1.0	0.6	0.3
-0.14	0.61	1.0	0.6	0.4	0.22	0.25	1.0	0.6	0.5	0.15	0.34	1.0	0.6	0.6
0.56	0.00	1.0	0.6	0.7	0.52	0.00	1.0	0.6	0.8	0.93	0.00	1.0	0.6	0.9
1.17	0.00	1.0	0.6	1.0	0.58	0.41	1.0	0.7	0.1	0.78	0.08	1.0	0.7	0.2
-0.26	0.56	1.0	0.7	0.3	-0.03	0.91	1.0	0.7	0.4	0.17	0.48	1.0	0.7	0.5
0.46	0.00	1.0	0.7	0.6	0.29	0.01	1.0	0.7	0.7	0.73	0.00	1.0	0.7	0.8
1.00	0.00	1.0	0.7	0.9	1.10	0.00	1.0	0.7	1.0	1.66	0.02	1.0	0.8	0.1
-0.03	0.97	1.0	0.8	0.2	0.90	0.07	1.0	0.8	0.3	-0.14	0.69	1.0	0.8	0.4
0.11	0.67	1.0	0.8	0.5	0.47	0.00	1.0	0.8	0.6	0.46	0.00	1.0	0.8	0.7
0.46	0.00	1.0	0.8	0.8	0.63	0.00	1.0	0.8	0.9	0.76	0.00	1.0	0.8	1.0
0.66	0.25	1.0	0.9	0.1	0.84	0.14	1.0	0.9	0.2	0.07	0.87	1.0	0.9	0.3
0.46	0.15	1.0	0.9	0.4	0.16	0.53	1.0	0.9	0.5	0.49	0.00	1.0	0.9	0.6
0.35	0.01	1.0	0.9	0.7	-0.03	0.80	1.0	0.9	0.8	0.23	0.02	1.0	0.9	0.9
-0.04	0.65	1.0	0.9	1.0	-1.95	0.05	1.0	1.0	0.1	-1.08	0.28	1.0	1.0	0.2
0.94	0.00	1.0	1.0	0.3	0.00	0.99	1.0	1.0	0.4	-0.70	0.00	1.0	1.0	0.5
-0.50	0.00	1.0	1.0	0.6	-0.70	0.00	1.0	1.0	0.7	-0.82	0.00	1.0	1.0	0.8
-0.84	0.00	1.0	1.0	0.9	-0.66	0.00	1.0	1.0	1.0					

Note: Cost, Max and Min are shorthands for the names of the ordinal overhang variables associated with cost, maximum and minimum reference prices.

APPENDIX E

COST_MAX_MIN_SIMPLE_IA_ONLY_5 MODEL OUTPUT

Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min	Coef.	P-val	Cost	Max	Min
1.77	0.00	0.2	0.2	0.2	0.32	0.00	0.2	0.2	0.4	0.31	0.00	0.2	0.2	0.6
0.47	0.00	0.2	0.2	0.8	0.68	0.00	0.2	0.2	1.0	1.61	0.00	0.2	0.4	0.2
0.20	0.00	0.2	0.4	0.4	0.17	0.00	0.2	0.4	0.6	0.53	0.00	0.2	0.4	0.8
0.73	0.00	0.2	0.4	1.0	1.55	0.00	0.4	0.2	0.2	0.20	0.00	0.4	0.2	0.4
0.19	0.01	0.4	0.2	0.6	0.47	0.00	0.4	0.2	0.8	0.66	0.00	0.4	0.2	1.0
1.53	0.00	0.4	0.4	0.2	0.26	0.00	0.4	0.4	0.4	0.18	0.00	0.4	0.4	0.6
0.43	0.00	0.4	0.4	0.8	0.71	0.00	0.4	0.4	1.0	1.46	0.00	0.4	0.6	0.2
0.06	0.00	0.4	0.6	0.4	0.11	0.00	0.4	0.6	0.6	0.60	0.00	0.4	0.6	0.8
1.03	0.00	0.4	0.6	1.0	1.32	0.00	0.4	0.8	0.2	-0.33	0.00	0.4	0.8	0.4
0.07	0.35	0.4	0.8	0.6	0.52	0.00	0.4	0.8	0.8	1.04	0.00	0.4	0.8	1.0
1.59	0.00	0.6	0.2	0.2	0.52	0.00	0.6	0.2	0.4	0.43	0.00	0.6	0.2	0.6
0.77	0.00	0.6	0.2	0.8	0.77	0.00	0.6	0.2	1.0	1.31	0.00	0.6	0.4	0.2
0.55	0.00	0.6	0.4	0.4	0.37	0.00	0.6	0.4	0.6	0.85	0.00	0.6	0.4	0.8
0.80	0.00	0.6	0.4	1.0	1.37	0.00	0.6	0.6	0.2	0.33	0.00	0.6	0.6	0.4
0.37	0.00	0.6	0.6	0.6	0.92	0.00	0.6	0.6	0.8	1.40	0.00	0.6	0.6	1.0
1.07	0.00	0.6	0.8	0.2	-0.11	0.00	0.6	0.8	0.4	0.29	0.00	0.6	0.8	0.6
0.86	0.00	0.6	0.8	0.8	1.73	0.00	0.6	0.8	1.0	-0.72	0.00	0.6	1.0	0.4
-0.20	0.00	0.6	1.0	0.6	0.66	0.00	0.6	1.0	0.8	1.79	0.00	0.6	1.0	1.0
1.48	0.00	0.8	0.2	0.2	1.24	0.00	0.8	0.2	0.4	0.94	0.00	0.8	0.2	0.6
0.92	0.00	0.8	0.2	0.8	0.78	0.00	0.8	0.2	1.0	1.61	0.00	0.8	0.4	0.2
1.23	0.00	0.8	0.4	0.4	1.08	0.00	0.8	0.4	0.6	1.22	0.00	0.8	0.4	0.8
1.21	0.00	0.8	0.4	1.0	1.63	0.00	0.8	0.6	0.2	1.18	0.00	0.8	0.6	0.4
1.11	0.00	0.8	0.6	0.6	1.39	0.00	0.8	0.6	0.8	1.68	0.00	0.8	0.6	1.0
1.41	0.00	0.8	0.8	0.2	1.26	0.00	0.8	0.8	0.4	1.39	0.00	0.8	0.8	0.6
1.67	0.00	0.8	0.8	0.8	2.08	0.00	0.8	0.8	1.0	0.13	0.00	0.8	1.0	0.2
0.02	0.04	0.8	1.0	0.4	0.34	0.00	0.8	1.0	0.6	1.01	0.00	0.8	1.0	0.8
1.95	0.00	0.8	1.0	1.0	-4.29	0.95	1.0	0.2	0.4	1.29	0.00	1.0	0.2	0.6
1.06	0.00	1.0	0.2	0.8	0.81	0.00	1.0	0.2	1.0	1.72	0.02	1.0	0.4	0.4
1.28	0.00	1.0	0.4	0.6	1.25	0.00	1.0	0.4	0.8	1.15	0.00	1.0	0.4	1.0
0.92	0.19	1.0	0.6	0.4	1.33	0.00	1.0	0.6	0.6	1.31	0.00	1.0	0.6	0.8
1.35	0.00	1.0	0.6	1.0	1.52	0.00	1.0	0.8	0.4	1.43	0.00	1.0	0.8	0.6
1.50	0.00	1.0	0.8	0.8	1.57	0.00	1.0	0.8	1.0	-0.05	0.81	1.0	1.0	0.4
0.16	0.00	1.0	1.0	0.6	0.39	0.00	1.0	1.0	0.8	0.83	0.00	1.0	1.0	1.0

Note: Cost, Max and Min are shorthands for the names of the ordinal overhang variables associated with cost, maximum and minimum reference prices.

REFERENCES

Aggarwal, R. (2014). Animal spirits in financial economics: A review of deviations from economic rationality. *International Review of Financial Analysis*, *32*, 179–187. doi:10.1016/j.irfa.2013.07.018

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, N.J: Wiley.

Agresti, A. (2018). *Statistical methods for the social sciences* (5th ed.). Boston: Pearson.

Akay, E. O. (2008). *Essays on investor behavior* (Doctoral dissertation, Texas Tech University).

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., … Krewinkel, A. (2022). Rmarkdown: Dynamic documents for R (Version 2.18). Retrieved December 1, 2022, from https://CRAN.R-project.org/package=rmarkdown

An, L. (2015). Asset pricing when traders sell extreme winners and losers. *Review of Financial Studies*, *29*(3), 823–861. doi:10.1093/rfs/hhv060

An, L., & Argyle, B. (2014). V-shaped disposition: Mutual fund trading behavior and price effects. Scholarly Paper. doi:10.2139/ssrn.2529141

Antony, A. (2019). Behavioral finance and portfolio management: Review of theory and literature. *Journal of Public Affairs*, *20*(2). doi:10.1002/pa.1996

Arch Linux. Retrieved December 17, 2021, from https://archlinux.org/

Aspara, J., & Hoffmann, A. O. I. (2015). Selling losers and keeping winners: How (savings) goal dynamics predict a reversal of the disposition effect. *Marketing Letters*, *26*(2), 201–211. doi:10.1007/s11002-013-9275-9

Aust, F., Xie, Y., Lovelace, R., & Aden-Buie, G. (2019). Citr: ’RStudio’ add-in to insert markdown citations (Version 0.3.2). Retrieved December 6, 2022, from https://mran.microsoft.com/snapshot/2020-10-27/web/packages/citr/

Bakar, A. A., Othman, Z. A., & Shuib, N. L. M. (2009). Building a new taxonomy for data discretization techniques. In *2009 2nd Conference on Data Mining and Optimization* (pp. 132–140). 2009 2nd Conference on Data Mining and Optimization. doi:10.1109/DMO.2009.5341896

Balkanska, D. V. (2018). Disposition effect and analyst forecast dispersion. *Review of Quantitative Finance and Accounting*, *50*(3), 837–859. doi:10.1007/s11156-017-0648-7

Bansal, A., & Jacob, J. (2022). Impact of price path on disposition bias. *Journal of Banking & Finance*, *143*. doi:10.1016/j.jbankfin.2022.106616

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, *55*(2), 773–806. doi:10.1111/0022-1082.00226

Barber, B. M., & Odean, T. (2013). The behavior of individual investors. In *Handbook of the Economics of Finance* (Vol. 2, pp. 1533–1570). doi:10.1016/B978-0-44-459406-8.00022-6

Barberis, N., & Thaler, R. H. (2003). A survey of behavioral finance. In *Handbook of the Economics of Finance* (Vol. 1, pp. 1053–1128). doi:10.1016/S1574-0102(03)01027-6

Barberis, N., & Xiong, W. (2009). What drives the disposition effect? An analysis of a long-standing preference-based explanation. *The Journal of Finance*, *64*(2), 751–784. doi:10.1111/j.1540-6261.2009.01448.x

Barberis, N., & Xiong, W. (2012). Realization utility. *Journal of Financial Economics*, *104*(2), 251–271. doi:10.1016/j.jfineco.2011.10.005

Baucells, M., Weber, M., & Welfens, F. (2011). Reference-point formation and updating. *Management Science*, *57*(3), 506–519. doi:10.1287/mnsc.1100.1286

Bay, S. D. (2001). Multivariate discretization for set mining. *Knowledge and Information Systems*, *3*(4), 491–512. doi:10.1007/PL00011680

Ben-David, I., & Hirshleifer, D. (2012). Are investors really reluctant to realize their losses? Trading responses to past returns and the disposition effect. *Review of Financial Studies*, *25*(8), 2485–2532. doi:10.1093/rfs/hhs077

Bharandev, S., & Rao, S. N. (2020). Does the association between abnormal trading volumes and historical prices explain disposition effect? *Asia-Pacific Financial Markets*, *28*, 141–151. doi:10.1007/s10690-020-09321-5

BibLaTeX. Retrieved December 1, 2022, from https://www.ctan.org/pkg/biblatex

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654. JSTOR: 1831029

Boles, T. L., & Messick, D. M. (1995). A reverse outcome bias: The influence of multiple reference points on the evaluation of outcomes and decisions. *Organizational Behavior and Human Decision Processes*, *61*(3), 262–275. doi:10.1006/obhd.1995.1021

Bouteska, A., & Regaieg, B. (2018). Investor characteristics and the effect of disposition bias on the Tunisian stock market. *Borsa Istanbul Review*, *18*(4), 282–299. doi:10.1016/j.bir.2018.05.004

Bradley, D., Pantzalis, C., & Yuan, X. (2016). The influence of political bias in state pension funds. *Journal of Financial Economics*, *119*(1), 69–91. doi:10.1016/j.jfineco.2015.08.017

Brettschneider, J., & Burgess, M. (2017). *Using a frailty model to measure the effect of covariates on the disposition effect*. CRISM Working Paper Series, 17. Retrieved from https://warwick.ac.uk/fac/sci/statistics/crism/research/17-05/17-05w.pdf

Brettschneider, J., Burro, G., & Henderson, V. (2021). Wide framing disposition effect: An empirical study. *Journal of Economic Behavior & Organization*, *185*, 330–347. doi:10.1016/j.jebo.2021.03.003

Brown, A., & Yang, F. (2017). Salience and the disposition effect: Evidence from the introduction of "cash-outs" in betting markets. *Southern Economic Journal*, *83*(4), 1052–1073. doi:10.1002/soej.12202

Bryan, J., & RStudio. (2021). Googlesheets4: Access Google Sheets using the Sheets API V4 (Version 1.0.0). Retrieved December 17, 2021, from https://CRAN.R-project.org/package=googlesheets4

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). doi:10.18637/jss.v080.i01

Bürkner, P.-C. (2022). Estimating monotonic effects with brms. Retrieved November 3, 2022, from https://cran.r-project.org/web/packages/brms/vignettes/brms_monotonic.html

Bürkner, P.-C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, *73*(3), 420–451. doi:10.1111/bmsp.12195

Bürkner, P.-C., Gabry, J., Weber, S., Johnson, A., Modrak, M., Badr, H. S., … Ben-Shachar, M. S. (2021, November 22). Brms: Bayesian regression models using 'Stan' (Version 2.16.3). Retrieved December 17, 2021, from https://CRAN.R-project.org/package=brms

Chang, T. Y., Solomon, D. H., & Westerfield, M. M. (2016). Looking for someone to blame: Delegation, cognitive dissonance, and the disposition effect. *The Journal of Finance*, *71*(1), 267–302. doi:10.1111/jofi.12311

Chen, G., Kim, K. A., Nofsinger, J. R., & Rui, O. M. (2007). Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. *Journal of Behavioral Decision Making*, *20*(4), 425–451. doi:10.1002/bdm.561

Cheng, T. Y., Lee, C. I., & Lin, C. H. (2013). An examination of the relationship between the disposition effect and gender, age, the traded security, and bull–bear market conditions. *Journal of Empirical Finance*, *21*, 195–213. doi:10.1016/j.jempfin.2013.01.003

Chhabra, R., & De, S. (2012). Are out-of-pocket costs overweighted relative to opportunity costs? A disposition effect - based investigation. Scholarly Paper. doi:10.2139/ssrn.2022986

Chiang, M.-H., & Huang, H.-Y. (2017). Do mutual fund flows drive the disposition behaviour of fund managers? *Investment Analysts Journal*, *46*(4), 311–323. doi:10.1080/10293523.2017.1373447

Ching, T., Collet, Y., Inc., F., Tino, R., Przemyslaw, S., Yuta, M., … Brüggemann, S. (2021). Qs: Quick serialization of R objects (Version 0.25.2). Retrieved December 17, 2021, from https://CRAN.R-project.org/package=qs

Chiu, S.-B., Shu, P.-G., & Jian, H.-Y. (2009). House money, counterfactual and the disposition effect. Scholarly Paper. doi:10.2139/ssrn.1464598

Chiyachantana, C. N., & Yang, L. (2013). *Reference point adaptation and the disposition effect*. Retrieved from https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4784&context=lkcsb_research

Choe, H., & Eom, Y. (2009). *How fast investors close positions? Liquidation decisions under prospect theory*. Retrieved from http://kmfa.or.kr/paper/annual/2009/2009_26.pdf

CmdStanR. Retrieved December 17, 2021, from https://github.com/stan-dev/cmdstanr

Comai, G. (2022). Zoteror - access the Zotero API in R. Retrieved December 6, 2022, from https://github.com/giocomai/zoteror

Constantinides, G. M. (1984). Optimal stock trading with personal taxes: Implications for prices and the abnormal January returns. *Journal of Financial Economics*, *13*(1), 65–89. doi:10.1016/0304-405X(84)90032-1

Coval, J. D., & Shumway, T. (2005). Do behavioral biases affect prices? *The Journal of Finance*, *60*(1), 1–34. doi:10.1111/j.1540-6261.2005.00723.x

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220. JSTOR: 2985181

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. London ; New York: Chapman and Hall.

Cox Proportional Hazards. (n.d.). Retrieved October 26, 2022, from https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/coxph.html

Crane, A. D., & Hartzell, J. C. (2010). Is there a disposition effect in corporate investment decisions? Evidence from real estate investment trusts. Scholarly Paper. doi:10.2139/ssrn.1031010

Crossref. (n.d.). Retrieved January 16, 2023, from https://www.crossref.org/

Da Costa, N., Goulart, M., Cupertino, C., Macedo, J., & Da Silva, S. (2013). The disposition effect and investor experience. *Journal of Banking & Finance*, *37*(5), 1669–1675. doi:10.1016/j.jbankfin.2012.12.007

Dai, M., Jiang, Y., Liu, H., & Xu, J. (2022). A rational theory for disposition effects. *Review of Economic Dynamics*. doi:10.1016/j.red.2021.11.003

da Silva, P. P., & Mendes, V. (2021). Exchange-traded certificates, education and the disposition effect. *Journal of Behavioral and Experimental Finance*, *29*. doi:10.1016/j.jbef.2020.100456

da Silva, P. P., Mendes, V., & Abreu, M. (2020, April). *The disposition effect among mutual fund participants: A re-examination* (No. 2020/0126). ISEG - Lisbon School of Economics and Management, REM, Universidade de Lisboa. Retrieved January 24, 2021, from https://ideas.repec.org/p/ise/remwps/wp01262020.html

da Silva, P. P., Mendes, V., & Abreu, M. (2022). The disposition effect among mutual fund participants: A re-examination. *The European Journal of Finance*, *28*(12), 1237–1256. doi:10.1080/1351847X.2021.1998176

da Silva Rosa, R., To, H. M., & Walter, T. S. (2005). Evidence contrary to the disposition effect amongst UK managed funds. Scholarly Paper. doi:10.2139/ssrn.676342

De, S., Gondhi, N. R., & Sarkar, S. (2011). Behavioral biases, investor performance, and wealth transfers between investor groups. Scholarly Paper. doi:10.2139/ssrn.2022992

De Winne, R. (2020). Measuring the disposition effect. Scholarly Paper. doi:10.2139/ssrn.3519938

Debondt, W., & Thaler, R. H. (1985). Does the stock-market overreact. *Journal of Finance*, *40*(3), 793–805. doi:10.1111/j.1540-6261.1985.tb05004.x de Groot, A. (2017). The disposition effect: A longitudinal study for private and institutional investors in the Belgian market in different market circumstances. Scholarly Paper. doi:10.2139/ssrn.2933418

Della Vedova, J., Grant, A. R., & Westerholm, P. J. (2016). Trading momentum: How shares and investors behave during momentum periods. Scholarly Paper. doi:10.2139/ssrn.2826568

Dhar, R., & Zhu, N. (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science*, *52*(5), 726–740. doi:10.1287/mnsc.1040.0473

Dorn, D., & Strobl, G. (2010). Rational disposition effects: Theory and evidence. Scholarly Paper. doi:10.2139/ssrn.1571534

Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In *Machine Learning Proceedings 1995* (pp. 194–202). doi:10.1016/B978-1-55860-377-6.50032-3

Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., … Schwen, B. (2021). Data.table: Extension of 'data.frame'. Retrieved December 17, 2021, from https://CRAN.R-project.org/package=data.table

Dutton, K. (2021). *Black-and-white thinking: The burden of a binary brain in a complex world*. New York: Farrar, Straus and Giroux.

Engelberg, J., Henriksson, M., & Williams, J. (2018). The portfolio-driven disposition effect. Scholarly Paper. doi:10.2139/ssrn.3126997

Fama, E. F. (1965). Random walks in stock market prices. *Financial Analysts Journal*, *21*(5), 55–59. JSTOR: 4469865

Fama, E. F. (1970). Efficient capital markets - review of theory and empirical work. *Journal of Finance*, *25*(2), 383–423. doi:10.2307/2325486

Feng, L., & Seasholes, M. S. (2005). Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Review of Finance*, *9*(3), 305–351. doi:10.1007/s10679-005-2262-0

Ferris, S. P., Haugen, R. A., & Makhija, A. K. (1988). Predicting contemporary volume with historic volume at differential price levels: Evidence supporting the disposition effect. *The Journal of Finance*, *43*(3), 677–697. doi:10.1111/j.1540-6261.1988.tb04599.x

Fox, J., Venables, B., & Salverda, A. D. a. A. P. (2021). English: Translate Integers into English (Version 1.2-6). Retrieved December 5, 2022, from https://CRAN.R-project.org/package=english

Frino, A., Lepone, G., & Wright, D. (2015). Investor characteristics and the disposition effect. *Pacific-Basin Finance Journal*, *31*, 1–12. doi:10.1016/j.pacfin.2014.10.009

Frydman, C., & Camerer, C. (2016). Neural evidence of regret and its implications for investor behavior. *Review of Financial Studies*, *29*(11), 3108–3139. doi:10.1093/rfs/hhw010

Frydman, C., & Rangel, A. (2014). Debiasing the disposition effect by reducing the saliency of information about a stock's purchase price. *Journal of Economic Behavior & Organization*, *107*, 541–552. doi:10.1016/j.jebo.2014.01.017

Frydman, C., & Wang, B. (2020). The impact of salience on investor behavior: Evidence from a natural experiment. *The Journal of Finance*, *75*(1), 229–276. doi:10.1111/jofi.12851

García, S., Luengo, J., & Herrera, F. (2015). Discretization. In S. García, J. Luengo, & F. Herrera (Eds.), *Data Preprocessing in Data Mining* (pp. 245–283). doi:10.1007/978-3-319-10247-4_9

Gärling, T., Blomman, M., & Carle, T. A. (2017). Affect account of disposition effect and consequences for stock prices. *Review of Behavioral Finance*, *9*(2), 187–202. doi:10.1108/RBF-07-2016-0049

Gemayel, R., & Preda, A. (2018). Does a scopic regime erode the disposition effect? Evidence from a social trading platform. *Journal of Economic Behavior & Organization*, *154*, 175–190. doi:10.1016/j.jebo.2018.08.014

Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics*, *116*(4), 1233–1260. doi:10.1162/003355301753265561

Git. Retrieved December 17, 2021, from https://git-scm.com/

GitHub. Retrieved December 17, 2021, from https://github.com/

GNU. Retrieved December 1, 2022, from https://www.gnu.org/

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *WIREs Cognitive Science*, *1*(1), 69–78. doi:10.1002/wcs.26

Goldstone, R. L., Steyvers, M., & Larimer, K. (1996). Categorical perception of novel dimensions. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Routledge.

Google Scholar. (n.d.). Retrieved January 16, 2023, from https://scholar.google.com/

Grinblatt, M., & Keloharju, M. (2001). What makes investors trade? *The Journal of Finance*, *56*(2), 589–616. doi:10.1111/0022-1082.00338

Grinblatt, M., Keloharju, M., & Linnainmaa, J. T. (2012). IQ, trading behavior, and performance. *Journal of Financial Economics*, *104*(2), 339–362. doi:10.1016/j.jfineco.2011.05.016

Grosshans, D., & Zeisberger, S. (2018). All's well that ends well? On the importance of how returns are achieved. *Journal of Banking & Finance*, *87*, 397–410. doi:10.1016/j.jbankfin.2017.09.021

Guo, J., Gabry, J., Goodrich, B., & Weber, S. (n.d.). Convergence and efficiency diagnostics for Markov Chains — Rhat. Retrieved November 3, 2022, from https://mc-stan.org/rstan/reference/Rhat.html

H2O.ai. Retrieved December 17, 2021, from https://www.h2o.ai/

Han, H. H. (2019). *Essays on the disposition effect and asset prices* (Doctoral dissertation, Florida State University). Retrieved January 24, 2021, from https://fsu.digital.flvc.org/islandora/object/fsu%3A722533/

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), (pp. 1–25). Categorical Perception: The Groundwork of Cognition (01/01/87). Cambridge University Press. Retrieved November 14, 2022, from https://eprints.soton.ac.uk/250386/

Harnad, S. (2003). Categorical perception. In *Encyclopedia of Cognitive Science*. Nature Publishing Group: Macmillan. Retrieved November 14, 2022, from https://eprints.soton.ac.uk/257719/

Haryanto, S., Subroto, A., & Ulpah, M. (2020). Disposition effect and herding behavior in the cryptocurrency market. *Journal of Industrial and Business Economics*, *47*(1), 115–132. doi:10.1007/s40812-019-00130-0

Haugen, R. A. (2001). *The inefficient stock market* (2nd ed.). Upper Saddle River, N.J: Pearson.

He, X. (2020). Subjective learning of trading talent: Theory and evidence from individual investors in the U.S. Scholarly Paper. doi:10.2139/ssrn.3732447

He, X. D., & Zhou, X. Y. (2016). Hope, fear, and aspirations. *Mathematical Finance*, *26*(1), 3–50. doi:10.1111/mafi.12044

Heimer, R. Z. (2016). Peer pressure: Social interaction and the disposition effect. *Review of Financial Studies*, *29*(11), 3177–3209. doi:10.1093/rfs/hhw063

Hens, T., & Vlcek, M. (2011). Does prospect theory explain the disposition effect? *Journal of Behavioral Finance*, *12*(3), 141–157. doi:10.1080/15427560.2011.601976

Ingersoll, J. E., & Jin, L. J. (2013). Realization utility with reference-dependent preferences. *Review of Financial Studies*, *26*(3), 723–767. doi:10.1093/rfs/hhs116

Ismay, C. (2022). Thesisdown. Retrieved December 1, 2022, from https://github.com/ismayc/thesisdown

Ivkovich, Z., Pearson, N. D., & Weisbenner, S. (2003). *The disposition effect and households' stock trades: The importance of relative performance and dividends*. University of Illinois at Urbana-Champaign. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5d1305eb34dd99f28798485469be7ca470a69c21

Ivkovich, Z., Pearson, N. D., & Weisbenner, S. J. (2007). Everything is relative: The disposition effect and households' stock trades. Scholarly Paper. doi:10.2139/ssrn.687643

Ivkovich, Z., Poterba, J., & Weisbenner, S. (2005). Tax-motivated trading by individual investors. *American Economic Review*, *95*(5), 1605–1630. doi:10.1257/000282805775014461

Ivkovich, Z., & Weisbenner, S. (2009). Individual investor mutual fund flows. *Journal of Financial Economics*, *92*(2), 223–237. doi:10.1016/j.jfineco.2008.05.003

Ivkovich, Z., & Weisbenner, S. J. (2005, March 1). Individual investors' mutual fund and common stock share selling decisions: Different strokes for the same folks. Scholarly Paper. doi:10.2139/ssrn.676947

Ivkovich, Z., & Weisbenner, S. J. (2006). 'Old' money matters: The sensitivity of mutual fund redemption decisions to past performance. Scholarly Paper. doi:10.2139/ssrn.903792

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65–91. doi:10.2307/2328882. JSTOR: 2328882

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, *51*(2), 296–312. doi:10.1016/0749-5978(92)90015-Y

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. doi:10.2307/1914185. JSTOR: 1914185

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*(1), 160–173. doi:10.1038/scientificamerican0182-160

Kapoor, S., & Prosad, J. M. (2017). Behavioural finance: A review. *Procedia Computer Science*, *122*, 50–54. doi:10.1016/j.procs.2017.11.340

Kassambara, A. (2022). Ggpubr: 'ggplot2' based publication ready plots (Version 0.5.0). Retrieved December 1, 2022, from https://CRAN.R-project.org/package=ggpubr

Kassambara, A., Kosinski, M., Biecek, P., & Fabian, S. (2021). Survminer: Drawing survival curves using 'ggplot2' (Version 0.4.9). Retrieved December 17, 2021, from https://CRAN.R-project.org/package=survminer

Kaustia, M. (2010). Prospect theory and the disposition effect. *Journal of Financial and Quantitative Analysis*, *45*(03), 791–812. doi:10.1017/S0022109010000244

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1-2), 81–93. doi:10.1093/biomet/30.1-2.81

Kendall, M. G. (1970). *Rank correlation methods* (4th ed.). London: Griffin.

Kim, M., Kim, T., & Kim, T. S. (2021). V-shaped disposition effect, stock prices, and post-earnings-announcement drift: Evidence from Korea. *Journal of Behavioral Finance*. doi:10.1080/15427560.2021.1975715

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text*. doi:10.1007/978-1-4419-6646-9

Klik, M. (2020). Fst: Lightning fast serialization of data frames. Retrieved December 17, 2021, from https://CRAN.R-project.org/package=fst

Kluyver, T., Angerer, P., Schulz, J., & Ram, K. (2021). IRkernel: Native R kernel for the 'Jupyter notebook'. Retrieved December 17, 2021, from https://CRAN.R-project.org/package=IRkernel

Koop, G. J., & Johnson, J. G. (2012). The use of multiple reference points in risky decision making. *Journal of Behavioral Decision Making*, *25*(1), 49–62. doi:10.1002/bdm.713

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165. doi:10.1093/qje/121.4.1133

Kotomin, V., & Varma, A. (2022). Do large losses loom larger than gains? Salience, holding periods, and the disposition effect. *Financial Review*, *57*(2), 397–427. doi:10.1111/fire.12288

Krause, A., Wei, K. C. J., & Yang, Z. (2009). Determinants of disposition and reverse disposition effects. Scholarly Paper. doi:10.2139/ssrn.1369114

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Boston: Academic Press.

Kuo, M.-H., & Chen, S. K. (2012). Prospect theory and disposition patterns: Evidence from Taiwan investors. *Studies in Economics and Finance*, *29*(1), 43–51. doi:10.1108/10867371211203846

LaTeX. Retrieved December 1, 2022, from https://www.latex-project.org/

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., … H2O.ai. (2021). H2o: R interface for the 'H2O' scalable machine learning platform. Retrieved December 16, 2021, from https://CRAN.R-project.org/package=h2o

Li, J., Massa, M., Zhang, H., & Zhang, J. (2019). Air pollution, behavioral bias, and the disposition effect in China. *Journal of Financial Economics*, *142*(2), 641–673. doi:10.1016/j.jfineco.2019.09.003

Lin, C.-H., Huang, W.-H., & Zeelenberg, M. (2006). Multiple reference points in investor regret. *Journal of Economic Psychology*, *27*(6), 781–792. doi:10.1016/j.joep.2006.03.004

Linnainmaa, J. T. (2011). Why do (some) households trade so much? *Review of Financial Studies*, *24*(5), 1630–1666. doi:10.1093/rfs/hhr009

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, *47*(1), 13–37. doi:10.2307/1924119

Linux. Retrieved December 17, 2021, from https://www.linux.org/

Lo, A. W. (2011). Fear, greed, and financial crises: A cognitive neurosciences perspective. Scholarly Paper. doi:10.2139/ssrn.1943325

Lo, A. W., Repin, D. V., & Steenbarger, B. N. (2005). Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review*, *95*(2), 352–359. doi:10.1257/000282805774670095

Loos, B., Meyer, S., Weber, J., & Hackethal, A. (2014). Individual investors' trading motives and security selling behavior. Scholarly Paper. doi:10.2139/ssrn.2309382

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 20, pp. 255–295). doi:10.1016/S0065-2601(08)60416-5

Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology*, *43*(2), 286–313. doi:10.1006/jmps.1999.1259

Lud, M.-C., & Widmer, G. (2000). Relative unsupervised discretization for association rule mining. In D. A. Zighed, J. Komorowski, & J. Żytkow (Eds.). G. Goos, J. Hartmanis, & J. van Leeuwen (typeredactors), *Principles of Data Mining and Knowledge Discovery* (Vol. 1910, pp. 148–158). doi:10.1007/3-540-45372-5_15

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. doi:10.1111/j.1540-6261.1952.tb01525.x

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*(2), 209–237. doi:10.1080/02699930802204677. pmid: 19809584

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Boca Raton: Taylor and Francis, CRC Press.

McLean, M. W. (2022). RefManageR: Straightforward 'BibTeX' and 'BibLaTeX' bibliography management (Version 1.4.0). Retrieved December 6, 2022, from https://CRAN.R-project.org/package=RefManageR

Meng, J. (2010a). *The disposition effect and expectations as reference point*. University of California, San Diego. Retrieved from http://www.wiwi.uni-bonn.de/kraehmer/Lehre/Beh_Econ/Papiere/Meng-Disposition-Effect.pdf

Meng, J. (2010b). *Three essays on behavioral economics* (Doctoral dissertation, University of California, San Diego). Retrieved January 24, 2021, from https://escholarship.org/uc/item/7zr471st

Meng, J., & Weng, X. (2018). Can prospect theory explain the disposition effect? A new perspective on reference points. *Management Science*, *64*(7), 3331–3351. doi:10.1287/mnsc.2016.2711

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review*, *48*(3), 261–297.

Modigliani, F., & Miller, M. H. (1963). Corporate income taxes and the cost of capital: A correction. *The American Economic Review*, *53*(3), 433–443. JSTOR: 1809167

Muhl, S., & Talpsepp, T. (2018). Faster learning in troubled times: How market conditions affect the disposition effect. *The Quarterly Review of Economics and Finance*, *68*, 226–236. doi:10.1016/j.qref.2017.08.002

Muhlenbach, F., & Rakotomalala, R. (2005). Discretization for continuous attributes. In *Encyclopedia of Data Warehousing and Mining* (pp. 397–402). doi:10.4018/978-1-59140-557-3.ch076

Müller, K., Wickham, H., James, D. A., Falcon, S., Hipp, D. R., Kennedy, D., … RStudio. (2022). RSQLite: SQLite interface for R (Version 2.2.19). Retrieved December 6, 2022, from https://CRAN.R-project.org/package=RSQLite

Nelson, D. A., & Marler, P. (1989). Categorical perception of a natural stimulus continuum: Birdsong. *Science*, *244*(4907), 976–978. JSTOR: 1704504

Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance*, *53*(5), 1775–1798. doi:10.1111/0022-1082.00072

Ooms, J., Lang, D. T., & Hilaiel, L. (2022). Jsonlite: A simple and robust JSON parser and generator for R (Version 1.8.4). Retrieved December 6, 2022, from https://CRAN.R-project.org/package=jsonlite

OpenCitations. (n.d.). Retrieved January 16, 2023, from https://opencitations.net/

Ozsoylev, H. N., Walden, J., Yavuz, M. D., & Bildik, R. (2014). Investor networks in the stock market. *The Review of Financial Studies*, *27*(5), 1323–1366. doi:10.1093/rfs/hht065

Pandoc. Retrieved December 1, 2022, from https://pandoc.org/

Peng, C. (2017). Investor behavior under the law of small numbers. Scholarly Paper. doi:10.2139/ssrn.3066369

Project Jupyter. Retrieved December 17, 2021, from https://www.jupyter.org

Quispe-Torreblanca, E., Gathergood, J., Loewenstein, G., & Stewart, N. (2020). Investor attention, reference points and the disposition effect. Scholarly Paper. doi:10.2139/ssrn.3422790

Quispe-Torreblanca, E., Hume, D., Gathergood, J., Loewenstein, G. F., & Stewart, N. (2021). At the top of the mind: Peak prices and the disposition effect. Scholarly Paper. doi:10.2139/ssrn.3879102

R Core Team. (2021). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Ramiah, V., Xu, X., & Moosa, I. A. (2015). Neoclassical finance, behavioral finance and noise traders: A review and assessment of the literature. *International Review of Financial Analysis*, *41*, 89–100. doi:10.1016/j.irfa.2015.05.021

Ramírez-Gallego, S., García, S., Benítez, J. M., & Herrera, F. (2016). Multivariate discretization based on evolutionary cut points selection for classification. *IEEE Transactions on Cybernetics*, *46*(3), 595–608. doi:10.1109/TCYB.2015.2410143

Ramírez-Gallego, S., García, S., Mouriño-Talín, H., Martínez-Rego, D., Bolón-Canedo, V., Alonso-Betanzos, A., … Herrera, F. (2016). Data discretization: Taxonomy and big data challenge. *WIREs Data Mining and Knowledge Discovery*, *6*(1), 5–21. doi:10.1002/widm.1173

Ranguelova, E. (2001). Disposition effect and firm size: New evidence on individual investor trading activity. Scholarly Paper. doi:10.2139/ssrn.293618

Ren, K. (2016). pipeR: Multi-paradigm pipeline implementation (Version 0.6.1.3). Retrieved December 17, 2021, from https://CRAN.R-project.org/package=pipeR

Ren, K. (2021). Rlist: A toolbox for non-tabular data manipulation (Version 0.4.6.2). Retrieved December 17, 2021, from https://CRAN.R-project.org/package=rlist

Richards, D. W. (2013). *The disposition effect, dual process theory and emotion regulation* (Doctoral dissertation, The Open University). Retrieved January 24, 2021, from http://oro.open.ac.uk/54524/

Richards, D. W., Fenton-O'Creevy, M., Rutterford, J., & Kodwani, D. G. (2018). Is the disposition effect related to investors' reliance on System 1 and System 2 processes or their strategy of emotion regulation? *Journal of Economic Psychology*, *66*, 79–92. doi:10.1016/j.joep.2018.01.003

Richards, D. W., Rutterford, J., Kodwani, D., & Fenton-O'Creevy, M. (2017). Stock market investors' use of stop losses and the disposition effect. *The European Journal of Finance*, *23*(2), 130–152. doi:10.1080/1351847X.2015.1048375

RStudio. Retrieved December 1, 2022, from https://www.rstudio.com/categories/rstudio-ide/

Sakaguchi, H., Stewart, N., & Walasek, L. (2019). The disposition effect varies with portfolio composition because people take gain-loss-domain-level sell decisions. Scholarly Paper. doi:10.2139/ssrn.3053331

Schlarbaum, G., Lewellen, W., & Lease, R. (1978). Realized returns on common stock investments - Experience of individual investors. *Journal of Business*, *51*(2), 299–325. doi:10.1086/295998

Seiler, M. J., Seiler, V. L., & Lane, M. A. (2012). Mental accounting and false reference points in real estate investment decision making. *Journal of Behavioral Finance*, *13*(1), 17–26. doi:10.1080/15427560.2012.653293

Seru, A., Shumway, T., & Stoffman, N. (2010). Learning by trading. *Review of Financial Studies*, *23*(2), 705–739. doi:10.1093/rfs/hhp060

Shapira, Z., & Venezia, I. (2001). Patterns of behavior of professionally managed and independent investors. *Journal of Banking & Finance*, *25*(8), 1573–1587. doi:10.1016/S0378-4266(00)00139-4

Sharpe, W. F. (1964). Capital-asset prices - A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442. doi:10.2307/2977928

Shefrin, H., & Statman, M. (1984). Explaining investor preference for cash dividends. *Journal of Financial Economics*, *13*(2), 253–282. doi:10.1016/0304-405X(84)90025-4

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, *40*(3), 777–790. doi:10.1111/j.1540-6261.1985.tb05002.x

Shiller, R. J. (1981). The use of volatility measures in assessing market efficiency. *The Journal of Finance*, *36*(2), 291–304. doi:10.1111/j.1540-6261.1981.tb00441.x

Shumway, T., & Wu, G. (2005). Does disposition drive momentum? Scholarly Paper. doi:10.2139/ssrn.771486

SQLite. Retrieved December 6, 2022, from https://www.sqlite.org/index.html

Stan. Retrieved December 17, 2021, from https://mc-stan.org/

Statman, M. (1999). Behaviorial finance: Past battles and future engagements. *Financial Analysts Journal*, *55*(6), 18–27. doi:10.2469/faj.v55.n6.2311

Statman, M. (2014). Behavioral finance: Finance with normal people. *Borsa Istanbul Review*, *14*(2), 65–73. doi:10.1016/j.bir.2014.03.001

Stoffman, N. (2008). *Individual and institutional investor behavior* (Doctoral dissertation, University of Michigan).

Strahilevitz, M. A., Odean, T., & Barber, B. M. (2011). Once burned, twice shy: How naive learning, counterfactuals, and regret affect the repurchase of stocks previously sold. *Journal of Marketing Research*, *48*(SPL), S102–S120. doi:10.1509/jmkr.48.SPL.S102

Studer, M., GmbH, M. S., Foundation, T. A. S., Builder, G., & Woolridge, B. (2021). XLConnect: Excel connector for R. Retrieved December 17, 2021, from https://CRAN.R-project.org/package=XLConnect

Sullivan, K., & Kida, T. (1995). The effect of multiple reference points and prior gains and losses on managers' risky decision making. *Organizational Behavior and Human Decision Processes*, *64*(1), 76–83. doi:10.1006/obhd.1995.1091

Summers, B., & Duxbury, D. (2007). Unraveling the disposition effect: The role of prospect theory and emotions. Scholarly Paper. doi:10.2139/ssrn.1026915

Summers, B., & Duxbury, D. (2012). Decision-dependent emotions and behavioral anomalies. *Organizational Behavior and Human Decision Processes*, *118*(2), 226–238. doi:10.1016/j.obhdp.2012.03.004

Svedsater, H., Karlsson, N., & Garling, T. (2009). Momentum trading, disposition effects and prediction of future share prices: An experimental study of multiple reference points in responses to short- and long-run return trends. *Applied Financial Economics*, *19*(8), 595–610. doi:10.1080/09603100801982620

Talpsepp, T. (2010). Does gender and age affect investor performance and the disposition effect? *Research in Economics & Business: Central & Eastern Europe*, *2*(1), 76–93. Retrieved from http://www.rebcee.eu/index.php/REB/article/view/25

Talpsepp, T. (2011). Reverse disposition effect of foreign investors. *Journal of Behavioral Finance*, *12*(4), 183–200. doi:10.1080/15427560.2011.606387

Talpsepp, T., & Vaarmets, T. (2019). The disposition effect, performance, stop loss orders and education. *Journal of Behavioral and Experimental Finance*, *24*. doi:10.1016/j.jbef.2019.100240

Tekçe, B., Yılmaz, N., & Bildik, R. (2016). What factors affect behavioral biases? Evidence from Turkish individual stock investors. *Research in International Business and Finance*, *37*, 515–526. doi:10.1016/j.ribaf.2015.11.017

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60. doi:10.1016/0167-2681(80)90051-7

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, *27*(1), 15–25. doi:10.1287/mksc.1070.0330

Thaler, R. H. (1999). The end of behavioral finance. *Financial Analysts Journal*, *55*(6), 12–17. doi:10.2469/faj.v55.n6.2310

Thaler, R. H., & Shefrin, H. (1981). An economic-theory of self-control. *Journal of Political Economy*, *89*(2), 392–406. doi:10.1086/260971

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model*. doi:10.1007/978-1-4757-3294-8

Therneau, T. M., Lumley, T., Elizabeth, A., & Cynthia, C. (2021). Survival: Survival analysis (Version 3.2-13). Retrieved December 17, 2021, from https://CRAN.R-project.org/package=survival

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. doi:10.1007/BF00122574

Vaarmets, T., Liivamägi, K., & Talpsepp, T. (2019). How does learning and education help to overcome the disposition effect? *Review of Finance*, *23*(4), 801–830. doi:10.1093/rof/rfy006

van den Brand, T. (2022). Ggh4x: Hacks for 'ggplot2' (Version 0.2.3). Retrieved November 22, 2022, from https://CRAN.R-project.org/package=ggh4x

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*(5), 207–213. doi:10.1016/S1364-6613(03)00095-0

Waiyasara, K., & Padungsaksawasdi, C. (2020). The V-shaped disposition effect in the stock exchange of Thailand. *The Journal of Asian Finance, Economics, and Business*, *7*(11), 55–65. doi:10.13106/JAFEB.2020.VOL7.NO11.055

Wang, L., Cao, S., Yuan, M., & Cheng, J. (2010). The reverse-disposition effect: Theory and evidence. In *2010 Third International Conference on Business Intelligence and Financial Engineering* (pp. 442–444). 2010 Third International Conference on Business Intelligence and Financial Engineering. doi:10.1109/BIFE.2010.107

Weber, M., & Camerer, C. F. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, *33*(2), 167–184. doi:10.1016/S0167-2681(97)00089-9

Weigand, R. A. (1993). *The cointegrating relationship between stock prices and trading volume: Evidence regarding the predictability of security returns.* (Doctoral dissertation, The University of Arizona). Retrieved January 24, 2021, from https://repository.arizona.edu/handle/10150/186294

Wheeler Huttunen, A. (2016). *A neuroeconomic investigation of risky decision-making and loss in the rat* (Doctoral dissertation, University of St Andrews). Retrieved January 24, 2021, from https://research-repository.st-andrews.ac.uk/handle/10023/9331

Wickham, H., & RStudio. (2021). Tidyverse: Easily install and load the 'Tidyverse'. Retrieved December 17, 2021, from https://CRAN.R-project.org/package=tidyverse

Wickham, H., & RStudio. (2022). Httr: Tools for working with URLs and HTTP (Version 1.4.4). Retrieved December 6, 2022, from https://CRAN.R-project.org/package=httr

Wickham, H., Seidel, D., & RStudio. (2022). Scales: Scale functions for visualization. Retrieved December 1, 2022, from https://CRAN.R-project.org/package=scales

Wierzbitzki, M., & Seidens, S. (2018). The causal influence of investment goals on the disposition effect. Scholarly Paper. doi:10.2139/ssrn.3275998

XeTeX. Retrieved December 1, 2022, from https://xetex.sourceforge.net/

Xie, Y., Allaire, J. J., Kim, A., Samuel-Rosa, A., Oles, A., Yasumoto, A., … Inc, F. (2022). Bookdown: Authoring books and technical documents with R Markdown. Retrieved December 1, 2022, from https://CRAN.R-project.org/package=bookdown

Xie, Y., Sarma, A., Vogt, A., Andrew, A., Zvoleff, A., Al-Zubaidi, A., … Kamvar, Z. N. (2022). Knitr: A General-purpose package for dynamic report generation in R. Retrieved December 1, 2022, from https://CRAN.R-project.org/package=knitr

Yang, Y., Webb, G. I., & Wu, X. (2010). Discretization methods. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 101–116). doi:10.1007/978-0-387-09823-4_6

Yannakakis, G. N., Cowie, R., & Busso, C. (2021). The ordinal nature of emotions: An emerging approach. *IEEE Transactions on Affective Computing*, *12*(1), 16–35. doi:10.1109/TAFFC.2018.2879512

Zahera, S. A., & Bansal, R. (2018). Do investors exhibit behavioral biases in investment decision making? A systematic review. *Qualitative Research in Financial Markets*, 00–00. doi:10.1108/QRFM-04-2017-0028

Zhu, H., Travison, T., Tsai, T., Beasley, W., Xie, Y., Yu, G., … Evans, B. (2021). kableExtra: Construct complex table with ’kable’ and pipe syntax. Retrieved December 1, 2022, from https://CRAN.R-project.org/package=kableExtra

Zimmermann, T. (2015). *Inductive learning and theory testing: Applications in finance* (Doctoral dissertation, Harvard University). Retrieved January 24, 2021, from https://dash.harvard.edu/handle/1/17467320

Zotero. Retrieved December 1, 2022, from https://www.zotero.org/