2025

Two essays on empirical asset pricing

BOSTON UNIVERSITY

QUESTROM SCHOOL OF BUSINESS

Dissertation

TWO ESSAYS ON EMPIRICAL ASSET PRICING

by

FANWEI WEI

B.S., Jilin University, 2018
M.S., Johns Hopkins University, 2020

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2025

Approved by

First Reader

Andrea Vedolin, PhD
Dean's Research Scholar
Professor of Finance

Second Reader

Fernando Zapatero, PhD
Department Chair of Finance
Richard D. Cohen Professor in Finance

Third Reader

Eugene Sorets, PhD
Adjunct Professor of Finance

Acknowledgments

I would like to express the biggest gratitude to my PhD advisor, Professor Andrea Vedolin, for introducing me to asset pricing field and guiding me through my research topics. Thank you for all the discussions and insights. I always walk away from our discussions with updated understandings about my research. I am also grateful to your patience for walking me through all the concepts and logic during my PhD studies. I want to say thank you to Professor Fernando Zapatero and Professor Eugene Sorets for serving on my committee. Your feedback and insightful comments help me reshape my research findings.

I am grateful to all the instructors and professors who kindly shared expertise in different subjects, helping me to build a well-rounded background knowledge for future research. Thank you to my cohorts and friends at Boston University for sharing both academic ideas and memorable moments in our everyday life.

Lastly, I want to express gratitude to my parents for their understanding and unconditional support. It is my luck to have parents like you two.

Fanwei Wei

TWO ESSAYS ON EMPIRICAL ASSET PRICING

FANWEI WEI

Boston University, Questrom School of Business, 2025

Major Professor: Andrea Vedolin, PhD
Dean's Research Scholar
Professor of Finance

ABSTRACT

This dissertation consists of two papers on asset pricing. The first paper constructs a new framework based on recombinant tree structure to recover asset price distributions and associated discount factors from state price information. The recovery framework is compatible with common price process and utility assumptions. The recovered results are closely aligned with data observed from market.

The second paper uses Gibbs sampler to estimate the transaction costs from price impact in addition to bid-ask spread, and re-examines anomaly profitability after newly estimated TC. Cost mitigation methods and maximum market capacity with mitigation techniques are also studied for each anomaly. TC from both spread and price impact rules out most acclaimed anomalies. However, proper cost mitigation methods can restore their profitability.

Contents

List of Tables

List of Figures

List of Abbreviations

CDF	.....	Cumulative Distribution Function
MLR	.....	Multinomial Logistic Regression
MVE	.....	Mean Variance Efficient
RN	.....	Risk Neutral
SDF	.....	Stochastic Discount Factor
SPX	.....	S&P 500 Index
TC	.....	Transaction Cost
TS	.....	Time Series

Chapter 1

Introduction

This dissertation contains two papers about empirical asset pricing. In the first paper, I recover physical price distribution and path probability of asset returns, along with marginal utility and subjective discount factor of a typical investor. The recovery is realized by a new recombinant tree framework which is compatible with a variety of price and utility assumptions. Price recovery studies started from Ross (2015) by assuming the researcher can observe state prices of all future states, including those not realized, to collect sufficient information for price recovery. Borovička et al. (2016) mentions Ross (2015) assumptions are too strong that it rules out all realistic price models, making it infeasible for the real world. They instead use Perron-Frobenius theory to separate the probability from stochastic discount factor (SDF) component. Jensen et al. (2019) further develops the recovery theorem by releasing assumptions on time-homogeneity and Markovian behaviors. Using Sard's theorem and time-separable utility assumption, Jensen et al. (2019) shows the sufficient conditions for a possible recovery with fixed number of states across different horizons. My research further shreds off time-separable utility assumption by modeling the underlying price process with expanding tree structures. In this new tree recovery framework, closed form solutions for path probabilities are available under risk neutral measure, or more generally, whenever marginal utility for each state is given. Recovered densities and risk aversions are also consistent with data observed from the real world. To evaluate the goodness of recovered estimates, I then construct delta-neutral option portfolios

from recovered price distributions. The recovery based pricing is more accurate than observed market price and such portfolios can generate consistent significant excess returns.

The second paper focuses on a more traditional empirical topic about anomaly returns, but with new estimates for transaction costs. In this study, I use Bayesian Gibbs sampler to estimate the price impact of U.S. stock market in recent 20 years, and re-examine the profitability of 9 anomalies in addition to Fama and French (1992) four factors. The significance tests are conducted after accounting for both proportional spread and non-proportional price impact with my estimated price impact. Hasbrouck (2009) first generalizes the Bayesian Gibbs method for equity TC estimates. This method requires only daily closing price data for target asset and thus gives more flexibility when intraday data is not available. For anomaly significance, Harvey et al. (2016) questions the significance of most anomalies and suggests higher thresholds of t-statistics should be used for significance tests. As a consequence, they show most anomalies are in fact insignificant under higher thresholds, even before TC analysis. Novy-Marx and Velikov (2016) studies 20 anomalies in addition to Fama and French (1992) four factors with linear transaction cost, but leave non-linear price impact unaccounted. In this study, I show such price impact can erase the profitability for large institutional traders. I also consider the effects of transaction cost mitigation methods and maximum capacities to attract new capital before the anomalies become statistically insignificant. With cost mitigation techniques, the Sharpe ratios decrease for all anomalies, but as turnover rates drop, more anomalies become statistically significant. Price impact also plays a role in the effectiveness of anomaly strategies, especially for large funds.

The dissertation is arranged in the following manner. Chapter 1 highlights the main structure of this dissertation and main findings from both research papers.

Chapter 2 consists the paper of price recovery under recombinant tree framework. The theoretical framework of new recovery method is explained in the first section, followed by empirical recovery results with past 15 years S&P 500 index data. The third section gives two applications based on recovery results and shows superiority of recovered belief as opposed to the market consensus. Chapter 3 discusses the paper of profitability of anomalies with transaction costs and price impact. I introduce the new cost models developed to estimate price impacts in the first section. The estimated results with past 20 years U.S. stocks data are illustrated in section 2. Section 3 compares a new measurement for return significance proposed by Novy-Marx and Velikov (2016) compared with traditional alpha, and evaluate the after cost profitability of acclaimed anomalies. The chapter is finished by analyzing different TC reduction techniques, and the maximum market capacity for each anomaly with those techniques. Chapter 4 concludes both papers with major empirical findings and future improvements.

Chapter 2

Price Recovery under Recombinant Tree Framework

2.1 Introduction

Equity prices are forward looking. Research and industry have long been interested in extracting summary statistics, notably expected return and implied volatility, of future returns from current asset prices. In this paper, I consider a more general question of recovering the underlying return distributions, together with risk preferences of a typical investor. This strand of research started from Ross (2015) by assuming an economy where all state prices, realized or not, can be observed. This strong assumption is required to address the main challenge of separating state probability from associated pricing kernel. Carr and Yu (2012) proves the same result in Ross (2015) using a numeraire portfolio. Borovička et al. (2016) highlights the unrealistic assumptions from Ross (2015) and use Perron-Frobenius theory (See Meyer (2000)) to isolate the probability from stochastic discount factor (SDF) component. Jensen et al. (2019) expands the recovery theorem by removing assumptions on time-homogeneity and Markovian behaviors. Using Sard's theorem and time-separable utility assumption, Jensen et al. (2019) incorporates state probability of different periods in an equation system and solves them in one go. My research further shreds off time-separable utility assumption by modeling the underlying price process with expanding tree structures, and return probabilities are solved sequentially for each

studied period.

The benchmark Ross (2015) recovery model can be illustrated in a one-period setting. Ross assumes a fixed finite number of states s labeled from $s = 1, ..., N$ for each period 0 and 1 with associated state transit probability $p^{s0,s1}$. Discount factor m is given by time-separable utility

$$m^{s0,s1} = \beta \frac{v_{s0}}{v_{s1}}, \tag{2.1}$$

where subjective DF $\beta > 0$ and marginal utility v_s are unique for each state s. Two components are bridged by state price equations, also called the fundamental equation in asset pricing

$$\pi^{s0,s1} = m^{s0,s1} p^{s0,s1}. \tag{2.2}$$

For an economy with N states, we have $(N^2 - N)$ unknowns from transit probabilities (unit sum of probability from each of N starting states reduces number of unknowns by N), 1 unknown subjective DF β and $(N - 1)$ unknowns from ratios of marginal utility. Thus, there are in total N^2 unknowns. Ross (2015) proves under the following assumptions:

(1) all N^2 state prices π, realized or not, can be observed from market data, leading to N^2 equations; and

(2) matrix of state prices $\Pi = \{\pi^{s0,s1}\}_{N \times N}$ is positive and irreducible,

there exist a unique set of solutions for subjective DF, state transit probability (or path probability) and marginal utility. The proof relies on Perron-Frobenius theorem, which bounds the maximum characteristic root of price transition matrix P. For multi-period settings, Ross further assumes both transit matrix P and marginal utility vector v are time homogeneous.

The Ross (2015) recovery first sheds light on separating discount factor from subjective return distribution, but is widely criticized for at least two reasons. Borovička

et al. (2016) first points out the Ross recovered probability "does not mean what you think it means". Recall Perron-Frobenius theory suggests for SDF $m_{t,t+1}$, there exists a positive eigenvalue λ and associated positive eigen-function f such that

$$\mathbf{E}_t\left(m_{t,t+1}f_{t+1}\right) = \lambda f_t, \tag{2.3}$$

where probabilities $p_{t,t+1}$ are implicit in the expectation. By decomposing the SDF $m_{t,t+1}$ into a martingale part m^* and a transitory part $\tilde{m}$ as

$$m^*_{t,t+1} = \frac{m_{t,t+1}f_{t+1}}{\lambda f_t}, \tag{2.4}$$

$$\tilde{m}_{t,t+1} = \frac{\lambda f_t}{f_{t+1}}, \tag{2.5}$$

we have $m_{t,t+1} = m^*_{t,t+1}\tilde{m}_{t,t+1}$ and $\mathbf{E}_t m^*_{t,t+1} = 1$. Note for state s at time t and corresponding state price π^s_t,

$$\begin{aligned}
&\int_{s_{t+1}} \pi^{s_t,s_{t+1}}_{t,t+1} f^{s_{t+1}}_{t+1} ds_{t+1} \\
&= \int_{s_{t+1}} m^{s_t,s_{t+1}}_{t,t+1} p^{s_t,s_{t+1}}_{t,t+1} f^{s_{t+1}}_{t+1} ds_{t+1} \\
&= \lambda f^{s_t}_t.
\end{aligned} \tag{2.6}$$

Ignoring subscripts, the decomposition gives

$$mp = \tilde{m}\cdot m^* p := \tilde{m}\tilde{p}, \tag{2.7}$$

where $\tilde{p} = m^* p$, and $\mathbf{E}m^* = 1$. In other words, m^* is the Radon-Nikodym derivative from measure p to $\tilde{p}$. Borovička et al. (2016) argues the Ross (2015) recovery from state price equations actually identifies the product of $\tilde{m}\tilde{p}$, rather than the product with supposed physical density on the LHS. The products are always the same, but not necessary the component terms. Only in the degenerated case where martingale

term $m^* = 1$ almost surely, the Ross (2015) recovery correctly identifies the physical probability. However, this case rules out almost all realistic models. The other stream of criticism on Ross (2015) recovery concentrates on its unrealistic assumptions. Even in the case of $m^* = 1$, it is almost impossible to observe all state prices originated from unrealized states on date 0. In aware of this, Ross uses observed state prices of different starting states across testing period as a proxy for the unobserved parallel universe at each studied starting time. This proxy requires time homogeneity on physical probability, which is also unrealistic to the market we observe and participate in.

Jensen et al. (2019) corrects and develops the Ross recovery by releasing time homogeneity constraint and discussing solution properties for different economies. They assume an economy with identical N states for each of T periods. Matrix $\Pi = \{\pi_t^s\}_{T\times N}$, $P = \{p_t^s\}_{T\times N}$ stands for state prices and transition probabilities from observed starting state at period 0 to terminal state s at period t, respectively. A typical investor has time-separable utility. Thus, the pricing kernel can be expressed as

$$m_{t,t+\tau}^{s0,s1} = \beta^\tau \frac{v_{s1}}{v_{s0}}, \tag{2.8}$$

for subjective DF β and marginal utility v. The state price equations can be written in matrix form

$$\Pi = DPV, \tag{2.9}$$

where discount factor diagonal matrix $D = diag\left(\beta, \beta^2, ..., \beta^T\right)$ and marginal utility ratio vector $V = \left(1, \frac{v_2}{v_1}, ..., \frac{v_N}{v_1}\right)'$. Using Sard's theorem in linear algebra, they prove for such economy of state N and period T,

(1) if $N > T$, there exist a continuum of solutions;

(2) if $N = T$ and the submatrix of Π excluding the first column has full rank, there exist at most N solutions; and

(3) if $N < T$, there is no solution almost surely, unless the the state price matrix Π is exactly generated by a given set of solution DPV. In this case, there is a unique solution.

The result incorporates Ross (2015) recovery as a special case when $N = T$ and illustrates when a recovery is possible. One limitation is that the economy is restricted to a fixed number of states across periods as forecast horizon stretches longer into future. In reality, we would expect an expanding price range as the recovery horizon increases.

My research contributes to the literature by changing the studied economy from matrix structure, i.e. fixed number of states across periods, to recombinant multinomial tree structures, thus allowing expanding price ranges across different periods. Since the tree model regulates price trajectory, I also aim at recovering the one-period path probability at each possible state (t, s_t), where Jensen et al. (2019) only considers state probabilities at each horizon from observation date. Once path probabilities are revealed, the state probability is given by the sum of all possible paths leading to the terminal state. Different from Jensen et al. (2019), recovering each path probability makes number of unknowns always surpasses number of state price equations regardless of the tree complexity, i.e. number of studied periods (tree layers) and possible outcomes (tree edges) expanding from each state (tree node). There is in general no hope to solve such equation systems without further constraints. To reduce number of unknowns, I assume parametric forms of price and utility functions, such that the problem is reformatted from solving path probability and marginal utility to solving a reduced number of process parameters. I also follow the previous literature to define market states by index levels and assume the utility is time homogeneous. Since price parameters can be different at each starting state on the tree, another way to interpret this model is recovering physical probabilities from state-wise recalibrated parametric

models at each future period and each possible state. The final state probability is given by the aggregate parametric processes weighted by previous state probabilities.

I also empirically test the tree recovery framework by a variety of price and utility processes on S&P 500 index (SPX) . The recovered underlying index return distribution has an average of 7% annualized return across past 15-year tested period and an excess annualized volatility for different parametric models compared with observed time-series (TS) data. The volatility is larger than the observed TS index volatility, but consistent with estimates from intraday trade-level data, which incorporates micro-level market noise and thus better reflects the price uncertainty of underlying asset. I also reported statistics for the recovered TS return distribution, and the TS distribution closely matches real data observed from market. The recovered Sharpe ratio centers around 0.5 with negative skewness around and excess kurtosis, indicating a left skewed, fat tailed recovered TS return distribution. One application of such recovery framework is to back out fair option prices from recovered price distribution and marginal utilities. I constructed a delta- and theta-neutral recovered option portfolio that buys low and sells high compared with market option prices. Such portfolios yield consistent positive Sharpe ratio across tested periods and different price models which indicates the recovered result contains more accurate return distribution information than the market price consensus.

The paper is organized in the following structure. Section I introduces the tree recovery framework settings, analyzes the solvability of different tree structures and provides closed-form solutions for path probability under risk-neutral (RN) measure. More generally, the closed-form solution under any known marginal utility is also proved. I then move on to general undecided physical measures and describe the algorithm to solve the recovered results numerically from recovery system. Candidate parametric models for price and utility processes are also listed out and discussed.

Section II shows empirical recovery results with most recent 15-year SPX data. I first discuss two different state price generating methods with their properties, followed by empirical recovery results under different specifications. Robustness tests of recovery results with different tree structures and data sources are also provided. The tree model recovery can generate consistent results across different specifications. Section III moves on with applications. I test profitability of the option portfolios based on recovered underlying density and use recovery result to estimate asset-spanned SDF under incomplete market. The consistent profitability of such portfolios indicates this new recovery provides a better estimate of index return distributions compared with market consensus.

2.2 Theory Framework

2.2.1 Fractional State Price Equations

The purpose of this study is to separate price probability from the associated discount factor. Different from previous study, the price range should allow more extreme possible returns as forecast horizons extend deeper into future. This is consistent with options data where more extreme strikes are available as maturity grows. Also, we are interested in the behavior of a typical market participant that only cares about her terminal wealth. Ross (2015) and Jensen et al. (2019) both focus on consumption based utilities. Although less studied, terminal wealth based utility better characterizes equity market investors as they mostly care about end-of-horizon portfolio return statistics. Besides, wealth based utility is better adapted to studies for dynamic asset allocation where investors adjust their portfolios over time to maximize long-term benefits. Terminal wealth utility is also more adapted to study decisions in risk management and price changes in crisis market regimes.

Throughout this paper, the reference asset is always SPX level. A typical investor

allocates her initial wealth w_0 optimally between index and risk-free bond, based on her terminal wealth utility $u(w_T)$. Specifically, at starting date, the investor maximizes her utility function over α shares of index to hold until end-of-period T:

$$\max_{\alpha} \mathbf{E}\left[u\left(W_T\right)\right], \quad s.t.\ \beta W_T = \left(W_0 - \alpha X_0\right) + \beta\alpha X_T, \tag{2.10}$$

where $\beta > 0$ is the subjective discount factor, X_0 is current index level, X_T is a random variable for index level at T, and $(W_0 - \alpha X_0)$ is dollar value in bond holding for the optimal portfolio. The expectation is taken over all possible terminal states $\omega_T^* \in \Omega_T$ at T, and maximization is over all possible investment decision α. The first order condition

$$\mathbf{E}\left[u'\left(W_T\right) \cdot \left(X_T - \frac{X_0}{\beta}\right)\right] = 0 \tag{2.11}$$

shows

$$X_0 = \mathbf{E}\left[\beta \frac{u'\left(W_T\right)}{\mathbf{E}u'\left(W_T^*\right)} \cdot X_T\right], \tag{2.12}$$

where both expectations are taken over all time T states, and optimal allocation α_T is implicit in terminal wealth $W_T = W_T\left(\omega;\ \alpha_T,\ X_T\left(\omega\right)\right)$. Assuming the market is efficient such that index level X_0 is an accurate proxy for its fair value, the fundamental equation in asset pricing yields SDF

$$m\left(\omega\right) := m\left(W_T\left(\omega\right)\right) = \beta \frac{u'\left(W_T\right)}{\mathbf{E}u'\left(W_T^*\right)} \tag{2.13}$$

for each time T. Carr and Yu (2012) shows under this market efficiency assumption, the recovered probability measure coincides with physical measure. Under terminal wealth utility, the discount factor can be expressed as the product of a state-invariant subjective discount factor and a marginal utility coefficient. Intuitively, the SDF always equals to a constant β if no variation exists in marginal utility across different states, i.e. in the RN measure. For any other physical measure, discount factor is

linearly controlled by the associated marginal utility level.

The SDF expression (2.13) may seem clunky at the first sight, but can be insightful after some simple re-arrangements. Multiplying state transit density from current state to state of $W_T = W\left(T,\ X_T\left(\omega\right)\right)$ by corresponding SDF, we have state price

$$\pi\left(\omega\right) = m\left(\omega\right) f\left(\omega\right) = \beta \frac{u'\left(W_T\right) f\left(\omega\right)}{\mathbf{E}u'\left(W_T^*\right)}, \tag{2.14}$$

or discretely for state transit probability q_s,

$$\pi_s = m_s q_s = \beta \frac{q_s u'_s}{\mathbf{E}u'_{s^*}}, \tag{2.15}$$

for terminal state $s = 1, ..., N$. Note summing up the RHS of (2.15) over all states at a given period gives exactly

$$\beta = \sum_{s=1}^{N} \pi_s, \tag{2.16}$$

and

$$\frac{\pi_s}{\sum_{s^*=1}^{N} \pi_{s^*}} = \frac{q_s u'_s}{\sum_{s^*=1}^{N} q_{s^*} u'_{s^*}}, \tag{2.17}$$

for $s = 1, ..., N-1$. Only the first $(N-1)$ fractional state price equations are independent since the last equation can be implied by the unit sum of all fractions. The first conclusion (2.16) shows subjective DF β is directly observable as sum of state prices at any given time. Thus, unlike Ross (2015) and Jensen et al. (2019), subjective DF need not be decided by the recovery system under terminal wealth utility. This is a shortcut result from this specific SDF structure and in general, subjective DF may not be observable from data with other SDF structures. The second equation (2.17) shows for any given time, the fraction in state prices equals fraction in probability weighted marginal utility. No additional assumption about time-separable utility is needed as investor only cares about terminal wealth from the index-bond portfolio. The LHS of (2.17) only consists of state price information and is observable from

options data at starting date, while all unknowns to be solved are on the RHS. The tree framework recovery lines up all such fractional state price equations for all studied periods and aims at finding probabilities together with associated marginal utilities, i.e. solving equation system

$$\frac{\pi_t^s}{\sum_{s^*=1}^N \pi_t^{s^*}} = \frac{q_t^s u_s'}{\sum_{s^*=1}^N q_t^{s^*} u_{s^*}'}, \tag{2.18}$$

for $s = 1, ..., N-1$ and $t = 1, ..., T$. Before moving forward to the recombinant tree framework, let us consider the validity of such equation system. On the one hand, there is no redundant equations in such system since at any given time, there are $(N-1)$ free fraction values in state prices, state probabilities and marginal utility levels. Given any $(N-2)$ equations cannot imply the rest. On the other hand, all equations in the system are consistent, i.e. any set of sub system does not rule out validity of the rest. To see this, note for all equations, the state price information for different period t is all observed from options data at the same date. Thus, all path probabilities $p_{t_1,t_2}^{s_1,s_2}$ from state s_1 to s_2 starting from any future date t_1 to t_2 are spot implied forward probabilities

$$p_{t_1,t_2}^{s_1,s_2} = \frac{p_{0,t_2}^{s_0,s_2}}{p_{0,t_1}^{s_0,s_1}}, \tag{2.19}$$

where the RHS terms are both spot probabilities extracted from starting date state prices. Alternatively, if we use additional data observed at a second observation date t_1 for LHS probability in (2.19), it requires spot implied forward probability (RHS in (2.19)) equals future spot probability. It is a very strong assumption and leads to inconsistency between equations from different periods. Instead, the recovery method only uses data from one period to rule out this conflict. Previous recovery studies all focus on matching the number of equations to number of unknowns for a finite number or even unique solution. However, since state price data are not directly observable from the market, and is instead generated by stochastic processes or fitted with tuning

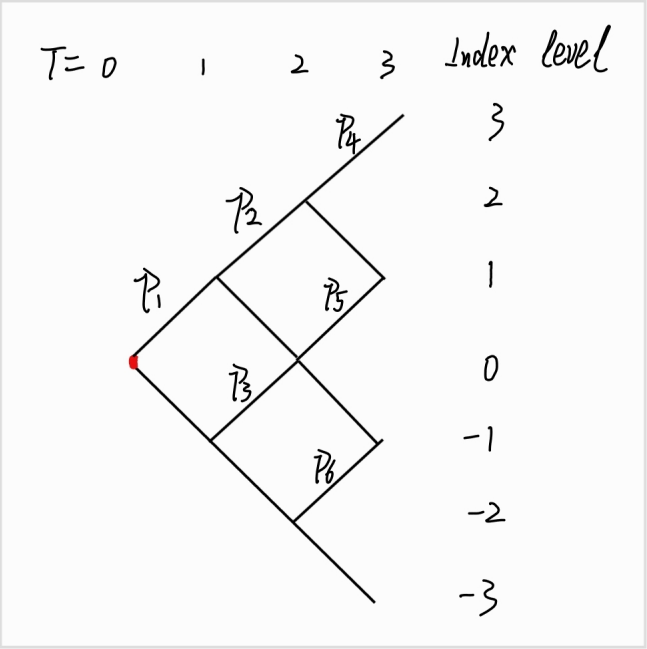

Figure 2·1: Price Recovery under Recombinant Binomial Tree

parameters, it makes more sense to have more state price equations than unknowns, and fit unknowns with an error function. I will discuss in more details about the state price generating process and alternatives in the empirical part. Supplementing constraints for exactly a unique set of solutions is intriguing mathematically, but may not be necessary from a financial point of view.

2.2.2 Recombinant Tree Models

The goal is to recover asset return distribution and pricing kernel in the recombinant tree settings. More specifically, the structure can also recover each sample path probability along the tree edges. In the simplest case, I illustrate the idea with the binomial tree.

Assuming index level follows a binomial tree model, where price either goes up or down by the same amount. The identical price change leaves a recombinant structure. Standing at recovery date 0, we observe the current state s_0 with associated index

level. State prices are generated for all future states s_t from options data observed strictly from date 0. We aim at recovering probability applied to each one-period path (tree edge) and marginal utility at each state (tree node). Compared with Jensen et al. (2019), tree model regulates the price trajectory and allows an expanding state range. As discussed before, all state price data are observed at starting time, and subjective DF β_t is known from the sum of all state prices at each given time. The recovery needs to solve for both path probabilities along each edge and marginal utility at each state. Then state probability $q_t^{s_t}$ is a natural outcome by aggregating multiplications of latest path probability $p_{t-1,t}^{s_{t-1},s_t}$ and starting state probability $q_{t-1}^{s_{t-1}}$ from last period as

$$q_t^{s_t} = \sum_{s_{t-1}} q_{t-1}^{s_{t-1}} \cdot p_{t-1,t}^{s_{t-1},s_t}, \tag{2.20}$$

where the sum is taken over all starting state s_{t-1} accessible to s_t, for any $s_t = 1, ..., N$. Simple counting shows for such 3-period binomial tree illustrated in Figure 2·1, there are 6 probability unknowns, one from each starting state and 7 marginal utility levels, one from each possible state in the tree. The recombinant structure largely reduces possible number of states and thus number of unknowns to be solved. Recall fractional state price equations only requires relative, rather than absolute value of marginal utility, we can assume marginal utility at starting state $v_0 := u_0' = 1$. Since each starting state provides one state price equation and the last equation from each node is already implied from unit sum across fractions, the recovery equation system has 6 equations with 12 unknowns, 6 from probability and 6 from marginal utility. The discrepancy indicates at lease 6 additional constraints are needed for a possible recovery.

For a general multinomial recombinant balanced tree with T levels and $(1 + 2n)$ edges from each node, for example, $n = 1$ for trinomial tree as illustrated in Figure 2·2,

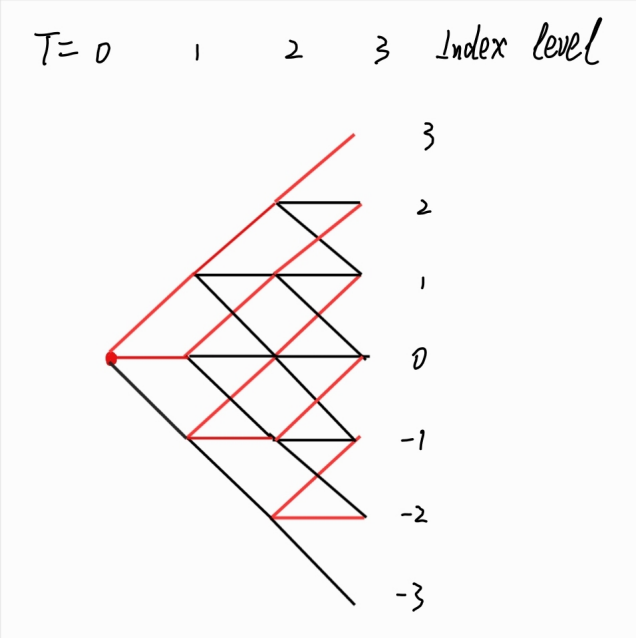

Figure 2·2: Price Recovery under Recombinant Multinomial Tree

there are

$$\sum_{t=1}^{T} (1 + 2nt) = (1 + T)(1 + 2nT) \tag{2.21}$$

different states (t, s_t) across recovered horizons. Ignoring last fractional equation from each period, number of independent equations from tree equals $nT(1 + T)$. Number of path probability unknowns equals number of edges minus number of starting nodes since tail probability for the lowest price change edge from each starting nodes is known from the unit sum of probability. Direct calculation gives $2nT(1 + nT - n)$ path unknowns for a tree of $(1 + 2n)$ possible price outcomes expanding from each node and T levels. For marginal utility, the bottom level of such $(T, 1 + 2n)$ tree has $(1 + 2nT)$ states and thus $2nT$ unknown ratios. In order to match number of equations to unknowns, we need additional

$$\begin{aligned} &2nT(1 + nT - n) + 2nT - nT(1 + T) \\ &= nT(2nT - 2n - T + 3), \end{aligned} \tag{2.22}$$

i.e. $\mathcal{O}(n^2T^2)$ constraints by subtracting number of equations from number of unknowns. As an illustration, to recover price from a trinomial tree ($n = 1$) with $T = 3$ periods displayed in Figure 2·2, there are 12 equations with 18 unknown path probabilities and 6 unknown marginal utility ratios. Thus, 12 additional constraints are needed for a recovery solution with terminal prices distributed across 7 possible outcomes. One may argue one limitation of such framework is unrealistic large number of constraints $\mathcal{O}(n^2T^2)$ needed for a possible recovery, especially when n is large. A workaround is to assume one-period price process from all starting state follows a known rule, for example, parametric processes or pre-defined training rules, where the exact parameters for each sub-tree can be different or re-trained. The path probability is then reformatted into probability of target range under such parametric model. One way to interpret this framework is to consider the tree model as a weighted average of multiple state-wise parametric price processes, each calibrated by data of the corresponding state, then weighted by its starting state probability. The new state probability is then aggregated from all possible paths leading to the target state. The method incorporates more information than simple parametric estimates and is immune to the increase of state numbers from each starting node. For example, assuming lognormal price process along each single edge, all paths extending from the same node are controlled by only two common parameters (μ, σ), regardless of how granular the target price range is divided by number of states.

The recovery for all one-period path probability, rather than state probability of different periods, may seem to add difficulty to the problem. However, it actually allows us to solve the equation system sequentially period-by-period, and thus reduces the system complexity. To see this, let us consider the relation between a set of solutions to path probabilities and marginal utility ratios (p, v) for a tree of level T and by adding one more layer to $T + 1$. All other specifications are otherwise

the same. Note investment decision α_t can be different for each time t, leading to different terminal wealth $W_t(\omega;\, \alpha_t, X_t)$ even for the same state ω. For all path probabilities up to T, financially speaking, investor's belief for future periods after T should not change beliefs in the realized past, even though mathematically, all (p, v) from different periods are jointly solved in one equation system. Specifically, all forward probability $p_{T,T+1}^{s_T,s_{T+1}}$ starting from T, which are based on events happening between T to $T+1$, should not reversely affect all historical probabilities in the previous realized T level tree, which are strictly based on events happening before T. More generally, the set of solution for path probability up to level T should also be part of the solution when an additional layer is added to the tree structure. For marginal utility ratios v up to time T, once path probabilities are fixed, we also get each state probability q_s. Since the number of equations always equals number of unknowns in v, the fractional state price equations (2.17) is reduced to an linear system of marginal utilities v_s for $s = 1, ..., N-1$. Thus, the marginal utility ratio vector associated with probability is also uniquely solved. Because all equations up to time T make up a subset of equations up to $T+1$, and probability solutions up to T must be inherited in probability solutions up to $T+1$, the same inheritance must also hold for marginal utility ratio solutions for a given level of terminal wealth. The above recursive relations between solutions of different tree levels reveals that we can slice the tree by levels and find solutions period-by-period, rather than jointly solving (p, v) across all levels from one combined equation system as in Jensen et al. (2019). This observation actually makes the system easier to solve. Specifically, by adding one more layer from T to $T+1$, we have additional $2n(1+T)$ equations, $2n(1+2nT)$ path probability unknowns and $2n(1+T)$ marginal utility ratio unknowns. Similar to previous analysis, additional $2n(1+2nT)$ restrictions are required for a possible set of solutions at level $T+1$.

2.2.3 Closed-form Solutions under RN Preference

Let us first consider the simplest case where the price follows a recombinant binomial tree for a RN investor. For each possible node (t, s_t) excluding the last state in each period, set $\pi_t = \sum_{s=1}^{N} \pi_t^s$, we have

$$r_t^s := \frac{\pi_t^s}{\pi_t} = \frac{q_t^s v_s}{\sum_{s_t} q_t^{s_t} v_{s_t}}, \quad \forall t, \tag{2.23}$$

where $q_t^{s_t} = \sum_{s_{t-1}} q_{t-1}^{s_{t-1}} p_{t-1,t}^{s_{t-1},s_t}$ and RN preference implies $v_s = 1$ for any s. The above fractional state price equations are reduced to

$$r_t^s = \frac{q_t^s}{\sum_{s_t} q_t^{s_t}} = q_t^s, \tag{2.24}$$

by unit sum of probability on denominator. To get path probability, following the centered symmetric state notation in Figure 2·1, starting from $r_0^0 = 1$ and $r_1^1 = p_{0,1}^{0,1}$, mathematical induction gives the up probability

$$p_{t,t+1}^{s,s+1} = \frac{1}{r_t^s} \cdot \left(\sum_{j=0}^{\lfloor \frac{t-s}{2} \rfloor} r_{t+1}^{s+1+2j} - \sum_{j=1}^{\lfloor \frac{t-s}{2} \rfloor} r_t^{s+2j} \right), \tag{2.25}$$

for any node (t, s_t) and $t \geq 0$ on the binomial tree. The down move probability is then

$$\begin{aligned} p_{t,t+1}^{s,s-1} &= 1 - p_{t,t+1}^{s,s+1} \\ &= \frac{1}{r_t^s} \cdot \sum_{j=0}^{\lfloor \frac{t-s}{2} \rfloor} \left(r_t^{s+2j} - r_{t+1}^{s+1+2j} \right). \end{aligned} \tag{2.26}$$

See proof in Appendix A. The first observation is that RN state probability equals the corresponding state price fraction due to $v_s = 1, \forall\, s$. This observation is independent of tree structure so long as investor is RN. The binomial tree further regulates the model by restricting up to two starting states that can result in each terminal outcome.

Thus, number of path probability unknowns equals number of states, enabling a unique solution without any assumptions needed on the price process.

Move on to general $(1+2n)$ multinomial tree with RN preference. State probabilities are still directly observable from $q_t^s = r_t^s$, and all path probabilities at $T=1$ exactly equals to the corresponding state probability. More generally, both path and state probabilities coincide if the state is solely accessible through the path. Following the same logic, top-mover path probability at each period is given by corresponding top-mover state probabilities

$$p_{t,t+1}^{nt,n(t+1)} = \frac{r_{t+1}^{n(t+1)}}{r_{t+1}^{nt}}, \quad \forall t \geq 2. \tag{2.27}$$

Starting from the second-to-top states, we have more than 1 unknown probability associated with each fractional state price equation. In order to get solutions, each equation needs to be associated with at least one unknown to make the most use of each independent data, and number of unknowns needs to match number of equations. There are many ways to impose restrictions on the excess number of unknowns. The recovery framework imposes restrictions such that they are both reasonable and can lead to an easy way to solve for path probability and marginal utility unknowns. Simple counting shows at level t, number of equations equals $2nt$ and number of starting state at starting period is $[1+2n\,(t-1)]$. Thus, one possible way is to assign 1 free parameter to drive all $(1+2n)$ 1-period path probabilities (i.e. reduce number of unknowns by $2n-1$) on each of first $2n\,(t-1)$ nodes, leaving the last node with remaining $2n$ independent probabilities to drive the tail behavior. In other words, we do not impose any functional restrictions on the tail probabilities and let it completely decided the state price data. Thus, for all edges expanding from nodes in period $(t-1)$, let us impose restrictions such that all probabilities expanding from the same node follow the same process of one common parameter

to be decided by the recovery system. With the most negative price change case at each period as the only exception, for which all $2n$ probabilities expanding from this node are independent. This is by any means not the only way to assign unknowns among limited equations. One upside of such assignment is that each state price data at level t is associated with unknown parameters from related starting states, making the most use of market data. The second strength is the framework gives high degree of freedom on the tail scenarios, making it more amenable to fit possible market catastrophe. As Barro (2006) and Weitzman (2007) point out, extreme but rare market downturns work like the dark matter in asset pricing models. Such rare events can change the perceived market return, and fail to take disasters into account can distort the model calibrations from the real data observed. Another advantage of this tree-structured, dynamically-calibrated parametric framework is although each sub-tree follows one specific model, giving a larger degree of freedom to calibrate catastrophic events allows this model to capture extreme market movements, which standard models like the Black-Scholes or other basic diffusion models may not be able to handle effectively.

Since probability of top path at each period is observable from state prices, we focus on finding path probabilities for all other edges. Moving to the second-to-top node, which corresponds to the second-best price change terminal state in unit period of time, the fractional equation gives probability of top-mover from the second-to-top node in the previous period. Equivalently, we can always solve for the probability of top-mover expanding from each node and express all other path probabilities as a function of the top-mover path probability from the same starting state. See Figure 2·2, all red edges stand for an unknown probability to be solved and all black edges are restricted as a function of the red edge probability from the same node. Mathematically, except for the last node from previous state $t-1$, set $p_{t,t+1}^{s,s+n} = p$ for

top-mover as the pivot probability to be solved by the recovery system and each of following path expanding from node (t, s) as

$$\begin{cases} p_{t,t+1}^{s,s+n-1} = f_{t,t+1}^{s,s+n-1}(p), \\ \qquad\qquad ... \\ p_{t,t+1}^{s,s-n} = f_{t,t+1}^{s,s-n}(p), \end{cases} \tag{2.28}$$

where all functions f reflect exogenous relations between pivoting top-mover probability and other corresponding paths. Once we know the top-mover probability p, the recovery system solves all path probabilities expanding from the same starting state (t, s). Generally, similar to the binomial case, mathematical induction gives

$$p_{t,t+1}^{s,s+k} = \frac{1}{r_t^s} \cdot \left[r_{t+1}^{s+k} - \sum_{j=1}^{(k+n)\wedge(nt-s)} r_t^{s+j} \cdot f_{t+1}^{s+j,s+k}\left(p_{t+1}^{s+j,s+j+n}\right) \right], \tag{2.29}$$

where the upper bound is given by

$$\begin{cases} s + j \leq nt \text{ at top mover in level } t \\ s + j \leq s + k + n \text{ at bottom mover from starting state } (t, s + j). \end{cases} \tag{2.30}$$

k is usually $+n$ for unknown pivot path, but ranges from $+n$ to $(1 - n)$ in the last expanding node since all edges in this node are freely decided by the system. The above closed form solution with exogenous relations f within each starting node allows us to solve for p at each level t sequentially from top to bottom, without solving any equation system.

2.2.4 Closed-form Solutions with Known Marginal Utility

The closed form solutions for path probabilities can be further generalized to any preference with known marginal utility. Suppose price process follows general $(1 + 2n)$ multinomial tree structure, investor's marginal utility v_s for all state s are exogenously

given, with RN as a special case of $v_s \equiv 1$. Then fractional state price equations (2.17) at period t plus unit sum of state probability gives

$$\begin{cases} r_t^s := \dfrac{\pi_t^s}{\pi_t} = \dfrac{q_t^s v_s}{\sum_{k=-nt}^{nt} q_t^k v_k}, \quad -nt \leq s \leq nt \ \ and \ s \neq 0, \\ \displaystyle\sum_{k=-nt}^{nt} q_t^k = 1, \end{cases} \tag{2.31}$$

for observed r_t^s and known v_s. Solving for q_t^s from the above system, we have

$$q_t^s = r_t^s \cdot \left[\frac{v_0/v_s}{1 + \sum_{k \neq 0} \left(\frac{v_0}{v_k} - 1 \right) r_t^k} \right], \tag{2.32}$$

for all state s at period t. Same as before, since only ratios of marginal utility rather than its absolute value is relevant, set marginal utility at starting state $v_0 = 1$, we have

$$q_t^s = \frac{r_t^s}{v_s} \cdot \left[\frac{1}{1 + \sum_{k \neq 0} \left(\frac{1}{v_k} - 1 \right) r_t^k} \right], \tag{2.33}$$

for any $-nt \leq s \leq nt$ at period t. Specifically, if investor is RN, $v_s \equiv 1$ and q_t^s degenerates to r_t^s. Then same as before, assuming exogenous functional relations (2.28), the recursive relations between path and state probability (2.20) gives closed form path probability

$$p_{t,t+1}^{s,s+k} = \frac{1}{q_t^s} \cdot \left[q_{t+1}^{s+k} - \sum_{j=1}^{(k+n) \wedge (nt-s)} q_t^{s+j} \cdot f_{t+1}^{s+j,s+k} \left(p_{t+1}^{s+j,s+j+n} \right) \right], \tag{2.34}$$

where all state probabilities q are given by (2.33). Specifically for binomial tree structure, up move path probability is

$$p_{t,t+1}^{s,s+1} = \frac{1}{q_t^s} \cdot \left(\sum_{j=0}^{\lfloor \frac{t-s}{2} \rfloor} q_{t+1}^{s+1+2j} - \sum_{j=1}^{\lfloor \frac{t-s}{2} \rfloor} q_t^{s+2j} \right), \tag{2.35}$$

for state probability (2.33).

2.2.5 General Unknown Preferences

Getting back to general case where both price distribution and preference are undecided. Introducing risk preference into the recovery framework augments solution from all path probabilities p to the set of path probabilities p together with state specific marginal utility v at each state (t, s_t). Following previous discussion, the recovery solution (p, v) is revealed period by period. Assuming we already have solutions up to period T, by adding one more period $(T+1)$, we have additional $2n\,(T+1)$ state price equations and $(1+2nT)$ starting states, i.e. $(1+2nT)$ new path probability unknowns from previous period. This implies we can allocate at most $(2n-1)$ unknowns on period specific utility function whenever a new period is added. Similar to the binomial case, for each starting states, we want at least 1 free parameter in probability unknowns to let the data speak, leaving at most $(2n-1)$ free parameters to govern marginal utility v across states. Since average number of total parameters for each starting state is $\frac{2n(T+1)}{2nT+1}$, which is less than 2, it is reasonable to assign 1 probability parameter to most of starting states. Specifically, let us assign parameters decided by the recovery system in the following way. Assuming marginal utility $v\left(W_t^s;\,\gamma\right)$ at state (t, s) is decided by parameter vector $\gamma = (\gamma_1, ..., \gamma_k)$, where number of utility parameters is $0 \leq k \leq 2n-1$. Except for the last starting node which represents tail probabilities, all path probability starting from node s_{t-1} at $(t-1)$ are driven by a price process with only 1 free parameter μ such that

$$p_{t-1,t}^{s_{t-1},s_t} = p_{t-1,t}^{s_{t-1},s_t}\left(\mu_{s_{t-1}}\right), \quad \forall s_{t-1} - 2n \leq s \leq s_{t-1} + 2n. \tag{2.36}$$

All other parameters in the price model are pre-calibrated exogenously from market data. The exogenous calibration of parameters not decided by recovery system is crucial to the recovery outcome. I always use the most recent data corresponding to each state for parameters pre calibration, and dynamically update the parameter

set at each recovery date. The allocation leaves $(2n-k)$ free parameters for the last starting state and we can assume all path probabilities starting from the previous bottom-mover are jointly driven by these $(2n-k) \geq 1$ parameters.

Before listing up candidate price processes and utility functions, let us consider how to tell the goodness of a recovered result (p, v), or what is the benchmark for recovered underlying asset price distribution? There is no definite answer to this. Unlike return, the underlying distribution that drives the return is never observed, what we observe is always one realization of it. Using observed frequency across different time as a proxy for probability requires time homogeneity, which is usually rejected by data and exactly the assumption this framework attempts to avoid. As we will see in the application part of this paper, a workaround is to test the predictive power of the recovered expected return or profitability of some portfolios constructed from recovered terminal price density. With recovered return as a more accurate estimate for future asset price behavior, the derivative prices implied by recovered distribution should also more accurately reflect their market value than observed market price. By constructing a portfolio based on such recovery implied derivative prices, a better recovery should result in a more significant or robust excess returns on such portfolios.

For any one-period subtree, the path probability along each edge is given by CDF of corresponding terminal price range. Specifically, let tree (T, n) be a tree of T periods and each node (starting state) has $(1+2n)$ edges, then terminal level T contains $(1+2nT)$ terminal states. Set price at initial state $(0, 0)$ as current index level X_0 and set price process CDF with parameter μ at period t as $F_t(x;\ \mu)$. At time 0, we observe possible terminal price range from options strike range as $(X_{min},\ X_{max})$ and price range length for each state is given by

$$d = \frac{X_{max} - X_{min}}{1+2nT}. \tag{2.37}$$

The states are equally divided within the strike range by index level for all tree levels so that tree is recombinant across different periods. Index midpoint at each state is given by the relative position compared with starting state 0

$$X_s = X_0 + d \cdot s, \quad -nT \leq s \leq nT, \tag{2.38}$$

and finally the path probability from CDF with parameter μ for each path connecting s_t and s_{t+1} is

$$\begin{aligned} p\left(s_t, s_{t+1};\, \mu\right) =& F\left[s_t + d \cdot \left(s_{t+1} - s_t + 0.5\right);\, \mu\right] \\ & - F\left[s_t + d \cdot \left(s_{t+1} - s_t - 0.5\right);\, \mu\right], \end{aligned} \tag{2.39}$$

for any $s_t - n \leq s \leq s_t + n$.

In the following study, I test different types of common price process models including jump diffusion models, time-series (TS) models, machine learning (ML) and non-parameter methods. Since only one free price process parameter is allocated for each subtree, for diffusion parametric models, we always leave drift parameter as the unknown to be estimated from recovery system. All other model parameters are estimated from market data in real time. This allows the system to control level of recovered return. For TS based price models, drift levels are exogenous calibrated and weight parameter on the TS conditional return volatility is carried over to the system. For ML and non-parametric models, they are structured or trained with one tuning parameter. The recovery system optimizes over possible tuning parameter range that best fits the fractional state price equations implied from options data.

2.2.6 Candidates for Underlying Price Process

The first candidate is classic geometric Brownian motion (GBM) price process. GBM is a stochastic process widely used to model stock prices in the Black-Scholes world. It

assumes asset log returns are continuously compounded and follow a normal distribution with constant volatility. GBM is analytically tractable for modeling returns with drift and uncertainty, and thus is widely used as benchmark model with closed-form solutions for asset pricing. Following previous notations, let

$$\ln \frac{X_{T+1}}{X_T} \sim \mathcal{N}\left(\mu, \sigma^2\right), \tag{2.40}$$

then one period price process is given by

$$X_{T+1} = X_T \cdot \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) + \sigma W_1\right], \tag{2.41}$$

where W_1 is a unit time standard Brownian motion, and path probability $p = p\left(s_t, s_{t+1}, \sigma;\, \mu\right)$ is given by (2.39). At each starting node (t, s), the drift parameter μ is estimated by the recovery system and state volatility σ is pre-estimated by the implied volatility from options data of the corresponding strike. As discussed before, such recovered price can be viewed as generated from an weighted averaged and dynamically estimated distribution, aggregated from each sub GBM process. Compared with traditional GBM model, this tree-structured aggregated distribution uses tree-structure to regulate price range and assign model weights across possible starting states, thus giving more flexibility on estimated results. Also, this method incorporates forward looking implied volatility data at different states, using more data inputs than the standard model.

The second price model is jump diffusion process. Diffusion with jumps is a natural extension from GBM and introduces discontinuity or sudden change effects in the price process. By allowing for abrupt changes in prices from crashes or other events, jump processes better capture the extreme behaviors especially tail risks that traditional GBM fails to represent. For a given maximum allowed number of jumps M in unit time, let W be a standard Brownian motion, $N_1, ..., N_M$ be M i.i.d. Poisson

processes with identical jump intensity λ and different jump size Y_i taking values uniformly from non-zero $-1 < y_1 < ... < y_M$. The uniform distribution of jump size is a direct result from i.i.d jump processes. Set total number and size of jumps at period t as

$$\begin{aligned} N(t) &= \sum_{m=1}^{M} N_m(t), \\ Q(t) &= \sum_{m=1}^{M} y_m N_m(t), \end{aligned} \tag{2.42}$$

then total jump size in unit time $N(t)$ has intensity $M\lambda$ and jump size

$$Q(t) = \sum_{m=1}^{M} Y_m \tag{2.43}$$

is a compound Poisson process. Then averaged jump size

$$\beta = \mathbf{E}Y_m = \frac{1}{M}\sum_{m=1}^{M} y_m, \tag{2.44}$$

and we have $Q(t) - \beta\lambda M \cdot t$ is a martingale. Set

$$dX_t = \mu X_t dt + \sigma X_t dW_t + X_{t-} d\left[Q(t) - \beta\lambda M \cdot t\right], \tag{2.45}$$

the one-period price process is given as

$$X_{T+1} = X_T \cdot \exp\left[\left(\mu - \beta\lambda M - \frac{\sigma^2}{2}\right) + \sigma W_1\right] \cdot \prod_{i=1}^{N(1)} (1 + Y_i), \tag{2.46}$$

with $p = p(s_t, s_{t+1}, \sigma, \lambda, \beta;\ \mu)$ as the path probability from state s_t to s_{t+1}. Similar to continuous diffusion case, state volatility is pre calibrated by implied volatility from corresponding option strikes. Jump parameters (λ, β) are pre-calibrated from intraday total and institutional trade volume data. Jump intensity is approximated by the fraction of institutional trade numbers as opposed to total trade numbers and

averaged jump size is pre-calibrated by the fraction of institutional trade volume of a given time period. Return level μ is the only parameter to be solved by the recovery system.

A natural extension of the above standard jump model is Kou (2002) double exponential jump model. The process extends the jump diffusion model by using a double-exponential distribution for the jump sizes, which calibrates upside and downside risks separately. The double exponential jump distribution gives the model more flexibility in calibrating fat tails and skewness in stock price distributions, making it a better fit for pricing options, especially out-of-the-money options. Let W be a standard 1-dimensional Brownian motion, N be a Poisson process with intensity λ. $\{V_i\}$ is a sequence of i.i.d. non-negative random variables such that $Y = \ln(V)$ is asymmetric double exponential distributed with density

$$f(y) = p\eta_1 e^{-\eta_1 y} \cdot \mathbf{1}_{\{y \geq 0\}} + q\eta_2 e^{-\eta_2 y} \cdot \mathbf{1}_{\{y<0\}}, \tag{2.47}$$

where $\eta_1 > 1, \eta_2 > 0$ and $p, q > 0, p + q = 1$. The model calibrates both upward and downward jumps with exponential distribution of mean $\frac{1}{\eta_1}$ and $\frac{1}{\eta_2}$, respectively. p, q stands for conditional upward and downward jump probabilities, conditioned on jump takes place. Set

$$dX_t = \mu X_t dt + \sigma X_t dW_t + X_{t-} d\left[\sum_{i=1}^{N(t)} (V_i - 1)\right], \tag{2.48}$$

we have

$$X_{T+1} = X_T \cdot \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) + \sigma W_1\right] \cdot \prod_{i=1}^{N(1)} (V_i), \tag{2.49}$$

$$\mathbf{E}V = \mathbf{E}e^Y = p\frac{\eta_1}{eta_1 - 1} + q\frac{\eta_2}{\eta_2 + 1}, \tag{2.50}$$

leading to $p = p(s_t, s_{t+1}, \sigma, \lambda, p, q, \eta_1, \eta_2;\ \mu)$ as the path probability from state s_t to

s_{t+1}. The exogenous calibration for all parameters except for μ is similar to the process in standard jump model. The only difference is double exponential model separates upward and downward jumps, which are calibrated separately by buy and sell counterparts of intraday total and institutional trade data observed at each state (t, s). Then all path probabilities are only functions of drift parameter μ at each starting state.

The second strand of price models are time series processes, especially the ARCH and GARCH processes that capture the heteroskedasticity of return volatility. The ARCH(1) model captures time-varying volatility, where the conditional variance of returns in one period depends on the lagged squared return. It models volatility clustering observed in financial markets and is useful for understanding heightened or reduced market volatility in different market regimes. Assume the one period underlying asset return follows ARCH(1) model such that

$$\begin{cases} r_t = \sigma_{t|t-1} \cdot \epsilon_t, \\ \sigma^2_{t|t-1} = \omega + \alpha r^2_{t-1}, \end{cases} \tag{2.51}$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$ and (α, ω) are parameters to be estimated. Note conditional one-period return volatility $\sigma_{t|t-1}$ is not observable, let us set

$$\eta_t = r^2_t - \sigma^2_{t|t-1}, \tag{2.52}$$

such that $\eta_t \perp \eta_{t-1}$, $\eta_t \perp r_{t-1}$, and $\mathbf{E}\eta_t = 0$. This transforms ARCH(1) of return process into AR (1) of squared return

$$r^2_t = \omega + \alpha r^2_{t-1} + \eta_t. \tag{2.53}$$

Set σ as the long term volatility of index return, we have

$$\sigma^2 = \omega + \alpha\sigma^2, \tag{2.54}$$

which gives $\omega = (1-\alpha)\,\sigma^2$, $0 \leq \alpha < 1$. This is the equivalent condition that AR(1) of r_t^2 is a week stationary process. From the definition of r_t in (2.51), we have

$$r_t \sim \mathcal{N}\left(0, (1-\alpha)\,\sigma^2 + \alpha r_{t-1}^2\right). \tag{2.55}$$

Set return CDF $F(x;\,\alpha) = \Phi\left(\frac{x}{\sqrt{(1-\alpha)\sigma^2+\alpha r_{t-1}^2}}\right)$ where Φ is the standard normal distribution CDF, we get $p = p(s_t, s_{t+1};\,\alpha)$ as the path probability from s_t to s_{t+1}. By converting ARCH(1) process of return to AR(1) process of squared return and assuming past returns, long term volatility are observed from the market, we effectively reduced parameters to only α and no other parameter needs to be exogenously calibrated.

A natural extension is GARCH model for the return process. The GARCH(1,1) model generalizes ARCH by allowing both past squared returns and past conditional variance to influence current volatility. This provides an additional layer of flexibility for capturing auto-correlated volatility patterns. The advantage of incorporating persistence of volatility shocks makes it a powerful tool for forecasting future volatility, especially for understanding market volatility clustering over time. Similar to the ARCH(1) model, assuming one period return follows GARCH(1,1) model such that

$$\begin{cases} r_t = \sigma_{t|t-1} \cdot \epsilon_t \\ \sigma_{t|t-1}^2 = \omega + \alpha r_{t-1}^2 + \beta\sigma_{t-1|t-2}^2, \end{cases} \tag{2.56}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$ and (α, β, ω) are parameters to be estimated. To reduce number

of parameters to be calibrated and introduce long term volatility σ, set

$$\eta_t = r_t^2 - \sigma_{t|t-1}^2, \tag{2.57}$$

we have r_t^2 follows ARMA(1,1) process

$$r_t^2 = \omega + (\alpha + \beta) r_{t-1}^2 + \eta_t - \beta \eta_{t-1}. \tag{2.58}$$

Taking expectation on both side in RN measure gives

$$\sigma^2 = \omega + (\alpha + \beta) \sigma^2, \tag{2.59}$$

which implies $\omega = (1 - \alpha - \beta) \sigma^2$, for $0 < \alpha + \beta < 1$. This is equivalent to ARMA(1,1) of r_t^2 is a week stationary process. Since $r_t \sim \mathcal{N}\left(0, \sigma_{t|t-1}^2\right)$, where

$$\sigma_{t|t-1} = (1 - \alpha - \beta) \sigma^2 + \alpha r_{t-1}^2 + \beta \sigma_{t-1|t-2}^2, \tag{2.60}$$

the return CDF is $F(x;\, \alpha, \beta) = \Phi\left(\frac{x}{\sigma_{t|t-1}}\right)$ and we can get conditional volatility sequentially from historical volatility $\sigma_{-(k-1)|-k}^2 = \sigma_{-k}^2$ of k periods in the past. Then $p = p(s_t, s_{t+1};\, \alpha, \beta)$ is the path probability from s_t to s_{t+1}. Because only one parameter can be brought into the recovery system, we need to pre-calibrate conditional volatility weight β by maximum likelihood method, leaving α to be estimated from the system.

The third strand of price process is non-parametric distribution from machine learning (ML) method or other non-parametric estimates. As an example of ML methods, we test the recovery results with one period discrete price distribution from multinomial logistic regression (MLR). MLR naturally fits the discrete recovery problem since it is designed to categorize returns into different outcomes with assigned probability based on a set of predictors. This approach is particular useful to clas-

sify various states of the market and make probabilistic predictions about the price directions. It also provides flexibility in modeling categorical outcomes and transitions between different market conditions. Consider MLR with n categories as $n-1$ independent logistic regressions, where the first category is chosen as the pivot, and all other $(n-1)$ categories are independently compared with the pivot. Thus, for starting state s_t, setting the probability of pivot path (e.g. top mover from s_t to s_t+n) as p, probabilities for all other following paths $s_t - n \leq s_{t+1} \leq s_t + n - 1$ are given by $p(s_{t+1}) = \alpha(s_{t+1}) \cdot p$, where $\alpha(s_{t+1})$ is the multiple of path probability for s_{t+1} as opposed to pivot path from each binomial logistic regression, respectively. After all ratios $\alpha(s_{t+1})$ are fixed by MLR, each path probability is only a function of unknown pivot probability p which is decided by the unit sum of all outcomes. In order to leave 1 free parameter for the recovery system, we can skip the last binomial logistic regression and set tail probability from each starting node as

$$\begin{aligned} p(s_t - n) &= 1 - \sum_{i=-(n-1)}^{n-1} p(s_t + i) \\ &= 1 - p \cdot \sum_{i=-(n-1)}^{n-1} \alpha(s_t + i), \end{aligned} \tag{2.61}$$

for $-n \leq i \leq n-1$. In this case, unit sum of probability is always guaranteed and only parameter pivot probability p is estimated by the recovery system. Another implicit input of MLR method is the training rule for each logistic regression. As an illustration, we trained the model by one-day lagged realized return and realized volatility from most recent 1-year data up to each recovery date, respectively.

The last candidate price model is non-parametric kernel estimate. Similar to MR methods, kernel estimates do not assume any specific parametric form of the underlying return, thus allowing a purely data driven result. Kernel estimators smooth data by averaging over neighboring sample points. It is like putting boxes centered at

each observed data points, and density is given by the sum of such boxes. Different kernel functions give different shapes of "boxes". Kernel estimation is flexible and can be adapted to any data distribution, capturing complex and unknown dynamics in asset prices. It is especially useful if we want the result to be fully data driven or no parametric model can fit data. One caveat of applying kernel estimates to such tree recovery framework is that the results are mainly driven by kernel bandwidth, i.e. price length at each state, not choice of kernels. But step size is decided by the strike range and tree structure, which cannot be calibrated or tuned for a given tree structure. However, we can still assume a functional form of kernel

$$K(x) = a\left(1 - x^2\right) + \left(\frac{1}{2} - \frac{2a}{3}\right), \tag{2.62}$$

for parameter $0 \leq a \leq \frac{3}{4}$, where $a = 0$ means uniform kernel and $a = \frac{3}{4}$ means Epanechnikov optimal kernel and the optimal is in estimate MSE sense. The probability of each path is then a function of parameter a. Set kernel CDF

$$\begin{aligned} F_n(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\frac{x-X_i}{h}} K(u)\, du, \end{aligned} \tag{2.63}$$

for bandwidth $h = \frac{X_{max} - X_{min}}{1+2nT}$, kernel function $K(u)$ and observation vectors $\tilde{X}_{n\times 1}$. Note we do note expect probability to change significantly for different kernel functions. The path probability from s_t to s_{t+1} estimated by kernel method is then $p = p(s_t, s_{t+1};\, a)$ where kernel function parameter a is decided by the recovery system.

2.2.7 Candidates for Investor's Utility Function

Aside from different assumptions for one-period price process, the second part that comes into each recovery equation is utility function of a typical investor at each state (t, s). In this study, I mainly test the effect of CRRA, CARA, HARA and Epstein and Zin (1989) recursive utility. For CRRA and CARA, there is only one parameter to calibrate, while there are 3 parameters for HARA. Recall the maximum number of allowed parameter in utility function is $(2n-1)$ for a tree of $(1+2n)$ edges at each node. So, we can only consider HARA when the tree structure has at least 5 possible outcomes for each starting state. For recursive utility, the investor cares not only the terminal wealth, but wealth from previous periods as well, the recovery equation for each observed state price also needs to be adapted accordingly.

For CRRA terminal wealth utility

$$u(W) = \begin{cases} \dfrac{W^{1-\gamma}}{1-\gamma}, & \gamma > 0, \gamma \neq 1, \\ \ln(W), & \gamma = 1, \end{cases} \tag{2.64}$$

and marginal utility

$$v(W) = u'(W) = W^{-\gamma}, \tag{2.65}$$

where risk aversion parameter to be estimated is $\gamma > 0$. CRRA utility implies that the investor's relative risk aversion remains constant regardless of wealth. CRRA is suitable for preference modeling if investors exhibit consistent risk-taking behavior across different wealth levels and is widely used in portfolio optimization problems. To see the constant relative risk aversion, note $u''(W) = \gamma W^{-(1+\gamma)}$ and thus $\frac{-u''(W)}{u'(W)} \cdot W = \gamma$. CRRA is also homothetic,

$$\frac{v(W_1)}{v(W_0)} = \left(\frac{W_1}{W_0}\right)^{-\gamma} = \frac{v(\alpha W_1)}{v(\alpha W_0)}. \tag{2.66}$$

It provides a shortcut for consumption based CRRA utility where consumption is proportional to terminal wealth. Homothetic property says under CRRA, the marginal utility ratios are always the same as the corresponding terminal wealth based utility case.

The second candidate is CARA utility

$$u(W) = 1 - e^{-\gamma W}, \tag{2.67}$$

where W is the terminal wealth and γ is the risk aversion parameter to be estimated. CARA utility simplifies analytical modeling, especially for normally distributed wealth as it can lead to closed-form solutions. It is widely used for modeling investor behavior when their risk tolerance does not change with wealth, making it useful in insurance and fixed-income studies. However, it is less realistic for wealthier investors who typically become less risk-averse as their wealth increases. As a compensation, we always normalize terminal wealth for CARA utility. Taking derivatives we have

$$u'(W) = \gamma e^{-\gamma W} \tag{2.68}$$

and

$$u''(W) = -\gamma^2 e^{-\gamma W}, \tag{2.69}$$

giving constant absolute risk aversion $\frac{-u''(W)}{u'(W)} = \gamma$. For consumption that is proportional to terminal wealth, CARA gives

$$\frac{v(\alpha W_1)}{v(\alpha W_0)} = \left(\frac{e^{-\gamma W_1}}{e^{-\gamma W_0}}\right)^{-\gamma} = \left(\frac{v(W_1)}{v(W_0)}\right)^{-\gamma}, \tag{2.70}$$

for $0 < \alpha \leq 1$.

HARA utility

$$u(W) = \frac{1-\gamma}{\gamma} \cdot \left(\frac{aW}{1-\gamma} + b\right)^{\gamma}, \tag{2.71}$$

for $a > 0$, $\frac{aW}{1-\gamma} + b > 0$ with parameter (a, b, γ) is a generalization of both CRRA and CARA. HARA utility is versatile because it can represent a wide range of investor preferences, including RN, CRRA and CARA as special cases. It allows modeling of changing risk aversion as wealth evolves, offering flexibility in capturing investor preferences over different wealth levels. This makes it suitable for life-cycle models and in situations where risk preferences can change over time. The absolute risk aversion is given by

$$\frac{-u''(W)}{u'(W)} = \left(\frac{W}{1-\gamma} + \frac{b}{a}\right)^{-1}, \tag{2.72}$$

and marginal utility

$$v(W) = a\left(\frac{aW}{1-\gamma} + b\right)^{\gamma-1}. \tag{2.73}$$

To see it is a generalization of the first two cases, the limiting behavior of $u(W)$ shows the investor is risk neutral when $\gamma = 1$, and has quadratic utility when $\gamma = 2$. If

$$\begin{cases} \gamma < 1 \\ a = 1 - \gamma \\ b = 0, \end{cases} \tag{2.74}$$

the HARA is degenerated to CRRA case and more specially when $\gamma \to 0$, the HARA coincides with log utility. If

$$\begin{cases} b = 1 \\ \gamma = -\infty, \end{cases} \tag{2.75}$$

the HARA is degenerated to CARA utility.

In addition to the above one period parametric utilities, I also consider the recursive utility. Epstein and Zin (1989) utility allows for a richer understanding of investor behavior by measuring risk preferences over different periods with certainty equivalent transformations. It also separates the risk preference with elasticity of intertemporal substitution (EIS) while in traditional CRRA, CARA and HARA util-

ities, they are fixed to be reciprocals to each other. This flexibility is particularly useful in long-term asset allocation and understanding how investors take actions to uncertain economic shocks. It also provides more accurate risk preference estimates in the presence of long-run risks. For initial wealth W_0 at each starting state, subjective discount factor β, EIS ρ^{-1} and risk aversion γ, the recursive utility measured at starting time is

$$\begin{aligned} u_0(W, u_T) &= \left[(1-\beta) W_0^{1-\rho} + \beta \left(\mathcal{R}(u_T)\right)^{1-\rho}\right]^{\frac{1}{1-\rho}} \\ &= \left[(1-\beta) W_0^{1-\rho} + \beta \left(\mathbf{E}\left(u_T^{1-\gamma}\right)\right)^{\frac{1-\rho}{1-\gamma}}\right]^{\frac{1}{1-\rho}}, \end{aligned} \tag{2.76}$$

for parameter $0 < \gamma < 1$, EIS $\rho^{-1} = \frac{dln(W_{t+1}/W_t)}{dln[u'(W_{t+1})/u'(W_t)]}$ and certainty equivalent $\mathcal{R}(u_T) = \left[\mathbf{E}\left(u_T^{1-\gamma}|\mathcal{F}_0\right)\right]^{\frac{1}{1-\gamma}}$. Miao (2014) shows for recursive utility with such certainty equivalent, the SDF is

$$m = \beta \cdot \left(\frac{W_T}{W_0}\right)^{-\rho} \left[\frac{u_T}{\mathcal{R}(u_T)}\right]^{\rho-\gamma}. \tag{2.77}$$

A special case is when W_T is deterministic, the last term becomes 1. Set

$$u_T(W_T) = (1-\beta)^{\frac{1}{1-\rho}} \cdot W_T, \tag{2.78}$$

such that u_0 and u_T share the same functional form, we have

$$\begin{aligned} u(W) &= u_0(W, u_T(W_T)) \\ &= \left[(1-\beta) W_0^{1-\rho} + \beta(1-\beta)\left(\mathbf{E}W_T^{1-\gamma}\right)^{\frac{1-\rho}{1-\gamma}}\right]^{\frac{1}{1-\rho}}, \end{aligned} \tag{2.79}$$

and

$$m_T^s = \beta \cdot \left(\frac{W_T}{W_0}\right)^{-\rho} \left[\frac{W_T^{\rho-\gamma}}{\left(\mathbf{E}W_T^{1-\gamma}\right)^{\frac{\rho-\gamma}{1-\gamma}}}\right], \tag{2.80}$$

where state s is implicit in terminal wealth variable $W_T(s)$. Multiply by state prob-

ability, the state price $\pi_T^s = m_T^s q_T^s$ and we can follow the same procedure to list up all state price equations and solve for the recovery results period by period. Note in this case, the subjective discount factor β is not observed from the market data and needs to be solved together with recovered (p, v).

2.3 Empirical Results

2.3.1 Data and Sample Selection

For empirical analysis of the tree recovery framework, I test the recovery results on both monthly and weekly forecast horizon. The baseline monthly forecast results use weekly SPX options data to structure a tree of up to 4 levels, where the tree level is decided by available Friday SPX expiries within the following one-month time. The bottom level then represents possible states and recovered distribution at monthly (around 4 weeks) horizon. The studied period for baseline monthly recovery covers recent 150 months from September 2008 to August 2023. Since standard SPX options, including weekly and longer expiry products, typically expire on Fridays (CBOE introduces daily expiry products in 2016), I selected last Friday of each month as the recovery date for each monthly recovery if at least one expiry falls into 24 to 35 days since the recovery date. The date range promises an approximate one-month forecast period and all months without appropriate monthly expiry are excluded from the studied period. I also exclude all recovery dates with only out-of-money or only in-the-money strikes for SPX options. Such extreme cases concentrate on dates before 2011 and almost all months in the recent 10 years are included.

I also test the recovery results for weekly forecast using daily expiry SPX options data. Since daily expiry SPX options are introduced in 2016, the weekly forecast recovery only uses data in recent 7 years until August 2023. Similar to baseline model, only Fridays with at least one expiry falling into 5 to 7 days from the recovery

date are selected to make sure the recovery is really on a weekly horizon. Each level of the tree corresponds to an available daily expiry within 7 days from the recovery date. Same for both recovery horizons, I test the tree framework with different number of edges expanding from the same node. The baseline model is set to have 5 equally spaced edges on each side expanding from the starting state. Recall for n edges on each direction, there are in total of $(1 + 2n)$ edges expanding from the same node. Thus, the baseline model has 11 possible terminal states for each of the starting nodes to guarantee a reasonable number of terminal states are available at the bottom level. Trees with 3 and 7 edges on each side are also tested. In general, tree structures with more possible terminal states are more precise for recovery distribution estimates with smaller standard errors, at the cost of increasing computation efforts for both state price generation and recovery equations solution processes.

The SPX options data are from OptionMetrics dataset. Trade level data of SPY is also used to calculate proxies for jump parameters and trade-level daily return volatility. To check robustness of tree framework, I also use VIX index level in addition to trade-level volatility to characterize market volatility. Daily level and return of both SPX, VIX indices are extracted from CRSP. SPX index level is also adjusted for dividends and stock splits. Risk free rate data is from FED website.

2.3.2 State Price Generators

Throughout the previous analysis, we have assumed all state prices at each state (t, s) are available from market data at the recovery date, i.e. date 0. However, there is no traded Arrow-Debreu securities and the state prices are not directly observable. Instead, I use two different methods, parametric and non-parametric, together with option prices data to first back out the RN state probability, then multiply the risk free rate for the RN state prices. Since state prices are invariant across different measures, we use generated state prices as input for the physical probability recovery.

The first parametric method comes from Bates (2000) stochastic volatility with jumps model. Assume RN asset price S_t follows

$$\begin{cases} \dfrac{dS_t}{S_t} = \left(r - d - \lambda\bar{k}\right) dt + \sqrt{V_t} dW_t^1 + k dN_t \\ dV_t = \left(\alpha - \beta V_t\right) dt + \sigma_v \sqrt{V_t} dW_t^2 \\ Cov\left(dW_t^1,\, dW_t^2\right) = \rho dt, \end{cases} \tag{2.81}$$

for risk free rate r and dividend yield d. Further assume the instantaneous return variance V_t is mean reverted with level α, speed β and a second Brownian motion source of uncertainty W^2, which is correlated with source of uncertainty W^1 for underlying return process. N_t is a Poisson counting process representing the jump component. The jumps are assumed to have intensity λ and multiplicative size $(1+k)$ with lognormal distribution

$$\ln\left(1+k\right) \sim \mathcal{N}\left(\frac{1}{2}\ln\left(1+\bar{k}\right) \cdot \delta^2,\, \delta^2\right). \tag{2.82}$$

The model parameters $\left(\alpha, \beta, \lambda, \delta, \bar{k}, \sigma_v, \rho\right)$ are dynamically calibrated with most recent 50 around-the-money call options at each recovery date. Once calibrated, we numerically calculate the characteristic function $\Psi\left(u\right)$ for digital option prices from Bates (2000) and use inverse Fourier transform

$$f\left(S\right) = \frac{1}{\pi} \int_0^\infty \left(\frac{S}{S_0}\right)^{-iu} \Psi\left(u\right) du \tag{2.83}$$

to back out RN density of asset price. Finally, the state price is revealed by adding the RN discount factor of corresponding recovery period. See Appendix A for more details.

The second way to generate state price is Jackwerth (2004) non-parametric fast-and-stable method. The method fits missing values of implied volatility σ_s from

observed implied volatility σ_s^{obs} as

$$\min_{\sigma_s}\left[\frac{1}{2\left(n+1\right)}\sum_{s=1}^{n}\left(\sigma_s''\right)^2+\frac{\lambda}{2n_{obs}}\sum_{s=1}^{n_{obs}}\left(\sigma_s-\sigma_s^{obs}\right)^2\right], \tag{2.84}$$

with tuning parameter λ, where σ_s'' is the second order derivatives of fitted volatility with respect to strike, n and n_{obs} are the number of states to fit and number of volatility observed from options data, respectively. The first term controls smoothness of fitting and the second term measures goodness of fitting by averaged sum of squared error.

The tuning parameter λ balances the trade off between two terms. High values of λ minimize fitting error but can result in jagged fitted volatility surface. For example, if the first smoothness term is ignored, the best fitting result is to fill in all missing volatilities with 0 and keep $\sigma_s = \sigma_s^{obs}$ for all state s with observed volatility, i.e. not fitting at all. Following Jackwerth (2004), we choose the highest λ that satisfies both conditions

(i). the recovered RN density is always non-negative at each state,

(ii). the derived Breeden-Litzenberger RN distribution is unimodal for strike moneyness within range $(0.8, 1.2)$.

The second condition decreases optimal λ to avoid jagged state prices data. See Appendix A for more details.

After getting fitted volatility surface, we first calculate Black-Scholes fair value of call option prices with fitted volatility, then use Breeden and Litzenberger (1978) formula to get the RN state price density, and finally state prices with corresponding discount factor.

Table 2.1 shows time series summary statistics of RN expected return recovered from both state price generators. As shown in the theoretical part (2.24), under terminal wealth utility and RN preference, state probability is only a re-weight of state

	Mean %	TS Vol %	TS Skew	TS Ex Kurt	TS Std Err %
Bates RN Return	1.53	16.38	-2.57	10.66	1.44
Jack RN Return	1.22	14.29	-3.61	18.21	1.26

Table 2.1: Comparison of RN Annualized Monthly Expected Returns from Two State Price Generators

price. Thus, we can check summary statistics of generated state prices by checking the corresponding RN return distribution. Under both generators, state prices imply similar expected annualized returns that are close to risk free rate. However, Jackwerth return distribution is much more left skewed and fat tailed than returns from Bates recovery. The exaggerated probabilities on the left tail is likely resulted from excess demand pressure from deep out-of-the-money options. Bates (2000) only uses around the money call options for calibration while Jackwerth (2004) uses both out-of-the-money calls and puts, thus are more susceptible to volatility skews, especially from the left end. As a consequence, I use Bates state prices as baseline specification in the following empirical results. Return distributions from Jackwerth state prices are also reported separately.

2.3.3 Parameters Pre-calibration

Since the recovery framework only allows for one parameter from pricing process to be solved by the recovery equations system together with parameters from utility process, we need to pre-calibrate all other parameters emerging from the assumed price process. One major downside of this recovery framework is that the final recovery result is sensitive to parameters decided outside the recovery process. Despite of this, this pre-calibration is inevitable due to the nature of excess number of unknowns than equations we have from limited state price information. To compensate for this exogenous factor, I always calibrate all model parameters with most up-to-date data. I also test the model with different calibrations or data sources for the same param-

eter to check the robustness of our recovery results. Let us break down into each parameter needed for recovery.

For GBM diffusion model, where return drift is taken into recovery system, we need to pre calibrate return diffusion parameter σ for each state (t, s), and each state s is associated with a terminal price range of the underlying asset. Since each state corresponds to at least one market tradable option contingent on the underlying asset that expires at time t with strike as the midpoint of the state price range, I calibrate the price diffusion parameter starting from state (t, s) by the implied volatility of corresponding options. To adjust for inflated implied volatility from demand pressure of extreme strikes, I cap the diffusion parameter σ by 60%, which is supported by SPX volatility observed at extreme periods. One advantage of applying implied volatility of each strike as the diffusion parameter is that the options implied volatility is by definition the underlying asset return volatility standing at the corresponding state in Black-Scholes world. In reality, we cannot observe state parameters of all states at the same period, but the implied volatility offers a peek for each imaginary states. Another advantage of using implied volatility comes from volatility skew. Volatility skew describe the excess implied volatility observed at two extreme strikes, compared with relative low and stable implied volatility for around the money strikes. The excess uncertainty is more pronounced on the extreme low return side. Using implied volatility with skew as the diffusion parameter exactly captures investor's anticipation of larger volatility in the states of extreme realized returns. Empirical data also supports surged volatility in reaction to surprising news or unexpected price moves.

For jump diffusion models where continuous drift parameter is calibrated by the recovery model, we need to settle down continuous diffusion parameter σ, jump intensity λ and multiplicative average jump size $(1 + \beta)$. Specifically, for double exponential jump model, we need to specify the upward jump probability and jump size

for upward, downward jumps respectively. Similar to the continuous model, continuous diffusion parameters are calibrated by capped implied volatility of corresponding options. One way to calibrate jump parameters is from bulk trade data point of view. Suppose price jumps are mainly from large institutional trades that result in sudden liquidity shocks, using institutional trade data from TAQ dataset can work as proxies for jump intensity and size. Specifically, at each recovery date, we get jump rate data as the fraction of institution trade number over total trade number of previous trading day. Upward jump size is calculated as the fraction of total institutional buy order dollar amount over daily buy order amount, and similarly for the downward jump size from sell data. Probability of upward jump is proxied by fraction of institutional buy order numbers over total institutional trade numbers of the previous day. Finally, the aggregated jump size in Poisson jump model is given by the weighted average of upward and downward jump sizes. The estimated jump rate parameter ranges from 0.2 to 1 and the aggregated jump size is averaged around -5% to -10% per year, which coincide with previous empirical studies. For double exponential jump model, once multiplicative jump sizes are calibrated, the exponential jump parameters η_1 and η_2 are simply the reciprocal of average jump sizes for each jump direction.

ARCH(1) model is only controlled by weight parameter on squared past return α and there is no exogenous parameter needs to be calibrated. However, we still need to estimate long term return volatility σ and one day lagged squared return. The long term volatility is estimated by annualized return volatility from past 252 trading days, and we always use the most recent 252 trading days up to the recovery date for each recovery. Different from long term volatility, the one day lagged return is state specific for each unrealized starting state at the recovery date. For recovery results from current recovery date to the first period, the past return is simply observed one-day lagged return from market data. Starting from the first period on the tree, for

each starting state (t, s), denote the set of all paths from period $(t-1)$ to t that ends in state s as $\Omega_{(t,s)}$. Since the recovery equation system is solved period by period, we have already know all path probabilities starting from $(t-1)$ by time t. Then, the estimated one-day lagged return is given as the weighed average return of all possible starting states ending in s as

$$r_{t-1}^{s} = \sum_{s_{t-1} \in \Omega_{(t,s)}} \frac{X_s - X_{s_{t-1}}}{X_{s_{t-1}}} \cdot p_{s_{t-1}}, \tag{2.85}$$

where $p_{s_{t-1}}$ is path probability connecting state (s_{t-1}, s) and $X_{s_{t-1}}$ is the index level at starting state s_{t-1}.

ARMA(1,1) model has two parameters: weight on squared return α and weight on past conditional volatility β for the conditional volatility process. The standard method is to first assume linear relations between two parameters

$$\beta = c_0 + c_1 \cdot \alpha, \tag{2.86}$$

for helper parameters c_0 and c_1. Then use maximum likelihood estimate to get both c_0 and c_1 for a linear representation of β, leaving only α to be estimated from the recovery system. In addition to long term volatility and one-period lagged squared return that are estimated in the same way as ARCH(1) model, we also need to estimate one-period lagged conditional return volatility $\sigma_{t-1|t-2}$ from a given period before the recovery date. See Appendix A for more details on parameters tuning and pre calibrations.

Fortunately, for non-parametric price processes including multinomial logistic regressions and kernel estimates, the tested price processes are designed to contain only one undecided parameter, thus no pre calibrations are needed for non-parametric methods. For MLR method, I test two different training rules: one-period lagged return and one-period lagged volatility, denoted as MLR and MLR2, respectively.

For each training rule, I train the model by most recent 252 trading days data at each recovery date. For MLR2, the daily return volatility data are either estimated with trade-level intraday volatility data or closing VIX level of corresponding trading date. As we can see, recovery results with different data sources do not change the recovered results much, which is a positive sign for the result robustness.

For parameters in utility functions, although they are all supposed to be estimated from the recovery system, we still need to estimate the wealth level at each possible state in the tree. To estimate the total wealth of an typical investor, we assume the optimal investment strategy is classical 40% equity index (recovered risky asset) plus 60% risk free bond, where share quantity of risky asset is estimated from SPY traded volume and index level is given by the state itself. For CARA utility, all wealth are rescaled to avoid its sensitivity to the wealth level. Because the marginal utility function only appears in ratio forms in recovery system, the exact absolute level of wealth and utlity does not affect results, only the marginal utility ratio at different states plays a role in the recovery framework.

2.3.4 Baseline Results

I empirically test the recovery framework to a variety of price and utility processes, tree structures, forecast horizons and inputs calibrated from different data sources. The aim is to test the robustness of price density recovery result under different settings. In the baseline model, each starting state has 11 possible ending states in unit period time, i.e. in the tree structure, each one-period subtree has $n = 5$ edges on each price change direction, plus the middle edge for no price change. Forecast horizon is set to one-month from the recovery date for the entire tree and tree levels are decided by number of available weekly expiries in the following month.

Table 2.2 shows time-series (TS) statistics of annualized recovered expected return across 150 months from September 2008 to August 2023, excluding months without

Bates	Mean %	TS Vol %	TS Skew	TS Ex Kurt	TS Std Err %
Jump HARA	8.20	13.80	-1.50	4.31	1.25
Jump CRRA	8.30	14.63	-1.74	4.87	1.33
Jump CARA	7.90	14.10	-1.54	4.55	1.28
Kou HARA	8.23	13.87	-1.57	4.58	1.26
Kou CRRA	9.29	14.03	-1.70	5.67	1.28
Kou CARA	7.38	14.18	-1.57	4.80	1.29
GBM HARA	7.76	12.31	-1.25	3.24	1.11
GBM CRRA	8.99	13.11	-1.39	3.52	1.19
GBM CARA	7.20	12.39	-0.91	2.15	1.13
MLR HARA	8.31	14.22	-0.61	1.41	1.30
MLR CRRA	8.99	15.24	-0.73	1.44	1.39
MLR CARA	6.73	15.00	-0.49	0.44	1.36
MLR2 HARA	7.73	15.00	-0.58	0.84	1.37
MLR2 CRRA	9.27	14.09	-0.85	1.16	1.31
MLR2 CARA	5.55	15.35	-0.88	2.32	1.39
ARCH1 HARA	7.92	12.72	-0.95	4.13	1.17
ARCH1 CRRA	8.49	12.55	-1.03	4.39	1.14
ARCH1 CARA	8.01	12.39	-0.73	3.88	1.12
GARCH1 HARA	8.68	12.21	-0.90	4.51	1.11
GARCH1 CRRA	8.51	13.38	-1.22	4.96	1.22
GARCH1 CARA	7.99	12.12	-0.84	3.98	1.10

Table 2.2: Summary Statistics of Monthly Expected Return from Recovery Results

sufficient SPX options data. The state prices are all generated with Bates (2000) stochastic volatility jump diffusion model. The seven price processes can be categorized into three types. Diffusion processes include continuous GBM model and two jump processes with Poisson, double exponential jump assumption, respectively. All diffusion models contain a continuous part of lognormal price trajectory with an exogenous diffusion parameter decided by the options implied volatility. Time series models include two candidates, ARCH(1) and GARCH(1,1). Both models also imply lognormal price process, but the diffusion parameter is decided either by a linear combination of long-term volatility and past squared return for ARCH(1) or the linear combination of above two components plus past conditional return volatility for GARCH process. Machine learning based non-parametric models include both MLR and MLR2, trained either by realized return or intraday volatility.

Across all specified price processes and utility assumptions, the annualized recov-

ered expected return are all around 6% to 8%, with annualized return volatility at around 14%. This consistency shows our tree recovery framework is robust across different underlying models with recovered SPX Sharpe ratios center around 0.5. This is in align with empirical findings of 0.5 to 0.7 in recent 15 years. Averaged recovered returns do not differ much within different diffusion models and time series models, but display more variety for different multinomial logistic regression (MLR) specifications. MLR2 expected returns, which are trained by past return volatility, ranges from 5.5% to 9.2% under different utilities. While on the other hand, MLR methods trained by past realized returns has more stable recovered expected returns across different utility specifications. This shows the choice of training rule is a deciding factor in recovered results. A more consistent rule like past return in MLR methods can generate more consistent recovered results. However, let us keep in mind that a vast number of possible factors can be applied to this machine learning price process framework for better recovery results. The factors with better predictive power and higher data quality are more promising in generating better recovery results. Here I only use past realized return and past intraday volatility as two illustration examples for the machine learning price process applied to the tree recovery framework.

Move on to annualized recovered TS volatility. For diffusion models, continuous GBM models yield lower return volatility compared with jump models. The annualized volatility increased from around 12.5% for GBM models to 14% for jump models, reflecting excess source of uncertainty from jump components. The 12.5% volatility from continuous model seems to underestimate SPX uncertainty compared with 15% from empirical data. The jump models compensate on this shortage and is thus more recommended when investors need to choose from diffusion models. TS models, including ARCH and GARCH, are also more prone to generate smoother expected returns with averaged volatility centered around 12% to 13%, compared with diffu-

sion models. The difference sheds light on how different data source can affect the recovery results. Diffusion models pre-calibrate the diffusion parameters directly with implied volatility from corresponding states, thus providing more accurate information of each possible future states. However, for ARCH(1) model where conditional return volatility at each state is decided by long term volatility and past realized return, k-th edge counting from top for each starting state (t, s) shares exactly the same path probability since both long-term volatility and past period realized return are the same. This fixed conditional volatility structure speeds up the optimization process for TS models but offer less variation across different states. Similar reasoning also applies to GARCH model. In summary, although reflecting less information from data inputs compared with diffusion models, TS models offer a solution for quick approximation to the recovery results.

All recovered expected returns are negatively skewed and fat tailed over tested period. To prevent from over fitting, the possible ending price range for each state is set to be symmetric around price associated with starting state. The price ranges are extracted from all market tradable option strikes and almost cover all possible price changes. As a result, this negative skewness really reflects heavier probability on the negative return side, and recovered state probability for extreme large return state is often recovered as zero. Market observed SPX return skewness usually lies between -1.5 to -0.5, which coincides with our recovery results. Our recovered return distributions also properly account for fat tail property observed for SPX market data. The excess kurtosis falls between 0.5 to 6 for all models, echoing previous studies that SPX return excess kurtosis is usually from 2 to 5. Both negative skewness and fat tail are more pronounced for diffusion and TS models. MLR based machine learning results put slightly less weight on tail probabilities. All standard errors of the recovered expected return under different model specifications are under 1.5%

Bates	Mean %	Undl Vol %	Undl Skew	Undl Ex Kurt	Undl Std Err %
Jump HARA	8.20	34.01	0.02	-0.28	1.64
Jump CRRA	8.30	34.36	0.02	-0.30	1.67
Jump CARA	7.90	34.24	0.02	-0.27	1.65
Kou HARA	8.23	34.19	0.02	-0.28	1.65
Kou CRRA	9.29	34.06	0.02	-0.28	1.64
Kou CARA	7.38	34.03	0.03	-0.27	1.64
GBM HARA	7.76	21.61	-0.28	0.89	1.08
GBM CRRA	8.99	21.82	-0.29	0.85	1.10
GBM CARA	7.20	21.96	-0.25	0.84	1.09
MLR HARA	8.31	28.69	-0.09	0.30	1.40
MLR CRRA	8.99	28.50	-0.08	0.26	1.39
MLR CARA	6.73	28.99	-0.10	0.26	1.42
MLR2 HARA	7.73	28.66	-0.08	0.30	1.40
MLR2 CRRA	9.27	28.56	-0.09	0.30	1.40
MLR2 CARA	5.55	28.89	-0.08	0.24	1.41
ARCH1 HARA	7.92	19.85	0.29	2.39	1.00
ARCH1 CRRA	8.49	19.81	0.29	2.46	1.00
ARCH1 CARA	8.01	19.85	0.29	2.45	1.00
GARCH1 HARA	8.68	21.67	0.05	0.47	1.08
GARCH1 CRRA	8.51	21.22	0.05	0.43	1.04
GARCH1 CARA	7.99	21.69	0.04	0.47	1.08

Table 2.3: Summary Statistics of Underlying Return Distribution from Recovery Results

across 150 sample months.

Table 2.3 shows averaged statistics of recovered underlying index return distribution. All values are reported as the average statistics under each model specifications over 150 tested recovery dates. Different from the TS distribution statistics in Table 2.2, the underlying statistics represents monthly return distribution across all possible states where observed return is one realization of it. This underlying distribution incorporates possibilities of all unobserved future outcomes and is thus never observed from real data.

The recovered underlying SPX return distributions for all model specifications have much larger volatility than their TS distributions of expected returns. For continuous GBM model and two time series models, the annualized volatility of underlying price is around 20%, while for jump models with additional uncertainty from market shocks, the underlying volatility goes up to around 35%. Although the

true price density is never observed and we cannot compare the recovered underlying volatility directly with such data, one interesting observation is SPX return volatility estimated from intraday trade level data is also around 35% to 40%. The excess uncertainty compared with statistics observed from TS data reflects micro-level noise from each trade and is more accurate as a proxy of true price process. For machine learning based recovery results, different training rules do not change the recovered volatility. Recovered underlying distributions are on average more dispersed than those from continuous models, but still do not match the volatility level estimated from jump price processes. For all recovered underlying distributions, unlike TS distributions, there is on average almost no skewness and excess kurtosis. Indeed, negative skewness and fat tails in return distributions are often attributed to rare but extreme market events, such as financial crises or large economic shocks. These rare disasters cause significant downward movements in prices, leading to a higher likelihood of observing large negative returns over time. When analyzing a time series of returns, the accumulated effects of rare disasters lead to asymmetric and heavy tailed TS return distributions. However, for an individual, static observation of the underlying distribution at any given point, the likelihood of such extreme events is very lower, and the distribution may appear more symmetric. Therefore, the negative skewness and fat tails are more likely observed in time series data over long periods, where rare disasters have a cumulative impact on the return distribution.

2.3.5 Recovery Results with Different Specifications

Table 2.4 and Table 2.5 show monthly recovery results with different state price generators and different input data sources for past return volatility. When state prices are generated from Jackwerth (2004) non-parametric method by fitting missing implied volatility, the TS statistics in general have similar patterns for different parametric models as Bates (2000) based recovery, only less consistent across different models.

The recovered expected returns over tested period center around 7.5% and annualized volatility are mostly from 14% to 17%, slightly more elevated than the Bates results. For diffusion models, although jump models still exhibits larger volatility than continuous GBM based recovered volatility, the difference is less pronounced compared with Bates (2000) recovery. Recovered TS excess kurtosis based on non-parametric state prices are also less consistent compared with Bates (2000) results. The difference is mainly resulted from options data used in both methods. Bates state price only uses call prices to calibrate Heston (1993) model with jumps for underlying asset return, while Jackwerth state prices are calibrated from both calls and puts. Thus, the volatility skew from both OTM calls and puts are more likely to pull more probability weights on state prices at two ends, especially for OTM puts. Indeed, the Jackwerth state prices are more likely to be larger than Bates state prices for extreme states at both ends. The exaggerated extreme case state prices then causes larger than expected state probabilities recovered for those states, leading to slightly elevated volatilities in recovered return distributions. Thus, in the above baseline results, we choose Bates parametric state prices as benchmark for our empirical tests.

I also test effect of different data sources for past volatility on the recovery results. The recovery results with MLR2 and GARCH price processes are trained with past return volatility data. The baseline models use daily SPY return volatility data estimated from trade level intraday data, which reflects uncertainties from both market demand and micro-level effects. In order to test the robustness of recovery model, I substitute all past volatility data with closing VIX level from previous trading day. All other specifications in recovery process are otherwise the same. The recovery results with VIX inputs are qualitative the same as results with baseline model, indicating the recovery framework is stable across different data sources.

	Mean Bates %	Mean Jack %	Mean VIX %	TS Vol Bates %	TS Vol Jack %	TS Vol VIX %
Jump HARA	8.20	8.01		13.80	15.48	
Jump CRRA	8.30	7.77		14.63	16.97	
Jump CARA	7.90	7.93		14.10	14.95	
Kou HARA	8.23	9.57		13.87	15.40	
Kou CRRA	9.29	8.82		14.03	17.38	
Kou CARA	7.38	7.49		14.18	15.75	
GBM HARA	7.76	7.30		12.31	13.18	
GBM CRRA	8.99	10.00		13.11	16.29	
GBM CARA	7.20	7.03		12.39	12.49	
MLR HARA	8.31	7.79		14.22	14.87	
MLR CRRA	8.99	8.51		15.24	16.73	
MLR CARA	6.73	7.04		15.00	16.43	
MLR2 HARA	7.73	7.22	7.75	15.00	16.19	14.94
MLR2 CRRA	9.27	8.77	8.55	14.09	15.13	14.71
MLR2 CARA	5.55	5.24	5.34	15.35	15.93	14.60
ARCH1 HARA	7.92	7.47		12.72	14.57	
ARCH1 CRRA	8.49	8.71		12.55	14.12	
ARCH1 CARA	8.01	7.66		12.39	13.34	
GARCH1 HARA	8.68	10.04	8.51	12.21	13.46	12.62
GARCH1 CRRA	8.51	9.03	8.84	13.38	14.44	13.46
GARCH1 CARA	7.99	7.84	8.02	12.12	13.25	12.22

Table 2.4: Summary Statistics of Monthly Expected Return under Different Specifications

	TS Ex Kurt Bates	TS Ex Kurt Jack	TS Ex Kurt VIX	Std Err Bates %	Std Err Jack %	Std Err VIX %
Jump HARA	4.31	6.40		1.25	1.40	
Jump CRRA	4.87	3.50		1.33	1.54	
Jump CARA	4.55	2.16		1.28	1.36	
Kou HARA	4.58	5.06		1.26	1.40	
Kou CRRA	5.67	4.40		1.28	1.59	
Kou CARA	4.80	4.24		1.29	1.43	
GBM HARA	3.24	3.53		1.11	1.19	
GBM CRRA	3.52	4.47		1.19	1.47	
GBM CARA	2.15	1.01		1.13	1.14	
MLR HARA	1.41	1.53		1.30	1.36	
MLR CRRA	1.44	0.38		1.39	1.53	
MLR CARA	0.44	0.34		1.36	1.49	
MLR2 HARA	0.84	0.05	1.95	1.37	1.48	1.35
MLR2 CRRA	1.16	0.61	1.82	1.31	1.41	1.33
MLR2 CARA	2.32	2.60	0.71	1.39	1.44	1.32
ARCH1 HARA	4.13	3.34		1.17	1.34	
ARCH1 CRRA	4.39	4.12		1.14	1.28	
ARCH1 CARA	3.88	3.13		1.12	1.21	
GARCH1 HARA	4.51	3.09	4.47	1.11	1.22	1.15
GARCH1 CRRA	4.96	3.66	5.48	1.22	1.31	1.22
GARCH1 CARA	3.98	2.62	4.22	1.10	1.20	1.11

Table 2.5: Summary Statistics of Monthly Expected Return under Different Specifications, continued

2.3.6 Recovery Results with Different Tree Structures

In this subsection, let us check the effect of different tree structures to the recovery results. Set number of edges from each starting state as $(1 + 2n)$ and tree level as T. Note the possible price range at each period is set by the available strikes of corresponding maturity, tree level is set by available weekly expiries within the recovery horizon, we get the bottom level has $(1 + 2nT)$ terminal states. Thus, n represents the precision of state partition on possible price range. Theoretically, larger n means finer partition and more accurate recovery results, however, since the recovery model is solved numerically, it also means larger computation effort. So we need to find a level of tree precision to balance recovery accuracy and computation efforts.

Table 2.6 and Table 2.7 show TS recovered expected return statistics over past 15 years under different n levels. All other specifications are otherwise the same as baseline model, i.e. Bates state price, monthly forecast horizon and intraday volatility inputs, for each tested underlying models. The baseline recovery uses $n = 5$ precision. By reducing number of possible outcomes to 3 on each direction from starting states, the recovered TS annualized volatility are slightly deflated, especially for models based on MLR and time-series price processes. The difference is likely resulted from the granularity of the partitioning. When $n = 3$, which uses larger partition length, the estimated distributions are smoother as the coarser bins aggregate more data within each sub-range. This tends to reduce volatility since rapid fluctuations between adjacent bins are not differentiated and averaged out. For the same reason, this smoothing can also understate extreme events or sharp changes in the distribution, leading to lower estimates of negative skewness and excess kurtosis. We can verify this argument in Table 2.7 that both skewness and ex kurtosis are closer to zero across different model specifications when $n = 3$. This is a sign that $n = 3$ is too small for

Bates	Mean 3	Mean 5	Mean 7	Vol 3	Vol 5	Vol 7
Jump HARA	7.98	8.20	8.15	14.07	13.80	14.21
Jump CRRA	8.16	8.30	8.06	14.13	14.63	14.25
Jump CARA	7.79	7.90	7.59	14.54	14.10	14.26
Kou HARA	7.92	8.23	8.52	13.77	13.87	14.36
Kou CRRA	8.21	9.29	8.97	14.38	14.03	14.24
Kou CARA	7.88	7.38	7.46	13.97	14.18	14.24
GBM HARA	6.65	7.76	7.50	12.60	12.31	15.67
GBM CRRA	6.90	8.99	8.69	13.75	13.11	15.67
GBM CARA	6.25	7.20	6.99	12.12	12.39	15.55
MLR HARA	8.64	8.31	8.04	13.86	14.22	16.42
MLR CRRA	8.59	8.99	8.63	14.94	15.24	16.47
MLR CARA	8.25	6.73	6.88	14.67	15.00	16.44
MLR2 HARA	7.29	7.73	7.75	14.91	15.00	16.33
MLR2 CRRA	7.28	9.27	8.80	14.32	14.09	16.37
MLR2 CARA	6.51	5.55	5.25	15.02	15.35	16.37
ARCH1 HARA	8.49	7.92	8.01	11.92	12.72	15.17
ARCH1 CRRA	7.02	8.49	8.31	12.60	12.55	15.22
ARCH1 CARA	6.71	8.01	8.21	11.66	12.39	15.20
GARCH1 HARA	9.12	8.68	8.57	12.68	12.21	14.70
GARCH1 CRRA	9.24	8.51	8.76	13.10	13.38	14.50
GARCH1 CARA	8.82	7.99	8.39	11.88	12.12	14.51

Table 2.6: Summary Statistics of Monthly Expected Return under Different Tree Edges

an accurate recovery results compared with the baseline $n = 5$ case.

On the other hand, for tree structures with more precise partitioning (smaller bins), they capture more local variations in price distributions, leading to higher volatility observed over the tested horizon. This increased granularity makes the estimate more sensitive to localized spikes or drops generated from the state price data. Consequently, the distributions are likely to show more pronounced peaks and fatter tails, resulting in higher excess kurtosis as small partitions are better at capturing extreme values and tail behavior. The tree structure with most precise partition $n = 7$ estimates TS recovered volatility at around 15% for different price and utility models. The estimated skewness and excess kurtosis are also more pronounced than results with coarser state partitions. However, we can see from Table 2.6 and Table 2.7 that recovered results do not differ much between $n = 5$ and $n = 7$ specifications. Theoretically, finer partitions are always preferred for higher recovery accuracy, but

Bates	Skew 3	Skew 5	Skew 7	Ex Kurt 3	Ex Kurt 5	Ex Kurt 7
Jump HARA	-0.55	-1.50	-2.03	0.94	4.31	4.25
Jump CRRA	-0.52	-1.74	-1.63	0.89	4.87	4.80
Jump CARA	-0.55	-1.54	-1.26	0.95	4.55	4.22
Kou HARA	-0.54	-1.57	-1.60	0.92	4.58	4.78
Kou CRRA	-0.52	-1.70	-2.08	0.88	5.67	6.25
Kou CARA	-0.54	-1.57	-1.34	0.93	4.80	5.00
GBM HARA	-0.94	-1.25	-1.69	1.91	3.24	3.03
GBM CRRA	-0.94	-1.39	-1.67	1.93	3.52	3.75
GBM CARA	-0.98	-0.91	-0.51	2.03	2.15	1.88
MLR HARA	-1.09	-0.61	-0.78	2.50	1.41	1.82
MLR CRRA	-1.20	-0.73	-0.75	2.89	1.44	1.65
MLR CARA	-1.15	-0.49	-0.27	2.70	0.44	0.10
MLR2 HARA	-1.13	-0.58	-1.05	2.54	0.84	0.70
MLR2 CRRA	-1.17	-0.85	-0.78	2.65	1.16	1.16
MLR2 CARA	-1.14	-0.88	-0.91	2.53	2.32	2.32
ARCH1 HARA	-0.27	-0.95	-1.03	0.65	4.13	4.41
ARCH1 CRRA	-0.24	-1.03	-1.25	0.86	4.39	4.25
ARCH1 CARA	-0.29	-0.73	-0.50	0.73	3.88	3.71
GARCH1 HARA	-0.72	-0.90	-0.53	1.94	4.51	4.85
GARCH1 CRRA	-0.66	-1.22	-1.71	1.89	4.96	5.14
GARCH1 CARA	-0.78	-0.84	-0.56	0.31	3.98	3.85

Table 2.7: Summary Statistics of Monthly Expected Return under Different Tree Edges, continued

also lead to longer computation time when optimizing the recovery system results numerically. The minor improve from $n = 5$ to $n = 7$ indicates the baseline models with 11 possible states for each starting state is an optimal choice for such recovery framework.

2.3.7 Recovery Results with Different Forecast Horizons

In addition to applying the recovery framework to monthly SPX return forecast, I also test recovery results with different forecast horizons. Table 2.8 shows TS statistics of annualized recovered expected return across 214 weeks from September 2016 to August 2023, excluding weeks without sufficient SPX options data. The weekly forecast applies the tree recovery system on daily SPX expiries within one week from the recovery date. Since CBOE introduced daily expiry options on SPX in 2016, our tested horizon only covers recent 7 years until September 2023. The state prices are

Bates	Mean %	TS Vol %	TS Skew	TS Ex Kurt	TS Std Err %
Jump HARA	6.11	14.11	-2.15	6.76	2.06
Jump CRRA	6.26	13.86	-2.16	6.96	2.02
Jump CARA	6.67	13.91	-2.15	6.80	2.03
Kou HARA	4.47	14.21	-2.05	6.11	2.07
Kou CRRA	5.22	14.12	-2.04	6.07	2.06
Kou CARA	5.21	14.21	-2.04	6.09	2.07
GBM HARA	7.70	18.38	-1.03	2.15	2.71
GBM CRRA	6.27	19.18	-1.08	2.09	2.80
GBM CARA	6.10	19.30	-1.08	1.85	2.82
MLR HARA	6.93	13.55	-2.47	8.64	2.00
MLR CRRA	5.60	13.73	-2.20	7.54	2.05
MLR CARA	7.52	13.75	-2.30	7.86	2.01
MLR2 HARA	7.40	12.99	-2.47	9.28	1.89
MLR2 CRRA	6.81	13.01	-2.44	9.21	1.90
MLR2 CARA	8.54	12.79	-2.58	10.12	1.91
ARCH1 HARA	6.27	13.63	-2.36	8.18	2.05
ARCH1 CRRA	6.78	13.28	-2.34	8.31	1.94
ARCH1 CARA	7.84	13.64	-2.55	9.29	1.99
GARCH1 HARA	8.52	13.37	-2.41	8.82	1.95
GARCH1 CRRA	7.67	13.48	-2.34	8.48	1.97
GARCH1 CARA	8.31	13.27	-2.28	7.96	1.94

Table 2.8: Summary Statistics of Weekly Expected Return from Recovery Results

generated from Bates (2000) parametric method.

Similar to monthly recovery, the annualized recovered expected returns are mostly around 5% to 8%, with annualized return volatility at around 14%. This result is less consistent with annualized expected return from monthly recovery, indicating shorter recovery horizons and data from daily expiries are less robust for the recovery framework and more sensitive to parameter tuning. Without higher quality data or more robust pre calibrations for model parameters, it is best to apply such recovery framework with monthly or longer horizons. Diffusion models on average delivers smaller expected return over the recent 7-year period compared with MLR and TS price processes based results. Similar to MLR based models in monthly recovery, different training rules significantly changes recovery results as recovered expected returns differ from MLR to MLR2.

Although same as monthly results, recovered expected returns are negatively

skewed and fat tailed over tested period, the level of skewness and peakness are much more pronounced for shorter forecast horizons. The discrepancy between the two sets of SPX annualized return distributions likely arises from the higher frequency of data in the weekly return estimates, which capture more granular, short-term fluctuations in the SPX performance. Weekly returns derived from daily SPX options expiries are more sensitive to sudden market movements, volatility clusters, and short-term shocks, which often result in negative skewness and fat tails. These short-term dynamics may include market corrections or negative sentiment shifts, leading to more pronounced downside risk and extreme price movements. In contrast, monthly returns with fewer data points tend to smooth out these short-term fluctuations, providing a less volatile and more symmetric view of the distribution, thus understating tail risk and skewness. Also, standard error of expected return estimates are around 2% for weekly recovery results which is much higher than 1.2% from recovery monthly horizon. In general, the estimate results with standard error less than 1.5% are considered robust. It is another sign that applying such recovery framework to monthly forecast, as in the baseline models, can generate more robust recovery results.

For averaged statistics of recovered underlying index return distribution on weekly horizon, we can find the underlying annualized volatilities for all model specifications are much larger than the underlying volatility estimated from monthly horizon. The volatilities ranges from 30% to 47%, which are much larger than the TS return volatilities, especially for the price models with jump component.

In summary, the empirical study demonstrates that the recovery framework consistently estimates annualized expected returns between 6% and 8% with volatilities around 14% across different price processes. Diffusion models with jumps show higher volatility and more pronounced fat tails, while time series models like ARCH and GARCH produce smoother returns. Recovery frameworks with time-series price pro-

Bates	Mean %	Undl Vol %	Undl Skew	Undl Ex Kurt	Undl Std Err %
Jump HARA	6.11	46.14	0.10	0.18	1.43
Jump CRRA	6.26	46.26	0.12	0.19	1.43
Jump CARA	6.67	46.24	0.12	0.18	1.43
Kou HARA	4.47	46.05	0.09	0.15	1.43
Kou CRRA	5.22	45.71	0.12	0.26	1.41
Kou CARA	5.21	45.97	0.12	0.19	1.42
GBM HARA	7.70	33.03	-0.10	0.83	1.09
GBM CRRA	6.27	33.53	-0.12	0.80	1.12
GBM CARA	6.10	32.54	-0.09	0.94	1.07
MLR HARA	6.93	34.14	0.11	0.70	1.12
MLR CRRA	5.60	32.46	0.06	0.90	1.05
MLR CARA	7.52	33.76	0.10	0.81	1.11
MLR2 HARA	7.40	33.72	0.09	0.77	1.12
MLR2 CRRA	6.81	33.94	0.08	0.74	1.12
MLR2 CARA	8.54	34.61	0.10	0.70	1.15
ARCH1 HARA	6.27	33.39	0.57	3.07	1.11
ARCH1 CRRA	6.78	33.38	0.51	3.03	1.10
ARCH1 CARA	7.84	33.57	0.50	2.89	1.11
GARCH1 HARA	8.52	34.09	0.19	1.04	1.15
GARCH1 CRRA	7.67	32.21	0.19	1.33	1.08
GARCH1 CARA	8.31	31.80	0.22	1.40	1.05

Table 2.9: Summary Statistics of Underlying Return Distribution from Weekly Recovery Results

cesses usually contain less input information, but can serve as quick approximations when recovering price density. Machine learning models display greater variability, especially when trained on return volatility, indicating the sensitivity of results to different training rules. Finer tree partitions is more capable of capturing extreme events, highlighting volatility and kurtosis, while shorter forecast horizons with weekly data yield more pronounced negative skewness and fat tails. Overall, the framework is robust in generating price density recovery for monthly or longer forecast horizons with appropriate tree structures.

2.4 Applications

2.4.1 Fair Option Prices

As discussed before, unlike realized return, we do not directly observe the underlying return distributions. What we can observe is always only one realization of the distribution. It makes us hard to find a benchmark to evaluate the goodness of fitting for the recovered underlying distributions. One workaround is to back out the fair intrinsic values of SPX options based on recovered price density, then use recovered discount factors to get the fair values of SPX option chains for the recovery date. If both the recovered density and marginal utility coincide the true density of SPX price and investor's preference, the back-out option prices from recovery should reflect their fair values at the recovery date. Then by creating an options portfolio that trades on price discrepancy between recovery based prices and market prices with proper hedges on delta risk and time decay, such portfolios should have excess profitability. Moreover, for a better recovery result that reveals the underlying distribution more accurately, the profitability should be more significant across time. We can then use such profitability on the constructed options portfolios to evaluate the recovery results.

Specifically, at each recovery date, we calculate the fair intrinsic value for each market tradeable options based on recovered terminal price distribution, then discounted by associated recovered discount factor to get the fair values of SPX option chains. We construct the test portfolio by going long on options, both calls and puts, that are undervalued in market price compared with our recovered theoretical price, and going short on options that are overvalued in market price. Since long leg is always bleeding from time decay with negative theta, and the short leg always has positive theta, we re-weight both legs to net out the time decay effect. For example, if original long leg has a theta of $-\Theta$ and $\alpha\Theta$ for the short leg for some positive α,

we rescale the long position by α so that the portfolio is independent of PnL from time decay. Finally, the portfolio position is delta hedged by SPX futures to net out PnL from the directional changes on index. Then the reported PnL of such portfolio really reflects pricing discrepancy from distribution-related exposures, including but not limited to volatility risks. I also exclude all deep out of the money options by filtering moneyness within 0.2 to 0.8 and exclude all products with less than 1-day to expiry. Those options can have pin risks near expiry and are risky in delta hedge. Finally, all positions are held for 1-day to avoid expiry settlements. If none of products are overvalued or undervalued on the recovery date, we skip the trade for such dates since theta risk cannot be hedged by options in only one trading direction. As discussed before, good recovery results should have more predictive power on price distribution, thus should have larger PnL on distribution-related exposure for delta and theta-neutral portfolios.

Table 2.10 reports profitability of constructed options portfolio based on each model specifications and risk aversions under baseline settings. The recovery results are all based on Bates (2000) state prices and tested period spans 150 months from September 2008 to August 2023, same as the baseline recovery settings. The first column specifies the price and utility assumptions for each recovery. Column 2 to 5 report number of trades and PnL statistics based on baseline recovery results with standard tree structure and monthly forecast horizons. Column 6 and 7 report portfolio Sharpe ratio based on recovery with coarser state partition and weekly forecast horizons, respectively. For the last two portfolio specifications, all other recovery settings are the same as the baseline recovery.

From Table 2.10, we can see for all portfolios from baseline recovery results with proper state partition and monthly forecast horizon, the Sharpe ratios for such mispricing trades are uniformly positive, ranging from 0.3 to 0.5. The number of trade op-

Bates	Num Trades	Mean PnL	Vol PnL	SR PnL	SR PnL 3	SR PnL Weekly	Risk Aversion
Jump CARA	76	22854	41232	0.55	-0.70	-1.26	5.10
MLR2 CARA	89	17696	35000	0.51	-0.97	0.03	5.55
Kou CARA	82	19220	41171	0.47	-0.56	-1.33	4.76
ARCH1 CARA	103	11728	26762	0.44	-0.91	0.58	5.30
MLR CARA	89	15879	36246	0.44	-0.96	0.50	4.71
GARCH1 HARA	102	14494	32900	0.44	-0.93	0.48	6.02
MLR HARA	88	15322	34661	0.44	-0.98	0.34	6.10
ARCH1 CRRA	112	10944	25759	0.42	-0.52	0.59	8.85
MLR2 CRRA	97	13959	33024	0.42	-0.88	0.16	7.91
Kou CRRA	100	13281	31270	0.42	-0.36	-1.60	8.53
MLR2 HARA	93	13748	32531	0.42	-0.66	-0.03	6.00
GARCH1 CARA	96	14296	34647	0.41	-0.86	0.01	4.90
GBM CARA	95	11544	30127	0.38	-0.57	0.02	5.02
Jump HARA	91	11945	32163	0.37	-0.48	-1.17	5.81
ARCH1 HARA	107	8967	25234	0.36	-0.89	0.06	5.70
Jump CRRA	100	11025	31894	0.35	-0.92	-1.25	8.48
GBM CRRA	107	9216	26848	0.34	-0.70	0.00	8.46
Kou HARA	99	10648	31398	0.34	-0.55	-1.47	5.97
MLR CRRA	101	7174	22050	0.33	-0.93	0.29	8.40
GARCH1 CRRA	106	11029	33051	0.33	-0.97	0.95	8.48
GBM HARA	104	9333	30256	0.31	-0.69	0.47	5.80

Table 2.10: PnL Summary of Delta Neutral Portfolios Based on Bates Recovery

portunities, which measures how many times our recovery results find both over- and undervalued options are also similar across different model assumptions. Although 0.5 is not an impressive Sharpe ratio for options strategy, remember the idea here is not to construct the most profitable strategy, but to test if such recovery framework can consistently generate more accurate return distributions compared with current market consensus across different model assumptions. The positive Sharpe ratios for all tested price and utility assumptions across the 15 years testing period shows such recovery framework is capable of revealing distribution information that are not captured by current market prices. Across different models, CARA utility consistently provides higher SR compared with models under the same price process but other utility assumptions. This indicates the typical market investor is better characterized as a CARA investor under such tree framework. The typical recovered risk aversion centers around 5, which is also in align with previous studies.

For portfolios constructed with coarser state partitions with each starting state has 7 possible price outcomes, i.e. $n = 3$, the profitability quickly vanish for all model assumptions. The Sharpe ratio for some model specifications even crashes to near -1, indicating the proper tree structure with enough possible terminal states is crucial for a successful recovery. This result is consistent with previous empirical part. When applying such portfolio constructions to weekly forecast recovery, the Sharpe ratios are less positive compared with their monthly counterparts for most model assumptions. Especially for jump diffusion models, the Sharpe ratios are uniformly negative, indicating inferior predictive powers for such recovery framework if applied to short horizons with jump price assumptions. This is another evidence echoing previous empirical part that such recovery method is best when applied to monthly or longer horizons unless better data quality or finer parameters pre calibration methods are available. In summary, the recovery framework can generate robust superior SPX

return distribution forecast than current market consensus when applied to longer forecast horizons and with proper tree structures. This framework is also applicable to different parametric assumptions about price process and typical investor's risk preference.

2.4.2 Minimum Variance SDF

Another application is to compare recovered discount factors with the minimum variance SDF spanned by tradable assets. For incomplete market where multiple SDFs can exist, we consider a special case of stochastic discount factor (SDF) m^* that is spanned by tradable assets. In this case, the asset universe consists all SPX options of the same expiry, tradable at the recovery date. Then this SDF can be expressed as

$$m^* = \alpha^T X, \tag{2.87}$$

where α is a parameter vector representing trade quantities and X represents the payoff vector of asset universe. To price this SDF, assuming the observed market prices equal their fair values, and note expected price of such spanned product

$$\mathbf{E}(m^* \alpha^T X) = \alpha^T \mathbf{E}(X^T X)\alpha = \alpha^T P \tag{2.88}$$

for market price vector P observed from market, we get

$$\alpha = \left[\mathbf{E}(X^T X)\right]^{-1} P, \tag{2.89}$$

conditioned on the payoff matrix is invertible. Substituting the parameter α back into expression for m^*, we arrive at

$$m^* = X^T \left[\mathbf{E}(X^T X)\right]^{-1} P. \tag{2.90}$$

	Mean MV SDF	Std MV SDF %
ARCH1 CARA	0.9870	3.46
ARCH1 CRRA	0.9870	3.41
ARCH1 HARA	0.9869	3.48
GARCH1 CARA	0.9870	3.68
GARCH1 CRRA	0.9869	3.60
GARCH1 HARA	0.9870	3.65
GBM CARA	0.9869	3.72
GBM CRRA	0.9870	3.59
GBM HARA	0.9870	3.66
Jump CARA	0.9875	4.78
Jump CRRA	0.9871	4.75
Jump HARA	0.9870	4.77
Kou CARA	0.9869	4.77
Kou CRRA	0.9869	4.77
Kou HARA	0.9876	4.77
MLR2 CARA	0.9870	4.67
MLR2 CRRA	0.9867	4.58
MLR2 HARA	0.9874	4.63
MLR CARA	0.9869	4.61
MLR CRRA	0.9870	4.57
MLR HARA	0.9870	4.57

Table 2.11: Summary Statistics of Minimum Variance SDF Estimated by Recovery Price Distributions

In practice, to estimate this asset spanned SDF, we need the density $f(S_T)$ for underlying asset price S_T. The recovery result, which provides the price distribution, is therefore directly applicable in estimating the asset spanned SDF. This highlights another application of the recovery framework, as it allows us to compute this SDF from recovered asset price distribution.

This asset spanned SDF is usually called minimum variance SDF m^* since any other SDF under the same incomplete market with n linearly independent assets should be more volatile than m^*. To see this, for any other SDF m such that price vector $P = \mathbf{E}(mX)$, the difference between m and m^*, denoted by $\epsilon = m - m^*$, must satisfy

$$0_{n\times 1} = \mathbf{E}\left[(m - m^*)\, X\right] = \mathbf{E}(\epsilon X). \tag{2.91}$$

This implies ϵ is orthogonal to the payoffs X, thus also orthogonal to the minimum

variance SDF m^*, or mathematically

$$\mathbf{E}(\epsilon m^*) = 0 \tag{2.92}$$

since it is a linear combination of X. The variance of any other SDF m can be written as

$$\begin{aligned}\mathbf{Var}(m) &= \mathbf{Var}(m^* + \epsilon) \\ &= \mathbf{Var}(m^*) + \mathbf{Var}(\epsilon) - 2\mathbf{E}(\epsilon)\mathbf{E}(m^*).\end{aligned} \tag{2.93}$$

Assuming

$$\mathbf{Var}(\epsilon) - 2\mathbf{E}(\epsilon)\mathbf{E}(m^*) \geq 0, \tag{2.94}$$

or more strictly $\mathbf{E}(\epsilon) \leq 0$, we conclude

$$\mathbf{Var}(m) \geq \mathbf{Var}(m^*). \tag{2.95}$$

This establishes that m^* has the minimum variance. Specifically if there exists a risk free asset in such incomplete market, we have

$$\mathbf{E}(\epsilon X) = \mathbf{E}(\epsilon) = 0, \tag{2.96}$$

reaffirming the minimum variance of such asset-spanned SDF. Table 2.11 reports the mean and standard deviation of minimum variance SDF spanned by tradable SPX options and estimated from recovery SPX price density. The average of m^* centers around 0.985 and standard deviation ranges from 3.4% to 4.8% for different model assumptions. The estimated statistics are consistent with previous empirical studies and stable across the 15 years testing period for all model specifications. We can also find the standard deviation of such SDF is indeed much smaller than SDF implied from common utility assumptions.

Chapter 3

Profitability of Anomalies with Transaction Costs and Price Impact

3.1 Introduction

This paper simulates the price impact coefficients and exams the most common anomaly strategies after considering transaction costs from both spread and price impact. The study shows under quarterly rebalance, none of the studied anomalies remains statistically significant after costs, which differs from the previous studies only considering frictions from linear spread. The study also shows multiple transaction cost mitigation techniques including lower rebalancing frequencies and buy-hold spread save the effectiveness of anomalies at the cost of lower Sharpe ratios. With cost mitigation methods in effect, some anomalies can attract more than 500-billion dollar before the marginal trader find it unprofitable.

Since Fama and French (1992), researchers have documented hundreds of anomalies. Hou et al. (2020) notices there are at least 452 cross-sectional anomalies. In this paper, the anomaly refers to a cross-sectional strategy that generates significant positive excess return relative to the Fama and French (1992) market, size, value and momentum four-factor model. McLean and Pontiff (2016) argues most equity strategies become less profitable after publications, and estimate a 32% deduction in return from post-publication tradings. Harvey et al. (2016) also mentions higher threshold of t-statistics should be adopted for significant tests, rather than 2, and as a

consequence, most claimed anomalies are in fact insignificant even before costs. More specifically, Hou et al. (2020) shows 65% of the claimed anomalies are insignificant even with the test hurdle of 1.96, and the failure rate raises to 82% if t-value goes up to 2.78.

Trading spread and price impact in large orders also hurt the anomalies. Novy-Marx and Velikov (2016) studies 20 anomalies in addition to Fama and French (1992) four factors with linear transaction cost, and concludes without cost mitigation techniques, excess returns for many previously identified anomalies are overwhelmed by the trading spread. However, applying multiple cost mitigation methods, especially buy-hold spread, can save the profitability of many strategies, and substantially boost their capacities to attract new capital.

Kyle (1985) motivates a linear pricing rule to describe the price impact, modeling the price change as a linear function of net trading volume. Breen et al. (2002) adapts the model by using the scaled variables, rather than absolute amount of price change and volume, to enable more meaningful comparisons between different stocks. The Breen et al. (2002) model uses transaction level data from TAQ and assumes the percentage return of stock i in each 30-minute time interval t is proportional to the net share turnover of stock i in the corresponding interval, i.e.

$$\frac{\Delta p_{i,t}}{p_{i,t-1}} = \lambda_i^B \cdot TO_{i,t}, \tag{3.1}$$

where the coefficient λ_i^B measures the magnitude of price impact for each stock i. Glosten and Harris (1988) decomposes the price impact into a fixed component resulting from the direction change of two consecutive trades and a variable component proportional to the signed trading volume

$$\Delta p_{i,t} = \alpha_i + \Phi_i \Delta d_{i,t} + \lambda_i^G q_{i,t} + \epsilon_{i,t}, \tag{3.2}$$

where direction indicator $d_{i,t}$ is set to be $+1$ for buy and -1 for sell orders, $q_{i,t}$ is the signed net share of stock i traded in period t, Φ_i and λ_i^G are corresponding coefficients. Empirical studies support a concave relation between price impact and net trading volume, but both specifications are either convex (Breen et al. (2002)) or linear (Glosten and Harris (1988)). Hasbrouck (1991) and Bertsimas and Lo (1998) argue non-concavity does not change the results significantly for retail traders, but leave the answers open for large liquidity demanders. For institutional investors, large trades are usually broken into smaller orders to minimize the price impact, at the risk of some tail orders may not be executed before expiration.

To infer transaction cost from incomplete data, Hasbrouck (2009) generalizes a Bayesian Gibbs method and applies it to Roll (1984) model. This method requires only daily closing price data from CRSP to output Gibbs estimates of spread for each stock-year, and thus enjoys much wider coverage than methods requiring higher quality data. This method also fits well on other regression models for coefficient estimation with missing data on regressor values. Hasbrouck (2009) then justifies his Gibbs/CRSP estimates by comparing it with the estimates from real transaction level data, and concludes a high correlation of 0.965 between the Gibbs estimates and TC estimated from real data.

Compared with Fama-MacBeth regressions on proxies, using MCMC methods including Gibbs sampler for coefficient estimation is more time-consuming, and requires relative long records of daily price data from CRSP. However, Novy-Marx and Velikov (2016) points out that cost estimates from proxy regressions are inaccurate to extrapolate due to their non-linearity and time varying nature. Moreover, Gibbs estimates enjoy larger data coverage and wider access than proxy methods requiring TAQ or proprietary datasets.

Novy-Marx and Velikov (2016) also provides a new measure to evaluate the

anomaly strategy performance as opposed to the baseline model. The generalized alpha, rather than traditional alpha, they proposed is designed with net-of-cost cross-sectional returns, and unambiguously represents the improvement new strategy contributes to the investment opportunity set of the baseline model. Followed by previous literature, their baseline model is Fama and French (1992) four-factor model constructed with 30 and 70 percentile NYSE break points.

This paper further takes cost from price impact into consideration. I first use daily price data on CRSP and Gibbs sampler method to calibrate two price impact models for each stock-year. My Gibbs/CRSP estimates for price impact are consistent with previous findings with higher quality data in terms of magnitude and correlation results from proxy regression studies. The restricted use of only CRSP data enables further studies with price impact for researchers with limited data access. Then, all long-short anomaly portfolios are reconstructed with costs from both spread and estimated price impact to exam the after-cost significance of different anomalies. Different from previous studies, I use the same decile break points to construct portfolios for both the test assets and baseline model to facilitate more meaningful comparisons between different categories of anomalies. Under the newly constructed baseline portfolio, quarterly rebalance and no cost mitigation techniques, I find none of the studied anomalies remains significant in terms of generalized alpha. Fortunately, applying multiple cost mitigation methods including lower rebalancing frequencies and buy-hold spread combined saves five out of the nine anomalies examined, with lower rebalancing frequency being the single most effective method to cut the cost. I also study the maximum market capacity for each anomaly strategy to attract new capital before the marginal trader find the profit insignificant. As a result, anomalies with more significant generalized alpha, such as net-issuance, also enjoy stronger capacities to attract new capital.

3.2 Trading Cost Models

In order to exam the after-cost significance of claimed anomalies, we first need to modify transaction cost models to incorporate the role of net trading volume and its concave effect on the price change. To capture the concavity between price impact and trading volume, I propose two modifications to Roll (1984) and Breen et al. (2002) models. In the original Roll model

$$\begin{cases} v_t = v_{t-1} + u_t \\ p_t = v_t + c \cdot d_t \end{cases} \tag{3.3}$$

where v_t is the intrinsic value of the asset (quote closing bid-ask midpoint of previous trade), u_t is a random innovation term reflecting disturbance from public information, p_t is the observed trade price, c is the proportional cost of trading measured by one half of quoted spread, and d_t is the trade direction indicator that takes value $+1$ for buy orders and -1 for sell orders. (3.3) yields

$$\Delta p_t = c \cdot \Delta d_t + u_t. \tag{3.4}$$

Assuming price sequence $\{p\}_t$ is observed by data, Roll (1984) calibrates the cost from his model by

$$c = \sqrt{-Cov(\Delta p_t, \Delta p_{t-1})}, \tag{3.5}$$

where $Cov(\Delta p_t, \Delta p_{t-1})$ is the first order covariance of price change and is estimated by the sample moments. This moment estimate takes no assumption on the distribution of u_t and is easy to implement. However, (3.5) only applies to price change data with non-positive first-order sample auto-covariance, which accounts for around half the cases. Harris (1990) assigns a prior value of zero for positive auto-covariance

price changes, and recognizes the result as a lower bound for the real cost. By further assuming the innovation term u_t is i.i.d. normal with zero mean and unknown variance σ_u^2, Hasbrouck (2009) advocates a Bayesian Gibbs approach to estimate c and σ_u^2, together with direction sequence $\{d\}_t$. He advocates this approach can yield estimates that are highly correlated with the cost extracted from real transaction level data despite of the computation burdens and is superior to the first moment estimate. My estimations follow this Bayesian Gibbs method with the modified models.

To allow price impact from net trading volume in the model, I modify the Roll model as follow

$$\begin{cases} v_t = v_{t-1} + u_t \\ p_t = v_t + c_t \cdot d_t + \lambda q_t^{\alpha} \cdot d_t, \end{cases} \tag{3.6}$$

where v_t is the midpoint of previous trading day's closing bid-ask quote, p_t is the adjusted previous closing price. Specifically, p_t is adjusted to previous ask if closing price of the current trading day is greater than intrinsic value v_t, and is adjusted to previous bid if price comes short of the value. No adjustment is needed if the two values are equal. Consequently, trade direction indicator d_t is set to 1 if the concurrent closing price is greater than intrinsic value, 0 if price coincides with value and -1 otherwise. Since closing price of each day is either the closing bid or closing ask, this adjustment matches buy orders to asks, sell orders to bids, and as a consequence, removes the price change effects from spread for two consecutive trades of the same direction. The new term $\lambda q_t^{\alpha} \cdot d_t$ describes the signed effect of price impact from net dollar trading volume q_t of trading day t, and is signed by direction indicator d_t. Scale coefficient λ reflects the magnitude of price impact, and exponential shape coefficient $\alpha \in (0, 1)$ captures the concavity effect from net trading volume. Under this modified Roll model (3.6), a trade of q dollar on the stock will move the price away from its

intrinsic value by $(c + \lambda q_t^{\alpha})$ dollar, where the first term is observed from half quoted spread and the second term is the price impact to be estimated.

The second model for price impact is inspired by Breen et al. (2002). The paper assumes a linear relation between percentage price change and percentage net trading volume

$$\frac{\Delta p_{i,t}}{p_{i,t-1}} = \lambda_i \cdot TO_{i,t} = \lambda_i \cdot \frac{q_{i,t}}{Q_i} := \lambda_i' \cdot q_{i,t}, \tag{3.7}$$

where $p_{i,t}$ is the last transaction price of asset i in period t, λ_i denotes the price impact coefficient for stock i, $q_{i,t}$ measures the net number of shares traded of stock i in period t, and Q_i is the number of shares outstanding for stock i. To estimate the price impact coefficient λ for each month τ, they use transaction level price volume data from TAQ, and define the time period to be consecutive 30-minute slots during the trading hours in the studied month. The introduction of percentage price change and trading volume enables more meaningful comparisons between different stocks, but also brings convex relations between the price impact and net trading volume. To see this, re-write the above equation (3.7) in differential format

$$\begin{cases} \frac{d(f+p_i)}{f+p_i} = \lambda_{i,\tau}' \cdot dq_i \\ f(p_i, 0) = 0, \end{cases} \tag{3.8}$$

i.e. no price impact with zero trade volume. It gives

$$f(p_i, q_i) = p_i(e^{\lambda_{i,\tau}' q_i} - 1), \tag{3.9}$$

where for each stock i, $f(p_i, q_i)$ denotes the price impact function, p_i is the state variable of price level, $\lambda_{i,\tau}'$ is defined as $\frac{\lambda_{i,\tau}}{Q_i}$ which is constant for each month τ and q_i is the net number of shares traded on this stock.

Since empirical studies support a concave relation between price impact and trad-

ing volume, Breen et al. (2002) model leads to a conservative (larger) estimate of non-linear cost. To adjust for it, I add an exponential coefficient α on the turnover rate

$$\frac{\Delta p_{i,t}}{p_{i,t-1}} = \lambda_i \cdot \left(\frac{q_{i,t}}{Q_{i,t}}\right)^{\alpha_i} := \lambda_i \cdot \tilde{q}_{i,t}^{\alpha_i}. \tag{3.10}$$

Re-write (3.10) in differential format, we have:

$$\begin{cases} \frac{d(f+p_i)}{f+p_i} = \lambda_i \cdot \tilde{q}_i^{\alpha_i} \\ f(p_i, 0) = 0, \end{cases} \tag{3.11}$$

which gives a new price impact function

$$f(p_i, q_i) = p_i(e^{\lambda_i \tilde{q}_i^{\alpha_i}} - 1). \tag{3.12}$$

The above model measures price impact given price level p and share trade size q for stock i in period τ. For exponential coefficient $\alpha \in (0, 1)$ and given price level p, the price impact function $f(q)$ is concave for small q before turning convex. This bounded range of concavity is reasonable for estimation since large block trades are usually split into smaller orders. Now with this modified model, estimating transaction costs with price impact is equivalent to estimating two coefficients α and λ for each stock-year.

The two cost models adopt different unit for trade size. Roll (1984) model measures trade size in dollar value, while Breen et al. (2002) describes volume by net number of shares traded. The specifications give us more flexibility in choosing cost models according to the description of volume data. It also worth mentioning in both models, we only consider market orders and assume all shares are traded at the same price within a single order. Time difference between order sent from traders and order being executed by the market maker is also ignored. One major simplicity resulting from these assumptions is the transaction cost measured by spread and price impact

combined is a deterministic value, rather than a random variable affected by decisions from other market participants. Engle et al. (2012) discusses the optimal execution strategy when the cost is a random variable, and use mean-variance analysis to study cost estimation allowing for order split and effects from other market participants.

3.3 Gibbs Sampler Estimations

3.3.1 Data

The sample period to simulate price impact coefficients for both models covers 20 years from 2002 to 2021, and the universe includes all common stocks available on CRSP with at least 100 trading days at the corresponding calendar year. I only pick stocks with at least 100 trading days because my Bayesian estimates are for each firm-year and need enough data in each year to update the posterior beliefs. The annual cost coefficients for each firm-year are then estimated for both models. In this study, cost coefficients of over 19,000 firms are estimated across 20 years, with most firms only existing for parts of the period and around 7,000 firms estimated for each year.

3.3.2 Modified Roll Model

The modified Roll model

$$p_t = v_t + (c_t + \lambda q_t^{\alpha}) \cdot d_t, \tag{3.13}$$

where q_t is the dollar trading volume indicates a block trade of q_t dollar will move share price from its fundamental value by $(c_t + \lambda q_t^{\alpha})$ dollar. To calibrate the model, I regress the log price on log of price impact function:

$$log(\lambda q_t^{\alpha}) = log\lambda + \alpha \cdot logq_t := \gamma + \alpha s_t. \tag{3.14}$$

Specifically, let p and v be the log of daily closing price and intrinsic value induced from the midpoint of the log bid-ask spread, fixed cost c_t equal to one half of log bid-ask spread, and r_t^m be the daily market return, we have

$$p_t = v_t + (c_t + \gamma + \alpha s_t) \cdot d_t + \beta_m r_t^m + u_t, \tag{3.15}$$

where $u_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2)$ with unknown innovation variance σ_u^2. In this model, daily log closing price p_t, intrinsic value v_t, direction indicator d_t and daily market return r_t^m for each trading day t are observed or calculated from CRSP daily level data, while parameters γ, α, β_m and variance of innovation term σ_u^2 for each sampled firm-year are left to be estimated. Note followed by Hasbrouck (2009), we add daily market return in the regression because any term to further separate the price change from costs can sharpen the cost estimates. Define data term $y_t = p_t - v_t - c_t d_t$, we have

$$y_t = \beta_m r_t^m + \gamma d_t + \alpha(s_t d_t) + u_t, \tag{3.16}$$

where $u_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2)$, sequences $\{y, r^m, d\}_t$ are observed from data, coefficients together with latent variable sequence $\{s\}_t$ are left unknown. By considering (3.16) as a regression with unspecified data $\{s\}_t$, instead of applying traditional OLS estimates when the net log volume is observed, we can use Bayesian Gibbs sampler to estimate coefficients and sequence of latent variables at the same time.

3.3.3 Procedures for Bayesian Gibbs Sampler

Before sampling, we first need to set up priors for parameters to be estimated. Good assumptions on prior distributions can lead to closed-form posteriors, and thus boost the computational efficiency. Besides, prior distribution parameters that are close to real values can also speed up the convergence. Note for joint normal likelihood and fixed data $\{r^m, d, s\}_t$, normal prior distributions for coefficients $\{\beta_m, \gamma, \alpha\}$ will lead

to closed-form conjugate normal posterior distributions, and inverted gamma prior distribution on innovation variance σ_u^2 will also give closed-form conjugate posterior. See Appendix B for the proof of conjugate posteriors and choice for parameter values. Geweke (2005) also shows for standard Bayesian normal regression model $y = Xb + e$ with both multivariate normal priors on b and e, when b is restricted on a finite domain, the posterior for b has the same parameters as the unrestricted case, only to be truncated to the same domain as in the prior. Equipped with the above results, we now set priors for each variable to be estimated and use Gibbs sampler to sample one variable each at a time, rather than sampling all variables together from a higher dimensional joint distribution. The procedure to implement Bayesian Gibbs sampler is as follows:

-Set up initial priors for coefficients $\beta_m, \gamma, \alpha, \sigma_u^2$ and latent data $\{s\}_t$ as below:

$$\begin{aligned}
\alpha &\sim N(1/2, (1/6)^2) \\
\gamma &\sim N(-12, 2^2) \cdot \mathbf{1}_{[-14,-10]} \\
\beta_m &\sim N(0.75, 10^2) \cdot \mathbf{1}_{[-3,5]} \\
s_t &\sim N(24, 60^2) \cdot \mathbf{1}_{[20,28]} \\
\sigma_u^{-2} &\sim \Gamma(2, 2 \times 10^{-5})
\end{aligned}$$

-Generate one sample for each variable from its prior in sequence, given all other variables known:

Sample s_t conditioned on all latest coefficients and $\{s_{-t}\}$ from posteriors, for each t in sequence.

Sample σ_u^2 conditioned on latest other coefficients and $\{s\}_t$

Sample β_m conditioned on latest other coefficients and $\{s\}_t$

Sample α conditioned on latest other coefficients and $\{s\}_t$

Sample γ conditioned on latest other coefficients and $\{s\}_t$

-Repeat above sampling procedure until both the averages of $\hat{\gamma}$ and $\hat{\alpha}$ converge.

Since the goal is to estimate the cost parameters, we only care about the estimation results for parameters α and γ. The estimated shape parameter for modified Roll model $\hat{\alpha}$ is uniformly around 0.5 for different companies and across the 20-year period, with a minimum of 0.45, mean/median of 0.49 and maximum of 0.54. The standard deviation is around 0.01. The result is consistent with the square root rule of trading volume for price impact in previous literature. The estimated scale parameter $\hat{\gamma}$ falls between 2.31×10^{-6} to 12.38×10^{-6} for different firm-years, with a mean of 5.30×10^{-6} and median of 5.21×10^{-6}. The result is in align with that of Hasbrouck (2009) which reports the scale of price impact has a median of around 7×10^{-6}. Despite of the consistency with previous results, we notice a lack of significant right skewness as the scale estimates reported by Hasbrouck (2009). Also, the standard deviation of $\hat{\gamma}$ is much less than the previous result ($\sigma_{\hat{\gamma}}^{Has}$ around 70, $\sigma_{\hat{\gamma}}^{Gibbs}$ around 1). The difference is possibly resulted from the studied data itself. In this study, I only consider stocks with at least 100 trading days in the given calendar year, and thus exclude stocks with extreme price impact approximations estimated from insufficient data, and also short-lived stocks that are very sensitive to the price impact. The exclusion of these anomalous stock-years then yields a much smaller sample standard deviation. With this in mind, the difference in skewness and sample variance are minor and do not harm the consistency between two results.

3.3.4 Modified Breen Model

Using the same Gibbs sampler procedure, we can also calibrate the modified Breen et al. (2002) model with price impact coefficients

$$\frac{\Delta p_t}{p_{t-1}} = \lambda \cdot \left(\frac{q_t}{Q_t} \right)^{\alpha}, \tag{3.17}$$

where q is the net volume traded in shares, Q is the total number of shares outstanding and Δp_t is adjusted such that it equals bid minus bid if $\Delta p_t < 0$, and ask minus ask otherwise. Thus, the fixed cost effect is removed after adjustment and TC per share equals $c + p\left(e^{\lambda \tilde{q}_t^{\alpha}} - 1\right)$, where $\tilde{q} = \frac{q}{Q}$.

To calibrate (3.10), I regress the log of absolute return on the log of price impact as

$$log|\frac{\Delta p_t}{p_{t-1}}| = log\lambda + \alpha \cdot log\tilde{q}_t + u_t \tag{3.18}$$

and denote it as

$$y_t = \gamma + \alpha s_t + u_t, \tag{3.19}$$

where $u_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2)$ with unknown σ_u^2. Data y_t is calculated from CRSP daily closing price data, and parameters $\alpha, \gamma, \sigma_u^2$ together with latent data sequence $\{s\}_t$ are to be estimated. Using conjugate distribution results as before, we set normal priors for parameters as follows:

$$\begin{aligned}
\alpha &\sim N(1/2, (1/6)^2) \\
\gamma &\sim N(-10, 2^2)\mathbf{1}_{[-12,-8]} \\
s_t &\sim N(10, 50^2)\mathbf{1}_{[8,12]} \\
\sigma_u^{-2} &\sim \Gamma(2, 5 \times 10^{-8}),
\end{aligned}$$

and the procedures are the same as estimating modified Roll model: sampling one variable each at a time to avoid drawings from the high dimensional joint distribution.

In modified Breen model (3.10), estimates for the shape parameter $\hat{\alpha}$, which controls the concavity of price effect, range from 0.39 to 0.68, with both mean and median around 0.56 and a standard deviation of 0.03. The result is higher than that from

the modified Roll model (3.6), echoing Almgren et al. (2005) that 0.6 is a more appropriate α estimate than 0.5. The estimates for scale parameter, which controls the magnitude of the price effect, are more volatile across different firm-years. The estimates range from 0.47×10^{-5} to 7.43×10^{-5} across the sample period and the entire universe, with mean and median both around 2.35×10^{-5} and a standard deviation of 0.79. The result is consistent with Korajczyk and Sadka (2004) in terms of magnitude and no right-skewness. One minor difference between the two result is the maximum value from Gibbs estimate is much less than that reported by Korajczyk and Sadka (2004). However, since Korajczyk and Sadka (2004) uses transaction level data from TAQ and covers the entire universe on CRSP, while our Gibbs sampler only learns from daily closing price data on selected stocks, this difference does not harm the justification of our results.

Proxy regression is another popular method to estimate the price impact despite of the need for higher quality data and inaccuracy in extrapolation. Korajczyk and Sadka (2004) uses nine firm-specific proxies to regress the scale price impact coefficient λ and finds the positive correlation between price impact and the corresponding firm size variable ME. Using this result, we visually check the correlation between our scale parameter estimates from CRSP/Gibbs method and corresponding firm sizes. The following Figure 3·1 shows average price impact levels from our estimates for both top and bottom ME deciles in each year. We can find large firms uniformly suffer from greater price impact levels in all years according to our estimates. Also, price impact levels for large firms are usually (more) elevated in distressed periods, which further justifies the estimated results.

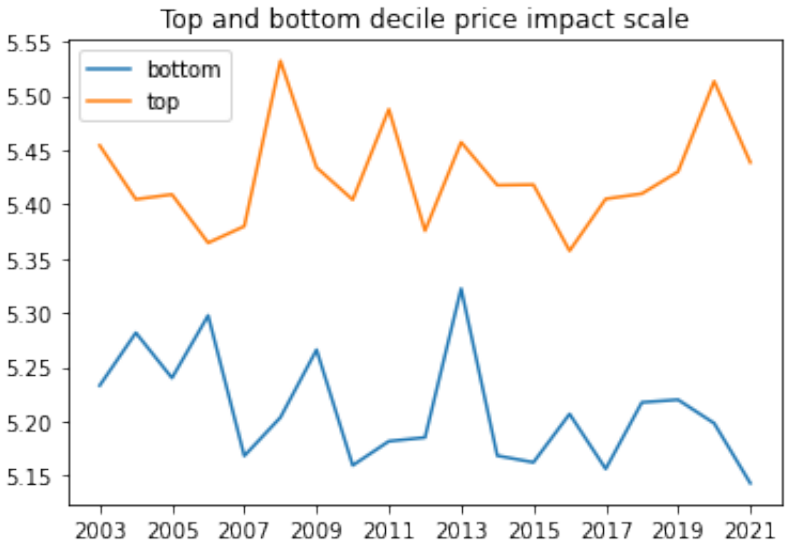

Figure 3·1: Level of Price Impact Estimates for Top and Bottom ME Decile Stocks in Each Year

3.4 Anomaly Significance under MVE Portfolios

3.4.1 Portfolio Constructions and Performance Evaluation

In this section, I exam the performance of nine anomalies in addition to size, value and momentum after fixed cost and price impact. The anomalies are listed in Table 3.1. All strategies are expressed as a time series of value-weighted long-short self-financing portfolio returns. The portfolios are constructed with top and bottom deciles with quarterly fundamentals data on Compustat and are rebalanced each quarter. Table 3.1 also indicates the name, reference and construction method for each anomaly.

All accounting data to calculate factor realizations are from Compustat, and the cost estimates are from Gibbs samplers as described in previous sections. To build the long-short factor portfolios, all stocks are sorted based on factor realizations before longing all the stocks in the top decile, and short selling the bottom one. All portfolios are value-weighted and are rebalanced at the end of each quarter. The return for

Anomaly	Reference	Signal
Size (ME)	Fama and French (1993)	Market Equity (ME)
Value (Val)	Fama and French (1993)	Book-to-market Equity (BE/ME)
Reversal (Rev)	Jegadeesh and Titman (1993)	Prior year's stock performance excluding the most recent month
Gross Profit (GP)	Novy-Marx (2013)	Gross Profit / Total Asset (GP/AT)
Accruals	Sloan (1996)	Sum of annual change in total current assets, debt in current liabilities, income taxes payable; negative change in total cash, current liabilities, depreciation, over average total assets for the last two years.
Asset Growth	Cooper et al. (2008)	Asset Growth (AT/AT_{-1})
Investment (Invest)	Lyandres et al. (2008)	Sum of annual change in gross total property, plant, and equipment, and annual change in total inventories, over lagged total assets.
Net Issuance (Net_Issue)	Fama and French (2008)	Annual change in adjusted shares outstanding (ADJEX×CSHO)
Return on Book Equity (Ret_BE)	Chen et al. (2011)	Income before Extraordinary Items/ Book Equity (IB/BE)
Value Profit	Novy-Marx (2015)	Sum of ranks in univariate sorts on Value and Gross Profitability
Value Reversal	Novy-Marx (2015)	Sum of ranks in univariate sorts on Value and Reversal
Value Reversal Profit	Novy-Marx (2015)	Sum of ranks in univariate sorts on Value, Reversal and Gross Profitability

Table 3.1: The Descriptions of Nine Anomalies Examined in Addition to Fama and French (1992) Four Factors

each factor portfolio is recorded as the sum of after-cost returns from the long, short portfolios combined. See Appendix B for detailed steps on portfolio net-of-cost return calculations.

Followed by Novy-Marx and Velikov (2016), generalized alpha, rather than traditional alpha unambiguously represent the improvement a factor makes on the investment opportunity set after accounting for trading costs. It is therefore a more appropriate standard to evaluate the net-of-cost profitability of an anomaly. Let MVE_X be the ex-post mean variance efficient (MVE) portfolio of a given investment opportunity set (assets) X or in a convenient abuse of notation, its return; $w_{\{y,MVE_{\{X,y\}}\}}$ be the weight on asset y in portfolio $MVE_{\{X,y\}}$, then the generalized alpha α^* is defined from the regression

$$\frac{MVE_{\{X,y\}}}{w_{\{y,MVE_{\{X,y\}}\}}} = \alpha^* + \beta^* MVE_X + \epsilon^*. \tag{3.20}$$

In other words, α^* for anomaly y is the abnormal return of the rescaled test portfolio $MVE_{\{X,y\}}$ relative to the baseline portfolio MVE_X, where $MVE_{\{X,y\}}$ is rescaled to one dollar of asset y in $MVE_{\{X,y\}}$. In case asset y does not improve the opportunity set, i.e. baseline portfolio MVE_X coincides with $MVE_{\{X,y\}}$, $w_{\{y,MVE_{\{X,y\}}\}}$ becomes zero and they define α^* to be zero as well. Novy-Marx and Velikov (2016) show their generalized alpha (3.20) agrees exactly with traditional alpha in the frictionless world, but reflects the marginal contribution new asset y makes to the original portfolio. Traditional alpha on the other hand, may falsely exaggerate the improvement new asset can make since cost can exceed the profit from trading the strategy. To see this, assume a new asset y is orthogonal to the baseline portfolio X with high gross return but even higher trading cost. Then the strategy has high gross (traditional) alpha, but the cost prevents its realization, and as a result, y does not improve the original investment opportunity set. On the other hand, short selling the strategy and collect

the costs generates positive net alpha. But earning the cost is also impossible for stock traders, making no investor wants to trade on this new asset. The negative generalized alpha successfully captures this result as opposed to the traditional standard. Another important detail to notice is that all anomaly assets in the test portfolio $MVE_{\{X,y\}}$ are formed by the top and bottom decile breakpoints. In this paper, I aim for studying the profitability of different anomaly categories, rather than anomalies with same economic interpretation but different breakpoints. To keep the consistency between portfolios on both sides of the regression (3.20), the anomaly assets in the baseline portfolio (regressor) are also constructed with the same decile breakpoints. To see the difference, Novy-Marx and Velikov (2016) uses top and bottom decile breakpoints to construct the test portfolios, but NYSE 30%, 70% breakpoints for the baseline portfolio. They then report the generalized alpha for (decile) momentum anomaly as 0.4 (t value of 3.12) as apposed to FF4. My construction using the same breakpoints on both sides of the regression avoids this potential conflict since for any y in asset universe X, $MVE_{\{X,y\}}$ always coincides with MVE_X, and α^* is by definition zero.

3.4.2 Summary Statistics for Anomalies and MVE Portfolio Analysis

In this study, I am interested in the after-cost profitability of different categories of anomalies. The sample period covers all quarters from 2002 to 2021. Under quarterly rebalance, Table 3.2 reports the quarterly average gross return before costs, net return after costs from both spread and price impact estimates, the standard deviation of net return and turnover rates. Transaction costs are calculated as the difference between the first two columns.

From Table 3.2, we can see all anomalies, including FF4 factors, have positive returns before and after cost adjustment, where momentum is constructed as reversal strategy. Aside from FF4 factors, three ranking anomalies Val_Prof, Val_Rev and Val_Rev_Prof proposed by Novy-Marx (2015) enjoy the highest net average returns,

	Mean Gross Ret	Mean Net Ret	Std Net Ret	TO rate	T-costs
MKT	0.1173	0.1173	0.1804	0.0000	0.0000
ME	0.7642	0.5430	0.6053	0.5147	0.2213
Val	0.4810	0.4167	0.3885	0.7221	0.0642
Rev	0.1466	0.1245	0.2509	1.2376	0.0221
GP	0.0632	0.0595	0.3710	0.5379	0.0037
Val_Prof	0.5212	0.4567	0.3879	0.7815	0.0645
Accruals	0.0442	0.0200	0.1277	1.9126	0.0242
Asset_Growth	0.2338	0.1943	0.2805	1.8194	0.0395
Invest	0.0634	0.0410	0.1181	1.7329	0.0224
Net_Issue	0.0426	0.0369	0.1178	0.3299	0.0057
Ret_BE	0.1102	0.0945	0.2390	1.1022	0.0157
Val_Rev	0.2689	0.2267	0.3349	1.1810	0.0422
Val_Rev_Prof	0.4530	0.3775	0.4894	1.2314	0.0755

Table 3.2: Summary Statistics under Quarterly Rebalance

all beyond 20% per year. However, the net returns for anomalies gross-profit, accruals, investment and net-insurance are less significant. In terms of turnover rate, six out of nine anomalies examined have a turnover rate higher than 100% annually, indicating any techniques to curb the turnover rates may largely increase their net returns. Accruals, asset-growth and investment's turnover rates are even near 200% per year, significantly undermining their strategy performance. On the other hand, anomalies net-issuance, gross-profit and Val_Prof enjoy lower turnover rates, indicating that they are less disturbed by the costs and are more consistent fundamental indicators than other anomalies.

The following Table 3.3 represents the weight of each asset in the MVE portfolios of the baseline FF4 model and test portfolios. Sharpe ratios of the MVE portfolios are also reported. Specifically, it reports the MVE portfolio weights for the baseline FF4 model and portfolio $\text{MVE}_{\{FF4,y\}}$ for each anomaly y in the second column. The last column reports the Sharpe ratio for each strategy MVE portfolio. In the baseline model, value and size anomalies dominate the MVE portfolio, taking over 60% and 30% respectively, while the reversal portfolio is not included, indicating that reversal strategy does not improve the investment opportunity set in addition to the FF3

	Weight	SR
FF4	[0.05, 0.33, 0.62, 0.0]	1.32
Val_Rev	[0.05, 0.33, 0.62, 0.0, 0.0]	1.32
Val_Rev_Prof	[0.05, 0.33, 0.62, 0.0, 0.0]	1.32
Accruals	[0.0, 0.23, 0.47, 0.0, 0.31]	1.33
Net_Issue	[0.09, 0.2, 0.35, 0.0, 0.36]	1.35
Invest	[0.09, 0.12, 0.35, 0.0, 0.44]	1.36
Val_Prof	[0.27, 0.19, 0.17, 0.0, 0.38]	1.49
Asset_Growth	[0.31, 0.05, 0.3, 0.0, 0.34]	1.50
GP	[0.21, 0.14, 0.37, 0.0, 0.28]	1.57
Ret_BE	[0.32, 0.09, 0.16, 0.0, 0.43]	2.16

Table 3.3: MVE Portfolio Weights for Baseline Model and Each Anomaly Strategy with Sharpe Ratios

model in the recent 20-year sample period. This result contradicts conclusion from Novy-Marx and Velikov (2016) that momentum-based anomalies are still significant relative to FF4. Surprisingly, despite of impressive quarterly return, two ranking anomalies Val_Rev and Val_Rev_Prof also fail to improve the FF4 model because the Sharpe ratios of the corresponding test assets remain the same as that of the MVE_{FF4} portfolio. One possible explanation is the ranking anomaly based on value and reversal can be interpreted as a re-weighted combination of FF4 components, and thus making no contribution to introduce new anomaly categories. Triple ranked anomaly Val_Rev_Prof brings in gross-profitability but the effect is diluted by FF4 value and reversal factors. This argument can be further justified by both anomalies Val_Prof and GP successfully improving the Sharpe ratios, with GP alone providing a larger improvement. Anomaly return-on-book-equity boosts the Sharpe ratio most and the tested anomaly takes over 40% weight in the new MVE portfolio. Large weight in the new MVE portfolio shows the big contribution the anomaly makes to the efficiency of combined MVE portfolio. Gross-profitability and asset-growth also improve the Sharpe ratio by around 0.2. An interesting observation is momentum strategy takes no weight in all MVE portfolios and cannot help to improve the investment opportunity set.

A further regression of the $\text{MVE}_{\{FF4,y\}}$ on MVE_{FF4} for each anomaly y shows none of the nine examined anomaly generates significant α^* relative to the Fama and French (1992) 4-factor model after accounting for TC from spread and price impact combined. It differs from earlier results that only considering linear trading cost and using traditional NYSE breakpoints, suggesting under quarterly rebalance and without any cost mitigation techniques, most claimed anomalies in earlier literature cannot generate significant (generalized) alpha as apposed to the 4-factor model. A further analysis shows for a fund of one-billion dollar size, muting price impact cost cannot save any of the tested anomaly. Combined with earlier studies, it implies for a fund around one-billion dollar size, different choice of baseline portfolio breakpoints is the dominating variable for the effectiveness of anomaly strategies. The price impact effect becomes more apparent as the fund size grows and I will discuss the effect of fund size, including its maximum capacity to attract new capital, in the following sections.

3.5 Cost Mitigation Techniques and Analysis

The regressions of anomaly portfolios (3.20) designed in the previous section show none of the tested anomaly remains significantly profitable after costs from spread and price impact combined. However, things turn out to be more optimistic in real applications. In practice, traders can submit limit orders rather than only passively accepting the market price to reduce impact of unfavorable price movements. For large institutional managers, they can also split the parent order into smaller child orders to further control the price impact it may have on the market and hide their trading incentives. Lastly, multiple cost mitigation techniques are also frequently applied in practice. I am interested in whether mainstream mitigation techniques including reducing rebalance frequency and buy-hold spread can effectively save the anomalies.

	Mean Gross Ret	Mean Net Ret	Std Net Ret	TO rate	T-costs
MKT	0.1203	0.1203	0.1746	0.0000	0.0000
ME	0.3821	0.3050	0.4111	0.2538	0.0771
Val	0.1185	0.1068	0.3650	0.2926	0.0117
Rev	0.1699	0.1560	0.3225	0.4829	0.0140
GP	0.1005	0.0962	0.2071	0.2714	0.0043
Val_Prof	0.2019	0.1922	0.2519	0.2989	0.0097
Accruals	-0.0961	-0.1029	0.1293	0.4541	0.0069
Asset_Growth	0.1965	0.1862	0.2549	0.4532	0.0103
Invest	0.0427	0.0366	0.2142	0.4027	0.0061
Net_Issue	0.0546	0.0499	0.1159	0.1968	0.0046
Ret_BE	0.0693	0.0639	0.2227	0.3236	0.0054
Val_Rev	0.2058	0.1869	0.7126	0.4429	0.0190
Val_Rev_Prof	0.2044	0.1879	0.3758	0.4424	0.0165

Table 3.4: Summary Statistics under Annual Rebalance

Earlier studies including Novy-Marx and Velikov (2016) finds other methods like only rebalancing part of the portfolio (staggered partial rebalancing) or concentrating trading among low-cost stocks are inferior to the two methods studied in this paper.

Reducing rebalance frequency from quarterly basis to annually can effectively control the ceiling of annual turnover rate and reduce the cost, but at the price of deviating the strategy portfolio from its target within the year. Table 3.4 shows the the annual average gross return before costs, net return after costs from both spread and price impact estimates, the standard deviation of net return, turnover rates and transaction costs as the difference between the first two columns for each anomaly.

Compared with their quarterly rebalance counterparts, the annual returns for most anomalies decrease, especially for anomaly accruals, whose average gross annual return becomes negative. It implies the loss caused by deviation from target portfolio exceeds the gross excess return for this anomaly even before paying the transaction costs. Similar to the original rebalancing strategies, ranking anomalies together with asset-growth still enjoy the highest positive returns, with all three ranking anomalies profiting over 20% and 18% before and after cost, respectively. As a compensation for lower returns, turnover rates for all strategies drop drastically, especially for ranking

	Weight	SR
FF4	[0.55,0.4, 0.0,0.06]	0.75
Accruals	[0.55,0.4, 0.0, 0.06, 0.0]	0.75
Val_Rev	[0.55,0.4, 0.0, 0.06, 0.0]	0.75
Val_Rev_Prof	[0.53,0.4,0.0, 0.0, 0.07]	0.75
Invest	[0.31,0.16, 0.0, 0.21, 0.32]	0.79
Val_Prof	[0.25,0.3,0.0, 0.0, 0.45]	0.82
GP	[0.35,0.24, 0.0, 0.0, 0.41]	0.90
Net_Issue	[0.28,0.17,0.0, 0.01, 0.54]	0.91
Asset_Growth	[0.18,0.08, 0.0, 0.25, 0.48]	1.07
Ret_BE	[0.2,0.25,0.0, 0.05, 0.51]	1.41

Table 3.5: MVE Portfolio Weights for Baseline Model and Each Anomaly Strategy with Sharpe Ratios under Annual Rebalance

anomalies. The turnover rates for most anomalies are reduced by two thirds, with even less than 20% for anomaly net-issuance, and less than 50% for all other anomalies.

Table 3.5 shows the MVE portfolio weights and Sharpe ratios for annually rebalanced strategies. It reports the mean-variance efficient portfolio weights for the baseline FF4 model and portfolio $\text{MVE}_{\{FF4,y\}}$ for each anomaly y in the second column. The last column reports the Sharpe ratio for each strategy MVE portfolio.

The Sharpe ratios for all portfolios are negatively affected by lower rebalancing frequency including the baseline MVE_{FF4} portfolio. The Sharpe ratio for the baseline model drops by over 40%, from 1.32 to 0.75. Same as in the quarterly rebalance case, ranking anomalies Val_Rev and Val_Rev_Prof cannot improve the Sharpe ratios against the FF4 because of the lack for new anomaly categories. Also, anomaly accruals takes no weight in the new MVE portfolio, making portfolio $\text{MVE}_{\{FF4,accruals\}}$ coincides with portfolio MVE_{FF4} since trading on accruals on average incurs negative net return. Anomalies return-on-BE and asset-growth remain to be the strategies that improve the baseline model the most in terms of the Sharpe ratio. Benefited from the low turnover rate, net-issuance also improves the MVE portfolio of FF4 by over 20% from 0.75 to over 0.9. More importantly, the regressions (3.20) show the above three anomalies net-issuance, asset-growth and return-on-BE generate significant general-

	Mean Gross Ret	Mean Net Ret	Std Net Ret	TO rate	T-costs
MKT	0.1203	0.1203	0.1746	0.0000	0.0000
ME	0.2388	0.2083	0.3440	0.2164	0.0305
Val	0.1399	0.1319	0.2877	0.2366	0.0080
Rev	0.1732	0.1601	0.3165	0.4514	0.0132
GP	0.0556	0.0535	0.1826	0.1952	0.0021
Val_Prof	0.1408	0.1357	0.1598	0.2114	0.0051
Accruals	-0.0791	-0.0850	0.1199	0.3877	0.0060
Asset_Growth	0.1795	0.1714	0.2506	0.3928	0.0081
Invest	0.0183	0.0136	0.1969	0.3449	0.0047
Net_Issue	0.0343	0.0319	0.0939	0.1578	0.0024
Ret_BE	0.0469	0.0432	0.2096	0.2434	0.0036
Val_Rev	0.1656	0.1514	0.5879	0.4007	0.0142
Val_Rev_Prof	0.1901	0.1769	0.3385	0.3909	0.0132

Table 3.6: Summary Statistics under Annual Rebalance and 10-20% Buy-hold Spread

ized alpha as opposed to the annual MVE_{FF4} portfolio. The result shows controlling the rebalance frequency can effectively save the profitability of these anomalies despite of compromised Sharpe ratios when compared with the corresponding baseline FF4 strategy.

The second most effective methodology-based cost mitigation technique is buy-hold spread. The buy-hold spread is implemented according to the "sS" rule. In the long-side portfolio, investor buys a stock if it enters the top s%, but does not liquidate it until it drops out of the top S%, where s < S. Similarly on the short side, stocks in the bottom s% are shorted, and are only bought back when they leave the bottom S% range. The spread between s-S is called the buy-hold spread. This methodology is easy to followed and monotonically reduces the turnover rate. Table 3.6 gives the returns and turnover rates under annual rebalance and 10-20% buy-hold spread. Specifically, it reports the annual average gross return before costs, net return after costs from both spread and price impact estimates, the standard deviation of net return, turnover rates and transaction costs as the difference between the first two columns for each anomaly.

Anomaly	Generalized α	t-val
Net_Issue	0.0745	3.326
GP	0.1209	3.425
Asset_Growth	0.2094	3.312
Ret_BE	0.1673	4.700

Table 3.7: Anomalies with Significant α^* under Annual Rebalance and 10-20% Buy-hold Spread

Compared with annually rebalanced portfolio without buy-hold spread, turnover rates for all tested anomalies are further reduced by 10% to 25%. Annual turnover rates for all tested anomalies besides FF4 are reduced to 40% or lower, at the price of slightly compromised net returns. The Sharpe ratio for baseline FF4 model slightly drops by less than 10% from 0.75 to 0.7, but one more anomaly MVE portfolio formed by factor gross-profitability (GP) generates significant generalized alpha, along with net-issuance, asset-growth and return-on-BE portfolios under annual rebalance without buy-hold spread. The MVE portfolio with GP boosts the Sharpe ratio by nearly one half from the baseline model, with a generalized alpha of more than 0.12 and t-value over 3.4. A wider 10%-30% spread can further save anomaly Val_Prof, together with other four anomalies as shown below. Moreover, t-values for GP, net-issuance and asset-growth all exceed 3, making their generalized alpha significant even under stricter threshold. The result indicates that the marginal performance for the top and runner-up deciles do not differ significantly, and thus holding stocks in the runner-up deciles can effectively reduce the trading amount without hurting the portfolio performance. Any larger spread beyond 10%-30% does not save anomalies formed by investment or other combined-ranking methods any further.

Table 3.7 and Table 3.8 report all significantly profitable anomalies under annual rebalance and two different buy-hold spreads. In both tables, column 1 displays the anomaly names, column 2 shows the generalized α and the last column shows the t-value for generalized α.

Anomaly	Generalized α	t-val
Val_Prof	0.0436	2.254
Ret_BE	0.1356	2.968
GP	0.0984	3.496
Net_Issue	0.0748	3.877
Asset_Growth	0.1737	3.182

Table 3.8: Anomalies with Significant α^* under Annual Rebalance and 10-30% Buy-hold Spread

3.6 Maximum Capacity to Attract Capital

I am also interested in the maximum capacity of the MVE portfolio for each anomaly to attract new capital before the marginal investor find it unprofitable. Since only five out of nine examined anomalies can generate significant generalized alpha under annual rebalance with 10%-30% buy-hold spread, I will focus on the maximum capacities for these five anomaly MVE portfolios. As shown in Figure 3·2 and Table 3.9, the generalized alphas for anomaly portfolios decrease as the fund sizes and price impacts increase. Figure 3·2 shows the trajectories of t-values for five anomalies under different log fund size in billion dollars before they become insignificant. All anomaly strategies are constructed under annual rebalance and 10%-30% buy-hold spread. Anomalies with higher t-values are on average more durable to the price impact cost effect from increasing fund size. Table 3.9 shows the maximum capacity each anomaly strategy has to attract new capital before the marginal investor find the anomaly long-short portfolio unprofitable. The sample period covers 20 years from 2002 to 2021 and all strategies are annually rebalanced with 10%-30% buy-hold spread.

In general, anomalies that can generate more significant generalized alpha are more persistent, and more attractive to new capital as fund sizes grow. The MVE portfolio with anomaly net-issuance is the most attractive among all strategies, attracting more than 1000 billion capital before its α^* is eaten up by the price impact. Anomalies

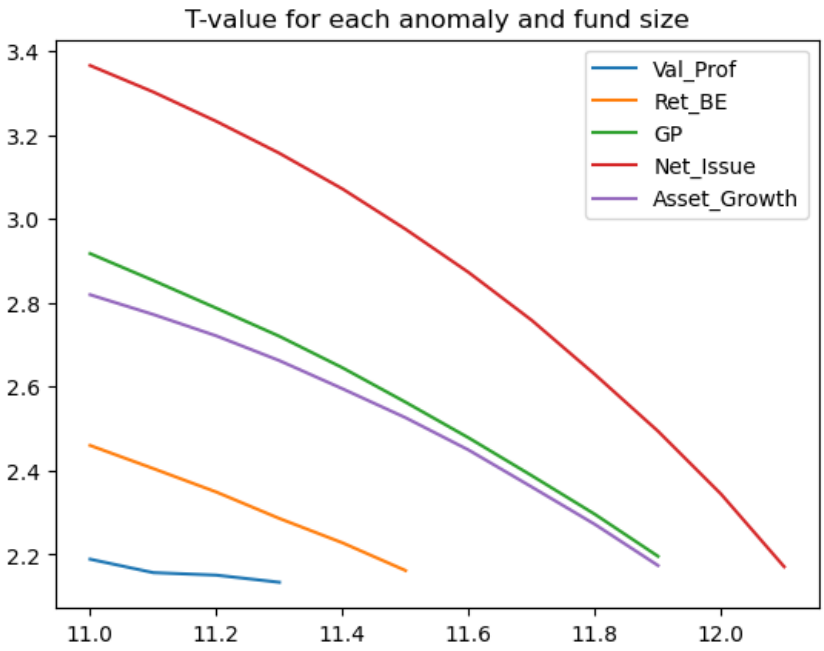

Figure 3·2: T-value of α^* for Each Significant Anomaly under Different Log Fund Size

	Val_Prof	Ret_BE	GP	Net_Issue	Asset_Growth
FS $B	199.53	316.23	794.33	1258.93	794.33

Table 3.9: Anomaly Strategy Capacities

GP and asset-growth stop generating generalized alphas before 800 billion dollars while the ranking anomaly Val_Prof is the least favorable among those five profitable anomalies, generating significant α^* only for funds less than 200 billion dollars. This is reasonable since even with 10-30% buy-hold spread, the MVE portfolio with Val_Prof only has a very slim α^* of 0.04, and quickly dies out as the fund size grows.

Chapter 4

Conclusions

4.1 Price Recovery Based on Recombinant Tree Framework

In this price recovery based on recombinant tree framework study, I introduce a new recombinant tree structure to recover physical asset price distributions and investor's risk preference. The tree framework allows an expanding price range across different horizons and sheds light on recovering probability for each sample path. For risk neutral investors, or more generally if marginal utilities are all given, we have closed form solutions for path probability. The closed form solution is only a function of state price observed from market data if tree structure is binomial, and needs functional form assumptions on the price process if more than two possible prices are allowed for each starting state. In general case where both probability and preference are unknown, the recovery system is solved period-by-period numerically and is capable of incorporating a wide range of price and utility assumptions.

I also numerically test the above recovery framework across different model specifications, parametric assumptions and data sources with recent 15 years SPX options data. The method consistently estimates return distributions with annualized mean between 6% to 8% and volatility around 14% across different specifications. Diffusion price models with jumps show higher volatility and more pronounced fat tails. Although recovery frameworks with time-series price processes usually contain less input information, they can serve as quick approximations when recovery speed is more emphasized. Machine learning models are flexible with different price training

rules and the recovery results are more accurate when the training rule has more predictive power on future return. The recovery framework is better adapted with monthly or longer forecast horizons. When applied to forecast weekly returns with daily expired SPX options data, the results are less negatively skewed and fat tailed due to the reduced ability to capture tail risks with short-term data. Tree structure also matters for the recovery results, with more granular partition in possible price ranges yielding more accurate recovery but also requiring more calibration efforts.

The recovery results are applied to check derivative market prices and some specific form of SDF. The possible pricing discrepancy from market prices and option prices implied by recovered distribution can lead to trading opportunities. The consistent profitability from portfolios constructed to capture such discrepancy shows the recovery framework provides more accurate forecast on return distributions than market consensus. This profitability is robust across different price and utility assumptions.

The recombinant tree recovery framework also has its limitations. The recovery results are sensitive to available expiries and strikes of recovered underlying asset on market. A more dispersed strike range usually results in a more dispersed set of price states at maturity. This dispersion on strikes then leads to a more dispersed recovered price distribution. Also, a good recovery result depends on parameters pre calibrated outside the recovery system. Future studies of such recovery system can be focused on improving the number of allowed parameters in the recovery system or methods to increase robustness of such system for less dependency on good parameter tuning.

4.2 Anomaly Profitability after TC with Price Impact

In chapter 3, this paper studies the profitability of nine anomalies in addition to Fama and French (1992) four factors after accounting for TC from spread and price impact combined. I find under quarterly rebalance and no cost mitigation methods, none

of the MVE portfolios constructed by tested anomaly and Fama and French (1992) factors combined can generate significantly positive generalized alpha relative to the baseline model. However, by using lower rebalancing frequencies and applying buy-hold spread, five anomalies become attractive since the turnover rate and transaction costs associated with it are effectively controlled from above, and the new MVE portfolio with anomaly largely improves the Fama and French (1992) MVE portfolio using the same construction rules. Between the two mitigation methods studied, lowering rebalancing frequency plays a more dominant role, reducing the turnover rate from baseline model by two thirds. On the other hand, a 10% buy-hold spread can further reduce the turnover rate by another 10% to 25%. I also find anomaly MVE portfolios with more significant generalized alpha are also more capable of attracting new capital before the profitability becomes insignificant, with the most persistent anomaly portfolio capable of attracting more than one trillion dollars before the marginal investor find it unattractive.

Appendix A

Appendix for Chapter 2

A.1 Proofs for Path Probability under RN Preference

A.1.1 Binomial Tree Structure

For price process follows a balanced recombinant binomial tree of level T as in Figure 2·1, let us label the current state (t, s_t) as $(0, 0)$ and for each subsequent period $0 < t \leq T$, s_t ranges from $-t$ to t. Higher state number corresponds to larger price up move. Then the top move state for each period t is also labeled as t, and it can only be reached from top state $(t-1)$ in the previous period. Thus, for top states $s = t$, forward path probability is given by spot state probabilities of the two consecutive maturities

$$p_{t,t+1}^{s,s+1} = \frac{q_0^{t+1}}{q_0^t} = \frac{1}{r_t^s}\left(r_{t+1}^{s+1} - 0\right), \tag{A.1}$$

which fits into

$$p_{t,t+1}^{s,s+1} = \frac{1}{r_t^s} \cdot \left(\sum_{j=0}^{\lfloor \frac{t-s}{2} \rfloor} r_{t+1}^{s+1+2j} - \sum_{j=1}^{\lfloor \frac{t-s}{2} \rfloor} r_t^{s+2j} \right). \tag{A.2}$$

For all other non top-mover edges, state price for terminal state $(t+1, s+1)$ can be expressed as the sum of state prices from starting state $(t, s+2)$ going down and (t, s) moving up:

$$r_{t+1}^{s+1} = r_t^{s+2}\left(1 - p_{t,t+1}^{s+2,s+3}\right) + r_t^s p_{t,t+1}^{s,s+1}. \tag{A.3}$$

Suppose all edges placed strictly upper than $p_{t,t+1}^{s,s+1}$ at period $t+1$ and all paths from previous periods $0 \leq t_0 < t$ follow the above formula (A.2) for any given starting

state s, plug in (A.2) for path $p_{t,t+1}^{s+2,s+3}$, we have

$$p_{t,t+1}^{s,s+1} = \frac{1}{r_t^s} \cdot \left(\sum_{j=-1}^{\lfloor \frac{t-s-2}{2} \rfloor} r_{t+1}^{s+3+2j} - \sum_{j=0}^{\lfloor \frac{t-s-2}{2} \rfloor} r_t^{s+2+2j} \right) \tag{A.4}$$

$$= \frac{1}{r_t^s} \cdot \left(\sum_{j=0}^{\lfloor \frac{t-s}{2} \rfloor} r_{t+1}^{s+1+2j} - \sum_{j=1}^{\lfloor \frac{t-s}{2} \rfloor} r_t^{s+2j} \right), \tag{A.5}$$

which also fits into (A.2). Starting from the first period and within each period, starting from top state to bottom, the mathematical induction gives the result.

A.1.2 Multinomial Tree Structure

If price process follows a multinomial tree structure of $(1 + 2n)$ edges from each node as in Figure 2·2, there are n possible outcomes in each price change direction per period. Following the same labeling as in binomial case, for each subsequent period $0 < t \leq T$, s_t ranges from $-nt$ to nt. Top mover state at period t is nt and can only come from top state $n\,(t-1)$ in previous period.

Assume probability of top mover path is $p = p_{t,t+1}^{s,s+n}$ and set it as the pivot path such that all other path probabilities starting from the same node can be expressed as a function of the pivot path

$$\begin{cases} p_{t,t+1}^{s,s+n-1} = f_{t,t+1}^{s,s+n-1}\,(p)\,, \\ \qquad\qquad \ldots \\ p_{t,t+1}^{s,s-n} = f_{t,t+1}^{s,s-n}\,(p)\,. \end{cases} \tag{A.6}$$

Then the forward pivot path probability is given by the state probabilities of two consecutive periods

$$p_{t,t+1}^{s,s+n} = \frac{q_0^{n(t+1)}}{q_0^t} = \frac{1}{r_t^s}\left(r_{t+1}^{s+n} - 0\right), \tag{A.7}$$

which fits into

$$p_{t,t+1}^{s,s+k} = \frac{1}{r_t^s} \cdot \left[r_{t+1}^{s+k} - \sum_{j=1}^{(k+n)\wedge(nt-s)} r_t^{s+j} \cdot f_{t+1}^{s+j,s+k} \left(p_{t+1}^{s+j,s+j+n} \right) \right], \tag{A.8}$$

for top mover $k = n$ and $s = nt$. Recall for each starting node $(s + j)$ and terminal node $(s + k)$, they can differ by at most n and starting state cannot be higher than the top state in its period, taking whatever is less, we have the upper bound in the second term by

$$\begin{cases} s + j \leq nt \text{ at top mover in level } t \\ s + j \leq s + k + n \text{ at bottom mover from starting state } (t, s + j). \end{cases} \tag{A.9}$$

For all other terminal states $(t + 1, s + k)$ with more than one starting states leading to it, its state price as the sum of state prices along all possible paths can be expressed as

$$r_{t+1}^{s+k} = r_t^s p_{t,t+1}^{s,s+k} + \sum_{j=1}^{(k+n)\wedge(nt-s)} r_t^{s+j} p_{t,t+1}^{s+j,s+k}, \tag{A.10}$$

where the first term represents path-wise state price from bottom lowest possible starting state s to $(s + k)$, and the second term represents all path-wise state prices starting from nodes placed above s. Suppose all path probabilities placed above $(s, s + k)$ at period $t + 1$ and all paths from previous periods $0 \leq t_0 < t$ follow the above formula, plug in (A.8) for all $p_{t,t+1}^{s+j,s+k}$ in the second term, we have the path probability $p_{t,t+1}^{s,s+k}$ also fits into the above (A.8). Similar to the binomial case, starting from upper left node in the tree, the mathematical induction gives the result.

In the case of given marginal utility, not necessarily RN, the proof for path probability formula has already been outlined in the main context. We only need to notice the r_t^s terms in RN path probability formula represents state probability. Substituting them with new state probability formula with known marginal utlity gives the result.

A.2 State Price Generator

A.2.1 Parametric Stochastic volatility State Price Generator

This appendix describe the Bates (2000) stochastic volatility jump diffusion method for state price and its implementation with fast Fourier transform (FFT). The process follows these high-level steps: the characteristic function (CF) for a stochastic jump-diffusion process is calculated, followed by applying the FFT to obtain call option prices. The model parameters are then calibrated by minimizing the sum of squared errors (SSE) between the model-generated option prices and observed market prices. After obtaining the optimal parameters, the model evaluates digital call option prices starting from each state, and the difference between these digital call prices is used to compute the risk-neutral state prices. The Bates model extends the Heston (1993) stochastic volatility model by adding jumps to stock prices. It is essentially a stochastic jump-diffusion process with both continuous volatility (from the Heston model) and discrete jumps (from Merton's jump model).

We first calculate the CF of the Bates model, which is the product of Heston CF and Jump CF:

$$\Psi_{Bates}(u) = \Psi_{Heston}(u) \cdot \Psi_{Jump}(u), \tag{A.11}$$

The jump part is governed by three parameters: jump intensity λ_J, average jump size μ_J and jump volatility σ_J. Then jump CF is given by

$$\Psi_{Jump}(u) = \exp\left\{-\lambda_J \mu_J iu\tau + \lambda_J \tau \left[(1+\mu_J)^{iu} \exp\left(\frac{1}{2}\sigma_j^2 iu(iu-1)\right) - 1\right]\right\}, \tag{A.12}$$

for given time to maturity τ. Heston's CF includes parameters mean reversion speed κ, long-term variance θ, volatility of volatility σ, initial variance v_0 and correlation between the stock and volatility ρ. The original Heston model can sometimes result in numerical issues, particularly in handling certain parametric values, due to the

exponential term in the characteristic function. To overcome this, the Little Heston Trap formulation was developed as a more stable method to compute the characteristic function. The core idea is to transform the characteristic function to a form that avoids the instability caused by potential overflow in exponential terms. In this formulation, the characteristic function is modified to ensure numerically stable results while retaining accuracy. The characteristic function of the Heston model, under this stable formulation, is expressed as

$$\Psi_{Heston}(u) = \exp\left(C + Dv_0 + iux\right), \tag{A.13}$$

where

$$C = (r-q)iuT + \frac{\kappa\theta}{\sigma^2}\left[(\kappa - \rho\sigma iu - d)T - 2\log\left(\frac{1-ge^{-dT}}{1-g}\right)\right], \tag{A.14}$$

$$D = \frac{\kappa - \rho\sigma iu - d}{\sigma^2}\left(\frac{1-e^{-dT}}{1-ge^{-dT}}\right), \tag{A.15}$$

$$d = \sqrt{(\rho\sigma iu - \kappa)^2 + \sigma^2(iu + u^2)}, \tag{A.16}$$

$$g = \frac{\kappa - \rho\sigma iu - d}{\kappa - \rho\sigma iu + d}. \tag{A.17}$$

This formulation, referred to as the "Little Heston Trap", ensures that even when certain parameter combinations would traditionally lead to instability, the function remains bounded and numerically stable. This makes it possible to accurately evaluate the characteristic function and thus derive option prices reliably.

After having Bates CF for a given set of parameters, the second step is to use Fast Fourier Transform (FFT) to compute call option prices based on the Bates CF. This is an efficient numerical method that simplifies integration in option pricing. Specifically, the call price $C(K)$ for strike price K is given by the inverse Fourier

transform of the characteristic function

$$C(K;\,\Theta) = \frac{e^{-r\tau}}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-iu\cdot\log(K/S_0)}}{\alpha^2+\alpha-u^2+iu\cdot(2\alpha+1)} \cdot \Psi_{Bates}(u;\,\Theta)du, \tag{A.18}$$

where α is a damping factor that helps with convergence, τ is time to maturity and S_0 is spot price of underlying asset at the recovery date. Θ is the set of Bates parameters used for option price calculation.

The third step is to calibrate for an optimal set of Bates parameters Θ^*. After having Bates call prices for a specific set of model parameters Θ, the parameter set Θ is calibrated by minimizing the sum of squared errors (SSE) between the Bates call option prices and those observed in the market

$$SSE\,(\Theta) = \sum_{i=1}^{N} \left[C_{Market}(K_i) - C_{Model}(K_i;\,\Theta)\right]^2, \tag{A.19}$$

and

$$\Theta^* = \arg\min_{\Theta} SSE\,(\Theta)\,. \tag{A.20}$$

The calibration process uses nonlinear optimization to find the parameter set that minimizes the SSE.

The final step is to get Bates state price associated the optimal parameters. Once the Bates model parameters are calibrated, we evaluate digital call option prices using the same characteristic function but slightly modified for digital options

$$C_{Digital}(K) = \frac{e^{-r\tau}}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-iu\cdot\log(K/S_0)}}{\alpha+iu} \cdot \Psi_{Bates}(u;\,\Theta^*)du. \tag{A.21}$$

Digital options pay a fixed amount if the stock price exceeds the strike price at maturity, which makes them useful for calculating state prices. The state price is

derived by taking the difference between the digital call prices at successive strikes

$$SP(K_i) = \max\left(0, C_{Digital}(K_{i+1}) - C_{Digital}(K_i)\right). \tag{A.22}$$

This gives the risk-neutral probability of the asset price being within a specific range (between two strike prices) at maturity.

The overall algorithm combines stochastic volatility (Heston model) and jump-diffusion (Merton's jump model) into a unified framework, i.e. Bates model, using Fourier transform techniques to calculate option prices and state prices. The model is calibrated to market data to ensure accuracy, making it highly effective in capturing both continuous and jump-related risks in financial markets. The use of the "Little Heston Trap" further enhances the stability and numerical reliability of the characteristic function computation.

A.2.2 Non-parametric Fast-and-Stable State Price Generator

This appendix describe the non-parametric Jackwerth (2004) fast-and-stable method to get state price from options data.Compared with parametric method described before, the fast-and-stable method does not assume functional form of the underlying price process, thus is free from any parameters calibration. The method is much faster in generating state prices than Bates method, but may subject to over smoothing when not enough options data are observed.

The first step in Jackwerth method involves fitting the missing or sparse implied volatility data. This is done by minimizing an objective function that balances smoothness of the volatility curve with the fit to observed implied volatilities. The objective function is numerically defined as

$$\frac{1}{2(1+I)} \sum_{k=1}^{I-2} \left(\sigma_{k+2} - 2\sigma_{k+1} + \sigma_k\right)^2 + \frac{\lambda}{2I} \sum_{k=1}^{I} \left(\sigma_k - \sigma_k^{obs}\right)^2, \tag{A.23}$$

where I is the number of observed implied volatilities, σ_k represents the interpolated or optimized implied volatilities, σ_k^{obs} are the observed implied volatilities. The first term controls smoothness, ensuring that the fitted volatility curve is smooth by penalizing large second derivatives of volatility with respect to strike prices. The second term represents goodness of fitting, weighted by the parameter λ, which adjusts the trade-off between smoothness and fit to observed data. The smoothness term minimizes the second derivative of the volatility curve, while the fit term ensures that the optimized volatility values are close to the observed ones. The tuning parameter λ controls the balance between fitting the observed data and ensuring smoothness. Higher values of λ prioritize data fitting over smoothness. In practice, we choose the highest λ, i.e. goodness of fitting, conditioned on non-negative recovered RN density and uni-modality within around moneyness strike range.

Once the implied volatility surface is fitted, the second step is to calculate theoretical European call option prices using the Black-Scholes formula

$$C(S_0, K, \tau, r, \sigma) = S_0 N(d_1) - K e^{-r\tau} N(d_2), \tag{A.24}$$

where

$$\begin{cases} d_1 = \dfrac{1}{\sigma\sqrt{\tau}} \cdot \left[\ln\left(\dfrac{S_0}{K}\right) + \left(r + \dfrac{1}{2}\sigma^2\right)\tau \right], \\ d_2 = d_1 - \sigma\sqrt{\tau}, \end{cases} \tag{A.25}$$

for spot price S_0 of underlying asset, strike K , time to maturity τ , risk free rate r and fitted implied volatility σ. The Black-Scholes formula provides a theoretical value for European call options, which is essential for the subsequent extraction of the risk-neutral density.

Once the theoretical call option prices are computed, the third step is to extract risk-neutral state density using the Breeden and Litzenberger (1978) formula. The Breeden and Litzenberger (1978) links second derivative of call option prices with

respect to strike to its risk-neutral density by

$$f(K) = e^{r\tau} \cdot \frac{\partial^2}{\partial K^2} C(K), \tag{A.26}$$

where $f(K)$ is the risk-neutral density and $C(K)$ is the call option price implied by Black-Scholes with fitted implied volatility. In practice, this is done numerically by taking the second derivative of the option prices with respect to strike prices using finite differences

$$f(K) \approx e^{r\tau} \cdot \frac{C(K+h) - 2C(K) + C(K-h)}{h^2} \tag{A.27}$$

for a small increment h in the strike price K. Finally, the state prices are obtained by multiplying the risk-neutral density and the discount factor derived from the risk-free rate

$$SP(K) = e^{-r\tau} f(K), \tag{A.28}$$

yielding the risk-neutral probability of the asset price centered at strike K for a given maturity τ.

This method combines the smooth interpolation of implied volatilities with the Breeden-Litzenberger approach to extract state prices, providing a robust framework for evaluating the risk-neutral market expectations of future asset prices. The algorithm ensures both accuracy in fitting observed data and smoothness of the resulting volatility surface, making it suitable for real-world option market applications.

A.3 Details on State Price Estimates and Parameters Tuning

This appendix describes details on data selection and parameter fine-tuning in both state price estimate and recovery processes.

A.3.1 Data Selection in State Price Estimates

For the baseline monthly state price estimates, we select SPX options data from September 2008 to August 2023. The edges expanding from each sub-tree is always 11 in the baseline model, i.e. 5 edges on each side. The recovery date are the last Friday of each month such that there exist at least one expiry falls into 24 to 35 days from the recovery date. If exists, this expiry is selected as the terminal level expiry for the tree, representing the monthly recovery horizon. If no such Friday in a given month, this month is skipped. We ends up with 150 months in the past 15 years and almost all skipped months are concentrated on the years before 2012. The level of trees for each selected recovery date is given by available standard Friday expiries at least 7 days from the recovery date and less than the terminal expiry. Thus, the level of tree for each recovery date can vary from 1 to 4, contingent on available weekly expiries in the following month. Starting from 2017, most recovery tree has full 4 levels.

For the weekly recovery, the logic is the same. We select all Fridays such that there is at least one expiry falls into 5 to 7 days from the recovery date, representing weekly forecast horizon. If exists, such expiry is decided as the expiry date for the terminal level of the tree. If there is no such expiry for a given week, the week is skipped. For ease of computation, we fix the maximum possible levels of each recovery tree to 3, even though theoretically, if all trading days in a week has expiry, the level can go up to 5. To select intermediate expiries for each level, we choose all dates that has at least 1 day to expiry to avoid tricky behaviour of super short term options. Indeed, super short-term SPX options, especially those close to expiry, can exhibit unusual price behavior due to heightened sensitivity to small changes in the underlying index, volatility, and time decay. With little time remaining, the option gamma can become extremely high, causing large swings in option prices from small movements in the

SPX. Additionally, the impact of bid-ask spreads can be more pronounced, leading to erratic price movements. As theta decay, which measures time value, accelerates, options can rapidly lose value, making prices appear unstable. Finally, liquidity issues near expiry can exacerbate price volatility and lead to inconsistent or "jumpy" pricing. To guarantee an approximately equal time interval between selected expiries, we select daily SPX expiries of either every Monday, Wednesday and Friday of the week, or if the Monday is a public holiday, every Tuesday, Thursday and Friday of the week. Since CBOE introduces daily expiry SPX options in 2016, most weekly expiry data are estimated with data in recent 7 years. We end up with 214 weekly recovery date and most of the recovery trees have full 3 levels. More levels on the tree give a finer partition in the terminal price range, thus more accurate discrete density estimates. As we can see in the empirical results part, recovery results with more possible terminal price states have less standard errors than those with limited levels or only 3 edges expanding from each node.

Since OptionMetrics dataset does not report the SPX forward price of corresponding maturity as such products may not exist, we need to calculate the call-delta-implied forward price from the strike that has delta closest to 0.5, rather than directly borrowing spot index level as a proxy for forward price. The observation that the SPX spot level often has a call delta less than 0.5 suggests the current spot price is not the true ATM level, especially when considering the impact of the dividend yield and the risk-free accumulation on the option's pricing. Using the forward price as the ATM level offers a more accurate reference point for options pricing. This approach aligns the delta of the call option closer to 0.5, improving the precision in measuring implied volatility and ensuring that the ATM option is truly at the midpoint between the probabilities of the option expiring in-the-money or out-of-the-money. The forward price, therefore, provides a more balanced and theoretically consistent basis for

option pricing, hedging, and volatility analysis.

For Bates (2000) state price estimator, the method only uses call options data as they are needed minimize the SSE from fast Fourier transform (FFT) estimated call option prices. The damping factor needed for FFT is set to $\alpha = 1.2$. Setting the damping factor slightly higher than 1 (which is the rule-of-thumb calibration for damping factor) improves the convergence of the FFT, reduces truncation errors and enhances numerical stability. By exponentially damping the payoff, it helps manage oscillatory behavior at large strikes and minimizes errors when integrating over a finite interval.

For Jackwerth (2004) state price estimator, it uses both out-of-the-money call and put data. In practice, we only use options with moneyness, i.e. estimated forward price over strike, falls into 0.9 to 1.1 to avoid errors from limited liquidity and over-inflated implied volatility for deep out-out-of-the-money options. Missing values for implied volatility are filled in with interpolation from adjacent strikes and extrapolate at two ends. Recall the tuning parameter λ in this fast-and-stable method controls goodness of fitting. Higher λ gives more accurate fitted implied volatility while enduring the risk of jagged volatility surface. In practice, we start from a high level of λ and calculate the corresponding implied volatility surface. The implied state probability is then given by Black-Scholes call price and Breeden-Litzenberger formula. We stop calibrating on the highest λ whenever the state probability is positive and unimodal.

A.3.2 Parameter calibrations for Recovery System

This section describes details in parameter pre calibrations for each price model. As we will see, a reasonable calibration for exogenous parameters not decided by the recovery system is crucial to final recovery results.

For all diffusion price models, we need to first calibrate diffusion parameter σ be-

fore expressing path probability as a function of only drift parameter μ. To account for demand pressure and liquidity shocks that inflate the implied volatility from options data, we cap the implied volatility by 0.6, which approximately represents the return volatility of SPX observed during crisis period. For jump parameters in Poisson and double exponential jump models, missing values of jump intensity is filled with 0. This reduces jump models for recovery dates with not enough intraday trade-level data to continuous GBM model. However, such data missing case is very rare across the tested period. The jump rate is also clipped to be greater than 0.5 times per year and less than 2 times per year, with the exact jump rate parameter prorated by the time to maturity of studied period. The multiplicative upward and downward jump size are calibrated by fraction of institutional large order numbers and is constraint to $(1, 1.05)$ for upward size and $(0.9, 1)$ for downward size, respectively. The missing values are filled with 1.025 for upward size and 0.95 for downward size. The calibration caps are consistent with average jump size implied by real data. The downward jump size is usually from -5% to -10%, even though at extreme cases like crisis, it can be -20%. The upward jump size is usually smaller in scale compared with downward size, and mostly falls into 2% to 5%. Our calibration reflects the fact that downward jumps are usually more pronounced due to fear sentiments. The caps also prevents extreme and unrealistic jump size implied by intraday data as it is only a proxy of the jump parameters. The upward jump probability, conditioned on there is a jump, is estimated by institutional buy order numbers over total institutional trade number from intraday data, and is cut off for the range $(0.4, 0.6)$, and missing values are filled with 0.5, reflecting equal chance on both sides. This estimate is for each individual market shock. From time series point of view, the conditional probability of an upward jump is typically estimated to be around 20% to 40%, with the remainder being downward jumps. This implies that when a jump happens, it is more likely to

be downward, reflecting the historical pattern of sharper market declines compared to rises. The asymmetry between upward and downward jump probabilities is consistent with the fact that markets often experience sudden sell-offs more frequently than sudden surges. Our jump parameter estimates have the same pattern over the tested horizon, where the upward jump number is less than the downward jump.

For time series models, in standard ARCH(1), we estimate the squared previous return weight parameter α in conditonal volatility from recovery system and assume there is no drift term for return process

$$r_t = \sigma_{t|t-1} \cdot \epsilon_t. \tag{A.29}$$

More generally, when the recovered return is consistently deviated from reasonable range, one fix is to add a drift term in the above return process. It is reasonable since the standard model only accounts for the RN measure, while we are actually interested in the physical return process with exogenous drift. For our recovery, in both monthly and weekly horizons, the recovered return matches observations from real index data, and we can set the pre calibrated drift parameter to be 0 in both cases. For GARCH(1,1) model, the only difference is from structure of conditional volatility which incorporates one-period lagged conditonal volatility in addition to past squared return, and we follow the same logic to set return drift to be 0 for both monthly and weekly recovery. However, we can keep in mind that for future applications to other expiries, we can tune the drift parameter whenever the recovered return has a parallel bias from real data across different recovery date.

For multinomial logistic regression (MLR) price models, the most important factor is training rule used to categorize one-period return. Such training rule can be very broad from different anomaly factors to any quantity that may have predictive power on future return. As an illustration, I trained two models separately with past one-

period return (MLR) and past one-period volatility (MLR2), respectively. For the return based MLR model, we train the model at each state with daily return data for most recent 252 trading days. Note the training data for different states on the recovery tree are only different for estimated unrealized returns after the recovery date and before the expiry corresponding to the state. Since the tree only have limited 1 to 4 layers, the difference in training dataset for each node is marginal compared with all daily data from past 1-year time. The n-outcome MLR model is trained as $(n-1)$ separate binomial logistic regressions with a common pivot outcome, i.e. one-vs-rest method. Then all probabilities are normalized to unit sum based on their probabilistic relations to the pivot outcome. Each individual binomial logistic regression is trained by coordinate descent method that optimizes the logistic loss function for logistic regression. In this approach, the optimization problem is decomposed into a series of one-dimensional sub-problems by iteratively updating each parameter while keeping the others fixed. Mathematically, for a logistic regression model, the objective is to minimize the regularized logistic loss function

$$L(\beta) = \sum_{i=1}^{N} \log\left(1 + \exp(-y_i(\beta_0 + \boldsymbol{x}_i^T \boldsymbol{\beta}))\right) + \frac{\lambda}{2}\|\boldsymbol{\beta}\|^2, \tag{A.30}$$

where $\boldsymbol{\beta}$ represents the model coefficients, $\boldsymbol{x}_i$ are the feature vectors, y_i are the target labels and λ is the regularization parameter. Coordinate descent iteratively updates each component of $\boldsymbol{\beta}$, solving the one-dimensional optimization problem for each coefficient. This method leverages the sparsity and separability of the problem, making it efficient for binary and one-vs-rest classification tasks.

For our recovery use case, we only train the first $(n-2)$ separate binomial logistic regressions for an MLR with n outcomes before normalization. Set the probability of first pivot path as p, then each of the rest $(n-1)$ path probabilities can be expressed as a ratio of p. The recovery system takes p as unknown parameter and solve it

together with marginal utility.

For the MLR2 model where data is trained by past one-period lagged return volatility, I estimate the return volatility by two different datasets. The first calibration of daily return volatility comes from TAQ trade-level data of SPY. Since the TAQ intraday volatility data measures trade based return variance of asset i on date t

$$IVAR_{i,t} = \frac{1}{t-1}\sum_{\tau=1}^{t}\left(Ret_{i,\tau} - \bar{Ret}_i\right)^2, \tag{A.31}$$

where $Ret_{i,t} = \ln \frac{P_{k,\tau}}{P_{k-1,\tau-1}}$ for k-th trade price $P_{k,\tau}$ at time τ, and $\bar{Ret}_i = \frac{1}{t-1}\sum_{\tau=1}^{t} Ret_{i,\tau}$, we calculate annualized volatility estimated from trade-level data as

$$\sigma_{i,t}^{intraday} = \sqrt{252 \cdot IVAR_{i,t} \cdot n_{i,t}}, \tag{A.32}$$

where $n_{i,t}$ is total number of trades on asset i at date t. Then all calculated intraday volatility are winsorized at 1% and 99% to exclude extreme or unrealistic values. The second calibration of daily return volatility is using closing VIX level at previous trading day. Theoretically, VIX represents the market's expectations for volatility over the next 30 days, derived from the implied volatilities of S&P 500 options. Although VIX does not directly measure the daily return volatility, it is a market wide index on return uncertainty. Using the previous close of the VIX index as a proxy for the daily return volatility of the SPX is a common practice because it can serve as a reasonable estimate of the anticipated daily volatility, as it reflects the collective sentiment and risk assessment of market participants up to the previous day's close. Although it may not capture intraday changes or sudden shifts in sentiment, it provides a consistent and readily available measure that aligns closely with observed SPX volatility patterns.

Since the recovery system of fractional state price equations does not have closed form solutions, we need to solve it numerically. In practice, for the set of solutions

(p, v) and recovery system $r_s = f(p, v)$ for $s = 1, ..., N-1$, define the objective function as

$$obj(p, v) = \sum_{i=1}^{N-1} [r_s - f(p, v)]^2 + L, \tag{A.33}$$

where the first term measures sum of squared error from fitting and the second term penalizes recovered daily absolute return that are greater than 2%. The penalty is set to be a linear function of the absolute distance between recovered return with a specific set of parameter with 2%. The penalty is added to regulate the unrealistic recovered return due to parameter calibrations, and solution (p, v) is the minimizer of the above objective function. The optimization problem is solved numerically by Sequential Least Squares Programming (SLSQP) method. SLSQP algorithm is a gradient based optimization technique used to solve minimization problems iteratively. It approximates the original nonlinear optimization problem by solving a sequence of quadratic programming (QP) sub-problems, where each sub-problem minimizes a quadratic approximation of the objective function subject to linear constraints. Mathematically, at each iteration k, the objective function is approximated by a quadratic model of the form

$$\min_{\Delta x} \quad \frac{1}{2}\Delta x^T H_k \Delta x + \nabla f(x_k)^T \Delta x, \tag{A.34}$$

where H_k is an approximation of the Hessian matrix of the Lagrangian, $\nabla f(x_k)$ is the gradient of the objective function evaluated at the current iteration x_k, and Δx represents the step direction. The solution to this sub-problem provides a search direction that is used to update the parameters iteratively. The algorithm efficiently utilizes the Jacobian gradient of the objective function, which helps in finding the search direction more accurately and improves the convergence rate by providing information on the slope of the objective function. The use of the gradient allows SLSQP to adjust the step size and direction effectively, especially near the optimum where the objective function's curvature plays a crucial role. We set the iteration to

stop when changes in the objective function is less than 0.001 or number of iterations exceeds 20 to avoid pitfalls of repetitive iterations without significant changes in objective function.

Appendix B

Appendix for Chapter 3

B.1 Proof of Conjugate Posteriors and Choice for Parameter Values

For normal Bayesian regression model

$$y_t = \beta_m r_t^m + \gamma d_t + \alpha(s_t d_t) + u_t, \tag{B.1}$$

where $u_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2)$ with unknown σ_u^2. With fixed data $\{r_m, d, s\}_t$, normal priors of coefficients $\{\beta_m, \gamma, \alpha\}$ and inverted Gamma prior for σ_u^2 gives conjugate distributions for the posteriors.

To see this, for parameter σ, prior $\sigma_u^{-2} \sim \Gamma(\alpha, \beta)$ and normal likelihood

$$p(\underset{\sim}{y}|\theta, \underset{\sim}{s}) = \left(\frac{1}{\sqrt{2\pi}\sigma_u}\right)^n \cdot e^{\frac{-1}{2\sigma_u^2}\Sigma_t u_t^2(y_t)}, \tag{B.2}$$

where

$$u_t = y_t - \beta_m r_t^m - \gamma d_t - \alpha s_t d_t, \tag{B.3}$$

and θ stands for all parameters to be estimated gives posterior

$$\begin{aligned} &p(\sigma_u^{-2}|\underset{\sim}{y}, \underset{\sim}{s}, \beta_m, \gamma, \alpha)_{\sigma_u^{-2}} \\ &\propto \left(\sigma_u^{-2}\right)^{\alpha+\frac{n}{2}-1} \cdot e^{-\sigma_u^{-2}\left(\beta^{-1}+\frac{1}{2}\Sigma_t u_t^2\right)}, \end{aligned} \tag{B.4}$$

i.e.

$$\sigma_u^{-2}|\underset{\sim}{y}, \underset{\sim}{s}, \beta_m, \gamma, \alpha \sim Gamma\,(\alpha + \frac{n}{2}, \frac{1}{\beta^{-1} + \frac{1}{2}\Sigma_t u_t^2}), \tag{B.5}$$

for n observations.

For s_t, the prior is $s_t \sim N\,(\mu_s, \sigma_s^2)$, with joint normal likelihood, we have the posterior is

$$\begin{aligned}
&p(s_t|\underset{\sim}{y}, \underset{\sim}{s_{-t}}, \beta_m, \gamma, \alpha, \sigma) \\
&\underset{s_t}{\propto} \exp\left[\frac{-1}{2\sigma_s^2}(s-\mu_s)^2 - \frac{1}{2\sigma_u^2}\Sigma_t\left(y_t - \beta_m r_t^m - \alpha s_t d_t - \gamma d_t\right)^2\right] \\
&\underset{s_t}{\propto} \exp\Big(\frac{-1}{2}\Big\{\left(\sigma_s^{-2} + \alpha^2\sigma_u^{-2}\right)s_t^2 - 2\big[\,\sigma_s^{-2}\mu_s + \sigma_u^{-2}\alpha y_t d_t \\
&\quad - \sigma_u^{-2}\alpha d_t\left(\beta_m r_t^m + \gamma d_t\right)\big]\cdot s_t\Big\}\Big),
\end{aligned} \tag{B.6}$$

i.e.

$$\begin{aligned}
&s_t|\underset{\sim}{y}, \underset{\sim}{s_{-t}}, \beta_m, \gamma, \alpha, \sigma \\
&\sim N\left(\frac{\sigma_s^{-2}\mu_s + \sigma_u^{-2}\alpha d_t\left(y_t - \beta_m r_t^m - \gamma d_t\right)}{\sigma_s^{-2} + \alpha^2\sigma_u^{-2}}, \frac{1}{\sigma_s^{-2} + \alpha^2\sigma_u^{-2}}\right).
\end{aligned} \tag{B.7}$$

Similarly for β_m, α and γ, under joint normal likelihood, we have for prior $\beta_m \sim N\,(\mu_\beta, \sigma_\beta^2)$, the posterior is given by

$$\begin{aligned}
&\beta|\underset{\sim}{y}, \underset{\sim}{s_t}, \gamma, \alpha, \sigma \\
&\sim N\left(\frac{\sigma_\beta^{-2}\mu_\beta + \sigma_u^{-2}\Sigma_t r_t^m\left(y_t - \gamma d_t - \alpha s_t d_t\right)}{\sigma_\beta^{-2} + \sigma_u^{-2}\Sigma_t (r_t^m)^2}, \frac{1}{\sigma_\beta^{-2} + \sigma_u^{-2}\Sigma_t (r_t^m)^2}\right);
\end{aligned} \tag{B.8}$$

for prior $\alpha \sim N\,(\mu_\alpha, \sigma_\alpha^2)$, the posterior is given by

$$\begin{aligned}
&\alpha|\underset{\sim}{y}, \underset{\sim}{s_t}, \gamma, \beta, \sigma \\
&\sim N\left(\frac{\sigma_\alpha^{-2}\mu_\alpha + \sigma_u^{-2}\Sigma_t s_t d_t\left(y_t - \beta_m r_t^m - \gamma d_t\right)}{\sigma_\alpha^{-2} + \sigma_u^{-2}\Sigma_t s_t^2}, \frac{1}{\sigma_\alpha^{-2} + \sigma_u^{-2}\Sigma_t s_t^2}\right);
\end{aligned} \tag{B.9}$$

and for prior $\gamma \sim N(\mu_\gamma, \sigma_\gamma^2)$, the posterior is given by

$$\begin{aligned} &\gamma | \underset{\sim}{y}, \underset{\sim}{s_t}, \alpha, \beta, \sigma \\ &\sim N\left(\frac{\sigma_\gamma^{-2}\mu_\gamma + \sigma_u^{-2}\Sigma_t d_t \left(y_t - \beta_m r_t^m - \alpha s_t d_t\right)}{\sigma_\gamma^{-2} + n\sigma_u^{-2}}, \frac{1}{\sigma_\gamma^{-2} + n\sigma_u^{-2}} \right). \end{aligned} \tag{B.10}$$

Reasonable choice of prior parameters can positively affect the convergent speed. To see this, for σ_u^{-2}, posterior mean

$$\alpha^*\beta^* = \left(\alpha + \frac{n}{2}\right) \cdot \frac{1}{\beta^{-1} + \frac{1}{2}\Sigma_t u_t^2}, \tag{B.11}$$

where

$$u_t = y_t - \beta_m r_t^m - \gamma d_t - \alpha s_t d_t. \tag{B.12}$$

Note data $\Sigma_t u_t^2$ is usually to the order of 5000, and number of observations is around 250 for each year, we have

$$\begin{aligned} \alpha^* &= \alpha + \frac{n}{2} \geq 125, \\ \beta^* &\leq 4 \times 10^{-4}, \end{aligned} \tag{B.13}$$

and σ_u^{-2} is usually very concentrated around the mean $\alpha^*\beta^*$.

For parameter α, posterior mean is

$$\frac{\sigma_\alpha^{-2}\mu_\alpha + \sigma_u^{-2}\Sigma_t s_t d_t \left(y_t - \beta_m r_t^m - \gamma d_t\right)}{\sigma_\alpha^{-2} + \sigma_u^{-2}\Sigma_t s_t^2}. \tag{B.14}$$

Note prior parameters of α has no cross term with data, and both data terms are rescaled by inverted innovation variance σ_u^{-2}, we notice the prior choice of σ_u^{-2} matters. Specifically, if σ_u^{-2} is too small, the data effects are overwhelmed by the priors and the post mean will be around $\frac{\sigma_\alpha^{-2} \cdot \mu_\alpha}{\sigma_\alpha^{-2}}$, equaling the prior mean μ_α.

To get the concavity of price impact effect, set exponential parameter $\alpha \in (0, 1)$. A reasonable prior is

$$\alpha \sim N(\frac{1}{2}, \frac{1}{36}), \tag{B.15}$$

such that under 3σ rule, $\alpha \in (0, 1)$. Earlier literature indicate $\alpha \approx \frac{1}{2}$, also supporting this choice. Besides, notice both data terms

$$\Sigma_t d_t s_t \left(y_t - \beta_m r_t^m - \gamma d_t\right), \tag{B.16}$$

and $\Sigma_t s_t^2$ are of the order 50,000, to unmute the data effect, we need

$$\sigma_u^{-2} \times 50000 \geq 10, \tag{B.17}$$

i.e.

$$\sigma_u^{-2} \geq 2 \times 10^{-4}. \tag{B.18}$$

Otherwise, the prior will dominate the data term, causing the posterior to be overwhelmingly affected by the potentially biased prior.

For parameter γ, the posterior mean is

$$\frac{\sigma_\gamma^{-2}\mu_\gamma + \sigma_u^{-2}\Sigma_t d_t \left(y_t - \beta_m r_t^m - \alpha s_t d_t\right)}{\sigma_\gamma^{-2} + n\sigma_u^{-2}} \tag{B.19}$$

Similarly, terms involving prior parameters and data are separated and are rescaled by σ_u^{-2}. A reasonable prior is

$$\gamma \sim N(-12, 4) \cdot \mathbf{1}_{[-14,-10]}. \tag{B.20}$$

Earlier literature suggest

$$\lambda = e^\gamma \approx 5 \times 10^{-6}, \tag{B.21}$$

also supporting this prior. Since the data term

$$\Sigma_t d_t \left(y_t - \beta_m r_t^m - \alpha s_t d_t\right) \tag{B.22}$$

is of the order 500, to unmute the data, we need

$$\sigma_u^{-2} \times 500 \geq 1. \tag{B.23}$$

Recall on average,

$$\sigma_u^{-2} = \alpha^* \beta^* \approx (\alpha + 125) \cdot \frac{1}{\beta^{-1} + \frac{1}{2}\Sigma_t u_t}, \tag{B.24}$$

and set $\alpha \ll 125$, e.g. 2, we have

$$(2 + 125) \cdot \frac{1}{\beta^{-1} + \frac{1}{2}\Sigma_t u_t} \geq 2 \times 10^{-3}. \tag{B.25}$$

Thus,

$$\beta \geq 1.6 \times 10^{-5}. \tag{B.26}$$

As a consequence, a reasonable choice of prior is

$$\sigma_u^{-2} \sim \Gamma\left(2, 2 \times 10^{-5}\right). \tag{B.27}$$

To get an idea of the level of s_t, note two data terms in the posterior mean of alpha

$$\sigma_u^{-2} \cdot \Sigma_t d_t s_t \left(y_t - \beta_m r_t^m - \gamma d_t\right) \tag{B.28}$$

and $\Sigma_t s_t^2$ should be of the same order which indicates γ and s_t should also be of the same order. To let the data speak,

$$\sigma_s^{-2} \leq \alpha^2 \sigma_u^{-2} \approx 6 \times 10^{-4}. \tag{B.29}$$

Thus, we choose σ_s to be 30 in this study. Under similar considerations, set prior for β to be

$$N(0.75, 100) \cdot \mathbf{1}_{[-3,5]}. \tag{B.30}$$

The procedure to choose the prior parameters for estimating Breen et al. (2002) model

$$\frac{\Delta p_t}{p_{t-1}} = \gamma \cdot \left(\frac{q_t}{Q_t}\right)^{\alpha} \tag{B.31}$$

are the same.

B.2 Net-of-cost Return Calculation

To calculate the net-of-cost return for each anomaly's value-weighted long-short portfolio, first notice the signed weight for each individual stock is given by

$$w_t = \frac{ME}{\Sigma_i ME_i}, \tag{B.32}$$

where ME_i stands for the market equity for each stock i, positive weights for the long side stocks and negative weights for the short side. Then the change of weight is given by

$$\Delta w_t = w_t - w_{t-1}\left(1 + r^g_{t-1}\right), \tag{B.33}$$

where r^g_t stands for the gross return of period t, the weight is for unit dollar investment and $\Delta w_t > 0$ means buying in more stocks, $\Delta w_t < 0$ means selling the corresponding stock in the rebalance. The transaction volume is

$$q_t = FS_t \cdot |\Delta w_t|, \tag{B.34}$$

where FS_t stands for the fund size. Now we have TC per dollar investment is

$$TC = \frac{1}{p}\left(\frac{s}{2} + \lambda q_t^{\alpha}\right), \tag{B.35}$$

and the liquidation cost ranked out of the target region is

$$LqC_t = |w_{t-1}\left(1 + r^g_{t-1}\right)| \cdot TC. \tag{B.36}$$

Equipped with cost representations, we have net-of-cost return for the short side and long side is

$$\begin{aligned} r^{n,s}_t = &\left[\Sigma_{i\in S}\left(w_{i,t}r^{g,s}_{i,t}\right) - \Sigma_{i\in S}|\Delta w_{i,t}| \cdot TC_{i,t} - LqC^s_{i,t}\right] \\ &\times \frac{1}{1 + \Sigma_{i\in S}|\Delta w_{i,t}| \cdot TC_{i,t} + LqC^s_{i,t}}, \end{aligned} \tag{B.37}$$

and

$$\begin{aligned} r^{n,l}_t = &\left[\Sigma_{i\in L}\left(w_{i,t}r^{g,l}_{i,t}\right) - \Sigma_{i\in L}|\Delta w_{i,t}| \cdot TC_{i,t} - LqC^l_{i,t}\right] \\ &\times \frac{1}{1 + \Sigma_{i\in L}|\Delta w_{i,t}| \cdot TC_{i,t} + LqC^l_{i,t}}, \end{aligned} \tag{B.38}$$

respectively. Set fund size re-scale coefficient for the long/short portfolio to be

$$res^l_t = \frac{1}{1 + \Sigma_{i\in L}|\Delta w_{i,t}| \cdot TC_{i,t} + LqC^l_{i,t}} \tag{B.39}$$

and

$$res^s_t = \frac{1}{1 + \Sigma_{i\in S}|\Delta w_{i,t}| \cdot TC_{i,t} + LqC^s_{i,t}}, \tag{B.40}$$

respectively, where negative weight stands for short selling the stock, then the final step is to re-scale the fund size by

$$FS_t = FS_{t-1} \times res^s_{t-1} \times res^l_{t-1} \times \left(1 + r^{n,l}_{t-1} + r^{n,s}_{t-1}\right). \tag{B.41}$$

References

Almgren, R., Thum, C., Hauptmann, E., and Li, H. (2005). Direct estimation of equity market impact. *Risk*, 18(7):58–62.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121:823–866.

Bates, D. S. (2000). Post-'87 crash fears in the s&p 500 futures option market. *Journal of Econometrics*, 94:181–238.

Bertsimas, D. and Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, 1:1–50.

Borovička, J., Hansen, L. P., and Scheinkman, J. A. (2016). Misspecified recovery. *The Journal of Finance*, 71(6):2493–2544.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–651.

Breen, W. J., Hodrick, L. S., and Korajczyk, R. A. (2002). Predicting equity liquidity. *Management Science*, 48:470–483.

Carr, P. and Yu, J. (2012). Risk, return, and ross recovery. *The Journal of Derivatives*, 20:38–59.

Chen, L., Novy-Marx, R., and Zhang, L. (2011). An alternative three-factor model. `http://dx.doi.org/10.2139/ssrn.1418117`. [Online; accessed 9-June-2011].

Cooper, M. J., Gulen, H., and Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *Journal of Finance*, 63:1609–1651.

Engle, R., Ferstenberg, R., and Russell, J. (2012). Measuring and modeling execution cost and risk. *The Journal of Portfolio Management*, 38(2):14–28.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 42(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.

Fama, E. F. and French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, 63:1653–1678.

Geweke, J. F. (2005). *Contemporary Bayesian econometrics and statistics*. John Wiley & Sons.

Glosten, L. R. and Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21:123–142.

Harris, L. E. (1990). Statistical properties of the roll serial covariance bid/ask spread estimator. *Journal of Finance*, 45:579–590.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29:5–68.

Hasbrouck, J. (1991). The summary informativeness of stock trades: An econometric analysis. *The Journal of Finance*, 4:571–595.

Hasbrouck, J. (2009). Trading costs and returns for u.s. equities: estimating effective costs from daily data. *Journal of Finance*, 64:1446–1477.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33:2019–2133.

Jackwerth, J. C. (2004). *Option-Implied Risk-Neutral Distributions and Risk Aversion*. Research Foundation of AIMR.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48:65–91.

Jensen, C. S., Lando, D., and Pedersen, L. H. (2019). Generalized recovery. *Journal of Financial Economics*, 133(1):154–174.

Korajczyk, R. A. and Sadka, R. (2004). Are momentum profits robust to trading costs? *The Journal of Finance*, 59:1039–1082.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53:1315–1335.

Lyandres, E., Sun, L., and Zhang, L. (2008). Investment-based underperformance following seasoned equity offerings. *Review of Financial Studies*, 21:2825–2855.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–31.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA.

Miao, J. (2014). *Economic Dynamics in Discrete Time*. The MIT Press, Cambridge and London.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108:1–28.

Novy-Marx, R. (2015). Backtesting strategies based on multiple signals. `https://www.nber.org/system/files/working_papers/w21329/w21329.pdf`. [Online; accessed July-2015].

Novy-Marx, R. and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29:104–147.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39:1127–1139.

Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2):615–648.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, 71:289–315.

Weitzman, M. L. (2007). Subjective expectations and asset-return puzzles. *American Economic Review*, 97:1102–1130.

CURRICULUM VITAE