This dissertation focuses on developing new statistical methods for analyzing and modeling financial time series. The first part of this dissertation discusses modeling of functional and distributional time series, assuming the series are driven by a finite dimensional underlying feature process. Functional time series are commonly observed in finance and difficult to model mainly due to high dimensionality. We instead focus on a low-dimensional latent feature process and connect it with the original functional time series through generalized state-space models. A state-space model assumes that observations are driven by an underlying dynamic state process and is widely used in many fields. We propose a generalized two-stage Sequential Monte Carlo (SMC) joint estimation framework to model functional time series driven by its feature process through state-space models and perform on-line estimations and predictions. In order to improve computation efficiency, we also implement parallel computing and re-design computation algorithms to integrate with non-linear optimization and SMC calculations.

Two financial applications are presented to demonstrate the robustness and efficiency of our proposed framework. The first application aims to extract and model
the daily implied risk neutral densities from observed call option prices. We view the underlying risk neutral density as a functional time series driven by its feature process and model it with a parametric mixed log-normal distribution through a state-space model. We conduct both simulation and empirical studies and compare prediction performance of our models with that of random walk models. Empirically, the proposed models improves prediction performance significantly. The second financial application studies daily cross-sectional distribution of 1000 largest market capitalization stock returns from year 1991 to year 2002. Similarly we view cross-sectional distribution as a functional time series driven by its feature process and model it with a four-parameter generalized skewed \(t\)-distribution. Using proposed two-stage SMC joint estimation framework, we build models separately for different market conditions, including the dot-com crisis. In both bearish and bullish markets, prediction performances of our models gain substantial improvement comparing with random walk models.

The second part of dissertation presents a new portfolio optimization strategy for minimum variance portfolios with constraints on short-sale and transaction costs. Unlike traditional mean-variance theory, our method minimizes only the portfolio risk and uses analysts’ consensus ratings to pre-screen stocks. An empirical study of S&P 500 stocks from year 1990 to year 2009 is conducted to demonstrate effectiveness. We show that portfolios constructed and optimized using our strategy deliver considerable improvement of performance in terms of Sharpe ratio comparing with benchmark portfolios and portfolios in past literatures.
Acknowledgements

First of all, I would like to thank Dr. Rong Chen for his great guidance as my advisor. I am deeply indebted to him for all he has done for me and will remember forever in my life. He not only taught me fundamentals of conducting scientific research but also set an excellent role model for me. For various reasons, there were several difficult times in my graduate study that made me think of giving up. Dr. Chen always trusted and supported me unconditionally even when I stopped believing in myself. Without his kind understandings and generosity, I am not able to gather enough patience and persistence to continue my study and would not stand a chance of finishing this dissertation here. I have always been honored to be his student both for the past and in the future.

Second, I would like to thank my committee members who helped and encouraged me.

Last but not least, I would like to express my deepest gratitude to my family. This dissertation would not have been done without their endless love, continued patience, and unconditional support.
Dedication

This dissertation is wholeheartedly dedicated to my beloved parents, Huayin Yan and Hongwen Sun, my wife, Lili Xia, and my son, Luke, who have always loved, supported, and believed in me.
## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Preliminary</td>
<td>4</td>
</tr>
<tr>
<td>2.1. Functional Time Series and State-Space Model</td>
<td>4</td>
</tr>
<tr>
<td>2.1.1. Functional Time Series Driven by Its Feature Process</td>
<td>4</td>
</tr>
<tr>
<td>2.1.2. State-Space Model</td>
<td>5</td>
</tr>
<tr>
<td>2.1.3. Kalman Filter</td>
<td>7</td>
</tr>
<tr>
<td>2.1.4. Sequential Monte Carlo Framework</td>
<td>7</td>
</tr>
<tr>
<td>2.1.4.1. Importance Sampling</td>
<td>8</td>
</tr>
<tr>
<td>2.1.4.2. Sequential Importance Sampling and Resampling</td>
<td>9</td>
</tr>
<tr>
<td>2.1.4.3. Parameter Estimation</td>
<td>15</td>
</tr>
<tr>
<td>2.2. Portfolio Optimization and Quadratic Optimization</td>
<td>16</td>
</tr>
<tr>
<td>2.2.1. Quadratic Programming</td>
<td>16</td>
</tr>
<tr>
<td>2.2.2. Mean-Variance Theory</td>
<td>17</td>
</tr>
<tr>
<td>2.2.3. Capital Asset Pricing Model</td>
<td>19</td>
</tr>
<tr>
<td>2.2.4. Portfolio Transaction Cost</td>
<td>20</td>
</tr>
<tr>
<td>2.2.5. Analysts' Consensus Rating</td>
<td>21</td>
</tr>
</tbody>
</table>
3. Modeling Dynamic Risk Neutral Density .......................................... 23
   3.1. Introduction ................................................................. 23
   3.2. Literature Review .......................................................... 24
   3.3. Theoretical Background ...................................................... 27
      3.3.1. Risk Neutral Density and Option Price ......................... 27
      3.3.2. Black-Sholes Model .................................................. 28
      3.3.3. Mixed Log-normal Distribution .................................... 29
   3.4. Method ................................................................. 31
      3.4.1. Model ................................................................. 31
      3.4.2. Sequential Monte Carlo with Parallel Computing ............. 32
   3.5. Simulation Study .......................................................... 37
      3.5.1. Data Simulation ......................................................... 37
      3.5.2. Least Square Estimation ............................................. 38
      3.5.3. Sequential Monte Carlo Parameter Estimation ............... 41
      3.5.4. Sequential Monte Carlo Prediction ................................ 41
   3.6. Empirical Study .......................................................... 43
      3.6.1. Data Description ...................................................... 43
      3.6.2. Least Square Estimation ............................................. 44
      3.6.3. Determination of State Dynamics .................................. 47
      3.6.4. Sequential Monte Carlo Parameter Estimation ............... 49
      3.6.5. Sequential Monte Carlo Prediction ................................ 51
   3.7. Conclusion ............................................................... 56

4. Modeling Dynamic Cross-Sectional Distribution ............................. 57
   4.1. Introduction ............................................................... 57
   4.2. Method ................................................................. 61
   4.4. Feature Process Estimation .............................................. 66
   4.5. Determination of the State Dynamics .................................. 67
4.6. Sequential Monte Carlo Prediction ........................................... 70
4.7. Conclusion ............................................................................. 72

5. Constrained Minimum Variance Portfolio Optimization With Analysts’ Consensus Ratings and Transaction Costs .................. 73

5.1. Introduction ........................................................................... 73
5.2. Method .................................................................................. 76
5.3. Empirical Study ..................................................................... 78
  5.3.1. Data Description ................................................................. 78
  5.3.2. Analysis of Analysts' Consensus Ratings ............................. 78
  5.3.3. Portfolio Optimization and Parameter Estimation .............. 80
  5.3.4. Portfolio Performance Evaluation ....................................... 81
5.4. Conclusion ............................................................................. 83
## List of Tables

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Environment</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>Parameter Estimation Result</td>
<td>38</td>
</tr>
<tr>
<td>3.3</td>
<td>SPY Option Price Data Summary</td>
<td>43</td>
</tr>
<tr>
<td>3.4</td>
<td>Time Series Model Coefficients $\phi_L$ for Year 2013 and 2014</td>
<td>49</td>
</tr>
<tr>
<td>4.1</td>
<td>MLE Estimator $\theta_t$ Summary Statistics</td>
<td>67</td>
</tr>
<tr>
<td>4.2</td>
<td>Average Prediction Errors Comparison</td>
<td>71</td>
</tr>
<tr>
<td>5.1</td>
<td>Summary of Analysts Consensus Rating 2004-2011</td>
<td>79</td>
</tr>
<tr>
<td>5.2</td>
<td>Portfolios Constructed for Comparison</td>
<td>82</td>
</tr>
<tr>
<td>5.3</td>
<td>Portfolios Optimization Result</td>
<td>83</td>
</tr>
</tbody>
</table>
List of Figures

2.1. Efficient Frontier and Capital Market Line ........................................... 19
2.2. Analysts Rating Chart ................................................................. 22
3.1. CUDA Program Structure .............................................................. 34
3.2. Vector Addition Performance Comparison ........................................... 35
3.3. Particle Filter Performance Comparison ............................................. 36
3.4. Risk Neutral Density - Simulation .................................................... 39
3.5. Times Series Plots Comparison of $\hat{\theta}_t$, $\hat{\theta}_{L,t}$ and $\hat{\theta}_{M,t}$ .... 40
3.6. Prediction Performance Comparison by MSE - Simulation .................... 42
3.7. SPY ETF Price in Year 2013 and 2014 .............................................. 44
3.8. SPY Call Option Price - Year 2014 .................................................. 45
3.9. Least Square estimated Risk Neutral Density Plot - Year 2014 .............. 46
3.10. Times Series Plots of $\hat{\theta}_{L,t}$ - Year 2014 .................................. 48
3.11. Times Series Plots of $\theta_{L,t}$ and $\theta_{M,t}$ - Year 2014 ................. 50
3.12. Risk Neutral Density Plot for SMC Fit - Year 2014 ......................... 51
3.13. Option Price Prediction Daily MSE Comparison - Year 2014 ............ 52
3.15. Option Price Prediction MSE Comparison by Moneyness - Year 2014 ... 54
3.16. Risk Neutral Density Plot for SMC Predictions - Year 2014 ............. 55
4.2. Basic Statistical Properties of Cross-Sectional Distribution ................. 65
4.3. Density plot of $f(x|0, 1, 2, 0.5)$, $N(0, 1)$ and $T(2)$ .......................... 66
4.4. Time Series Plot for Maximum Likelihood Estimator $\hat{\theta}_t$ ............. 68
4.5. SMC Daily Prediction Error Plot ..................................................... 71
5.1. 2004-2011 S&P 500 Stocks Consensus Ratings vs Returns .................. 79
5.2. Parameter Estimation for $\lambda$
Chapter 1

Introduction

Thanks to the rapid development of science and technology, the amount of data generated has been increasing tremendously in every aspect since beginning of 21st century. IDC reports by John and David (2012) and David et al. (2018) predict that global collective sum of data would accumulate 40 zettabytes (ZB) by the end of year 2020 and 175 ZB by 2025. Such explosion of data calls for more robust computational power and more sophisticated statistical models. Moreover, in financial industry, most of data is time-sequenced and requires immediate response. Sometimes decisions even have to be made within milliseconds such as in high-frequency trading. Thus more efficient solutions are needed to analyze financial times series to explore relationships between past and present events, reveal underlying dynamics over time and make future forecasts in real time.

In the first part of this dissertation, we specifically focus on modeling functional times series that are driven by its feature processes (FTS-FS) and propose a generalized two-stage Sequential Monte Carlo joint estimation framework through state-space models. In the meantime, to overcome the computational difficulties of handling high dimensionality and non-linearity, we introduce parallel computing algorithms and schemes and integrate them with Sequential Monte Carlo calculations to allow fast on-line estimation and prediction.

Two applications are conducted to demonstrate the robustness and efficiency of our model. The first application presents a generalized framework to estimate implied dynamic risk neutral densities from observed call option prices. As a powerful and popular tool to extract information from financial derivatives, risk neutral density is however not directly observable by market participants. Traditional approaches usually
use complicated numerical methods and fail to recognize and utilize the intrinsic time-evolving dynamical relationship. We view risk neutral density as a FTS-FS and model it using a modified parametric mixed log-normal distribution. We then obtain feature process estimates with classic least square estimation method and explore the dynamics through time series analysis. By introducing a generalized state-space model, we apply Sequential Monte Carlo parameter estimation to re-estimate the coefficients in the time series models. In a simulation study, prediction performance of our proposed model using calibrated parameters dominates that of least square estimation based random walk model. In a real data study, however, our model delivers an uneven performance and possible reasons are then discussed. The second application studies daily cross-sectional distribution of stock returns. The cross-sectional effects of expected stock returns have been comprehensively studied for several decades. Many theories and models have been proposed to explain the differences such as Capital Asset Pricing Model (CAPM) by Sharpe (1964) and Lintner (1965) and Fama French three factor model by Fama and French (1992). However, few researches manage to recognize the underlying dynamics nor stochastic properties of the cross-sectional distribution over time. We view the dynamic daily cross-sectional distribution as a FTS-FS and model it using our two-stage Sequential Monte Carlo joint estimation framework. A parametric skewed t-distribution is chosen to approximate daily cross-sectional distribution. In the empirical study of the 1,000 largest market capitalization stocks from year 1991 to year 2002, prediction performance of our model is considerably superior to that of a random walk model.

In the second part of this dissertation we present a new portfolio optimization strategy of minimum variance portfolios with pre-screened stock selection and constrained short-sale and transaction costs. Classic mean-variance theory minimizes portfolio risk on given level of expected returns. In practice, it is difficult to accurately estimate expected asset returns and their covariance matrix together and even smallest estimation errors could be carried out and greatly impact predictions. In the meantime, empirical studies reveal that portfolio formed with smaller beta stocks could have higher return than portfolio with larger betas. Inspired by these findings we propose a new portfolio
optimization strategy of minimum variance only portfolios and pre-screen stocks using analysts’ consensus ratings. With additional constraints on short-sale and transaction costs, in an empirical study of S&P 500 index stocks, portfolios constructed using our strategy notably outperform benchmark portfolios and portfolios in existing literatures.
Chapter 2

Preliminary

2.1 Functional Time Series and State-Space Model

2.1.1 Functional Time Series Driven by Its Feature Process

Functional data are commonly seen in finance, such as daily curve of interest rates and option volatility curves. If the measurement data functions $Y$ are indexed by time $t$, it is referred as functional time series. Fewer related researches have been conducted comparing with vector or scalar time series. Generally it is difficult to analyze such data structure because of the high dimensionality and temporal dependence between observations. Traditional approaches usually involve dimension reduction techniques to approximate the original functional time series using lower dimensional time series models. Instead, we design a new generalized method via state-space model to model functional time series.

Let $Y_t$ be observed functional time series data and $Y_t = \{y_t(\tau), \tau \in \Omega\}$ itself is a function defined on $\Omega$. Suppose there exists a latent dynamic process $X_t$ such that

$$
P(Y_t|X_t, Y_{t-1}) = P(Y_t|X_t)$$

$$
P(X_t|X_{t-1}, Y_{t-1}) = P(X_t|X_{t-1}),$$

(2.1)

where $X_t = (X_1, \ldots, X_t)$ and $Y_{t-1} = (Y_1, \ldots, Y_{t-1})$. $Y_t$ is then called functional time series driven by its feature process (FTS-FP). The idea behind FTS-FP is to reduce the original high dimension functional data into a lower dimensional latent feature process. By studying its dynamics with regular time series models such as ARMA and GARCH, the original functional time series could be estimated and predicted provided information loss is minimal. Their relationship can be explicitly described using generalized state-space model with Sequential Monte Carlo techniques.
If $Y_t$ is a density function satisfying $y_t(\tau) \geq 0$ and $\int y_t(\tau) \, d\tau = 1$, it is also called a **distributional time series**. To model $Y_t$, we may think the density function $y_t(\tau)$ belongs to an existing parametric family $\pi_t(\cdot|X_t)$. Similarly as FTS-FP, $X_t$ itself is a latent dynamic process such that

$$Y_t = \pi_t(\cdot|X_t) \tag{2.2}$$

$$X_t = f(X_{t-1}, \varepsilon_t),$$

where $f(\cdot)$ could be as simple as a low dimensional time series model. Such functional time series structure is called **distributional time series driven by its feature process** (DTS-FP).

Note in both FTS-FP and DTS-FP, $X_t$ itself is a first-order Markov process evolving over time and is independent of $Y_t$. At each time point $t$, current observation $Y_t$ only depends on past observations $Y_{t-1}$ through current latent state $X_t$. Therefore both FTS-FP and DTS-FP models can be further simplified in terms of a generalized state-space model:

$$P(Y_t|X_t) = q_t(Y_t|X_t, \eta)$$

$$P(X_t|X_{t-1}) = f_t(X_t|X_{t-1}, \theta), \tag{2.3}$$

where $\eta$ and $\theta$ are unknown parameters.

### 2.1.2 State-Space Model

Since firstly introduced in 1960s, state-space model has been extensively studied and widely used in many areas such as engineering and finance. In statistics, it is often used to analyze behaviours of stochastic dynamic systems.

In general, a state-space model is defined by two related time series equations:

$$X_t = s_t(X_{t-1}, \varepsilon_t) \tag{2.4}$$

$$Y_t = h_t(X_t, \varepsilon_t).$$

The state variable, $X_t$, defines a latent continuous process that evolves over time through transition function $s_t$ and known state innovation $\varepsilon_t$. The observation process,
$Y_t$, defines what is to be observed and only depends on current state variable $X_t$ through observation function $h_t$. Equivalently, they can be written in probability terms as:

$$
x_t \sim q_t(\cdot|x_{t-1})
$$

$$
y_t \sim f_t(\cdot|x_t).
$$

(2.5)

![Diagram](image)

The main goal of a state-space model is to obtain the hidden state variables $X_t$ through given observations $Y_t$. Hence, it is equivalent to calculate the posterior distribution:

$$
p(x_t|y_t) \propto \prod_{s=1}^{t} f(y_s|x_s)q_s(x_s|x_{s-1}).
$$

Most of the stochastic dynamic systems, such as motion tracking and financial derivative pricing, require that estimation and prediction be calculated in real time. This can be easily solved with an explicit solution of $X_t$. In ideal cases when $s_t$, $h_t$ are linear and $\varepsilon_t$, $e_t$ are Gaussian errors, calculation is straightforward through so-called Kalman Filter. In practice, however, it is common to have non-linear relationships and hence estimation is usually done through numerical computation in a recursive manner such as EM and Monte Carlo methods.

In state-space model framework, three main problems are:

- Filtering: the objective here is to use all historical observed information to estimate current state $x_t$ under marginal posterior distribution $p(x_t|y_t): E[h(X_t)|Y_t]$.

- Prediction: similarly, the objective is to estimate a future state $x_{t+1}$ under future marginal posterior distribution $p(x_{t+1}|y_t): E[h(X_{t+1})|Y_t]$.

- Smoothing: smoothing focus on estimating past states $X_{t-1}$ under marginal posterior distribution $p(x_{t-1}|y_t): E[h(X_{t-1})|Y_t]$.

The following part of this chapter focuses mainly on filtering, prediction and parameter tuning of models based on state-space framework.
2.1.3 Kalman Filter

As discussed above, when a state-space model has linear relationship with independent Gaussian noise, an explicit solution of current state \( X_t \) (filtering) could be obtained through linear algebra. It is firstly introduced by Kalman (1960) and named as *Kalman Filter*.

We start with the following state-space model:

\[
X_t = H_t X_{t-1} + W_t w_t \\
Y_t = G_t X_t + V_t v_t,
\]

where \( H_t, G_t, W_t, \text{ and } V_t \) are all known matrices with white noise \( w_t, v_t \sim N(0, I) \). It is obvious that the marginal posterior distribution \( p(x_t|y_t) \) follows a multivariate Normal distribution \( N(\mu_t, \Sigma_t) \). Given a past state \( \mu_{t-1} \) and \( \Sigma_{t-1} \), Kalman Filter is written as:

\[
P_t = H_t \Sigma_{t-1} H_t' + W_t W_t' \\
S_t = G_t P_t G_t' + V_t V_t' \\
\mu_t = H_t \mu_{t-1} + P_t G_t' S_t^{-1} (Y_t - G_t H_t \mu_{t-1}) \\
\Sigma_t = P_t - P_t G_t' S_t^{-1} G_t P_t.
\]

Hence filtering can be calculated recursively by updating \( \mu_t \) and \( \Sigma_t \) as model evolves over time. Prediction and smoothing formulas could also be derived in the same manner.

Kalman-Filter application is limited in practice as linearity assumption is too strong to achieve. In nonlinear cases, it is difficult to get an explicit closed form of high dimensional integration of \( x_t \):

\[
E[h(X_t)|Y_t] = \int \cdots \int h(x_t)p(x_t|y_t) \, dx_1 \cdots dx_t.
\]

Therefore we adopt *Sequential Monte Carlo* framework to do numerical approximation.

2.1.4 Sequential Monte Carlo Framework

Since firstly introduced in 1990s by Gordon et al. (1993), Sequential Monte Carlo (SMC) methods have been widely accepted and applied in many scientific and engineering fields.
such as signal and image processing, object tracking, financial risk analysis and artificial intelligence. For complicated non-linear state-space models, SMC methods provide a great way to inference states and posterior densities.

The basic principles of SMC are based on sequential importance sampling (SIS). In state-space problems, to approximate posterior distribution $p(x_t|y_t)$, SMC methods sequentially generate samples (also called particles) from a trial distribution $g_t(\cdot)$, attach particles to their sampling paths and compute importance weights recursively in each iteration. A resampling step is inserted when necessary to rejuvenate samples by replacing particles of low importance weights with those of high importance weights. Moreover, when state-space models depend on some unknown parameter $\theta$, SMC methods provide ways called parameter estimation to fine tune parameters before states estimation and prediction.

As of today there have been several SMC techniques such as standard particle filter (Gordon et al. (1993)), sequential imputation (Kong et al. (1994); Liu and Chen (1995)) and Monte Carlo filter (Kitagawa (1996)). They are superior in scalability and generality to other techniques when solving state-space problems. SMC methods do not have any limitations on prior distributions even when they are non-linear and non-Gaussian. With proper implementation, all states and posterior distributions could be approximated and the precision only depends on particle size. Therefore, with development of new algorithm and computer hardware, SMC techniques could provide faster and more precise results than ever before.

2.1.4.1 Importance Sampling

In a general Monte Carlo problem, suppose that we aim to estimate some complex target distribution $\pi(x)$ and expectation with respect to $\pi(x)$:

$$E_x h(x) = \int h(x)\pi(x) \, dx.$$  

Vanilla Monte Carlo techniques simply draw random samples $\{x^{(i)}, 0 \leq i \leq N\}$ from
\(\pi(x)\) and approximate the integral using its sample mean:

\[
\hat{E}_\pi h(x) \approx \frac{1}{N} \sum_{i=1}^{N} h(x^{(i)}).
\]

Obviously, as long as sample size \(N\) is sufficiently large estimation precision is satisfiable. However, in most cases \(\pi(x)\) is not directly known or only known up to its normalising factor. Instead, we may evaluate the expectation with respect to some known probability distribution \(g(x)\) (also called importance distribution):

\[
E_\pi h(x) = \int h(x)\pi(x)\,dx = \int h(x)\frac{\pi(x)}{g(x)}g(x)\,dx = E_g[h(x)w(x)].
\]

Thus we now draw samples \(\{x^{(j)}, 0 \leq j \leq N\}\) from \(g(x)\) and compute a weighted average based on importance weight \(w^{(j)}\):

\[
\hat{E}_\pi h(x) \approx \frac{1}{\sum_{j=1}^{N} w^{(j)}} \sum_{j=1}^{N} w^{(j)}h(x^{(j)}).
\]

Here sample mean, as an unbiased estimator, is not favored as it requires exact calculation of \(\frac{\pi(x^{(j)})}{g(x^{(j)})}\), which involves evaluation of actual normalising constant in \(\pi(x)\) and \(g(x)\). Notice that in Equation (2.8) importance weight \(w^{(j)}\) is only needed as proportional value without evaluating the actual normalising constant. Apply the Strong Law of Large Numbers:

\[
\hat{E}_\pi h(x) \xrightarrow{\text{a.s.}} \frac{N \to \infty}{N} E_\pi h(x).
\]

When applying importance sampling technique to state-space models, each time a new observation \(y_t\) is obtained, one has to generate the entire particle population path again and recompute the importance weights for all states, which brings increased computational burden in each iteration. To simplify this problem a sequential version of importance sampling procedure is designed below.

### 2.1.4.2 Sequential Importance Sampling and Resampling

The idea behind sequential importance sampling is to treat sampling process recursively so that the particle population path could be reused. Each time a new observation \(y_t\) is obtained, new particle would be generated based on its past path \(\{(x_{t-1}^{(i)}, y_t), j =\)
Instead of recalculating the entire particle path from the very beginning. To achieve this, the target distribution $\pi(x)$ itself has to be designed or approximated in a recursive form with respect to time $t$ ($1 \leq t \leq T$):

$$\pi(x) = \pi(x_1)\pi(x_2|x_1)\cdots\pi(x_T|x_{T-1}). \quad (2.10)$$

Therefore, instead of directly sampling from $\pi(x)$, we use another trial distribution $g(x)$ that can be expressed sequentially in the following manner

$$g(x) = g_1(x_1)g_2(x_2|x_1)\cdots g_T(x_T|x_{T-1}), \quad (2.11)$$

whose pattern matches the target distribution $\pi(x)$ in Equation (2.10). In this way, each time a new data $y_t$ is observed, we could sample from $g_t(\cdot)$ and attach to its sampling path sequentially. Importance weight can be calculated as

$$w(x) = \frac{\pi(x_1)\pi(x_2|x_1)\cdots\pi(x_T|x_{T-1})}{g_1(x_1)g_2(x_2|x_1)\cdots g_T(x_T|x_{T-1})}. \quad (2.12)$$

Clearly a similar recursive relationship also exists for $w_t(x_t)$

$$w_t(x_t) = w_{t-1}(x_{t-1})\frac{\pi(x_t|x_{t-1})}{g_t(x_t|x_{t-1})}. \quad (2.13)$$

Since it is difficult to find the marginal distribution $\pi(x_t)$ and decompose $\pi(x)$ using Equation (2.10), we could use a sequence of auxiliary distributions $\{\pi_t(x_t), t = 1, 2, \cdots, T\}$ as approximations to original marginal distributions $\{\pi(x_t), t = 1, 2, \cdots, T\}$. Thus Sequential Importance Sampling (SIS) can be formally defined by a two-step recursive procedure for $t > 1$:

1. Sample $x_t$ from $g_t(x_t|x_{t-1})$ and attach to its past sampling path $x_t = (x_{t-1}, x_t)$.

2. Compute incremental weight

$$u_t = \frac{\pi_t(x_t)}{\pi_{t-1}(x_{t-1})g_t(x_t|x_{t-1})} \quad (2.14)$$

and importance weight $w_t = w_{t-1}u_t$.

To apply SIS framework on state-space models, we firstly decompose posterior distribution $p(x_t|y_t)$ recursively using Bayes' Theorem:

$$p(x_t|y_t) = p(x_{t-1}|y_{t-1})p(y_t|x_t)p(x_t|x_{t-1}) \frac{p(y_t|y_{t-1})}{p(y_t|y_{t-1})}. \quad (2.15)$$
Similarly, marginal posterior distribution $p(x_t|y_t)$ also has following recursive relationship:

$$p(x_t|y_t) = \frac{p(y_t|x_t)p(x_t|y_{t-1})}{\int p(y_t|x_t)p(x_t|y_{t-1}) \, dx_t} \propto p(y_t|x_t) \int p(x_t|x_{t-1})p(x_{t-1}|y_{t-1}) \, dx_{t-1}. \quad (2.16)$$

Therefore, for state-space models a trial sampling distribution $g^*_t(x_t|y_t)$ can be designed sequentially as

$$g^*_t(x_t|y_t) = g^*_{t-1}(x_{t-1}|y_{t-1})g_t(x_t|x_{t-1}, y_t). \quad (2.17)$$

In this way, each time when a new data $y_t$ is observed, we may simply sample from $g_t(\cdot)$, attach to its path and derive importance weight sequentially:

$$w_t^{(i)} = \frac{p(x_t^{(i)}|y_t)}{g_t(x_t^{(i)}|y_t)} = w_{t-1}^{(i)} \frac{1}{p(y_t|y_{t-1})} \frac{q_t(x_t^{(i)}|x_{t-1}^{(i)})f_t(y_t|x_t^{(i)})}{g_t(x_t^{(i)}|x_{t-1}^{(i)}, y_t)} = w_{t-1}^{(i)} u_t^{(i)}, \quad (2.18)$$

where $w_t$ is the normalised importance weight and $u_t$ is the normalised incremental weight. The detailed procedures of application of SIS to state-space models are summarized in Algorithm 1.

In practice, the variance of the weights will stochastically increase according to studies by Kong et al. (1994). Hence after some iterations it is likely that most of the importance weights will decrease to 0 and only a few of them would take all probability masses. The entire particle population would diminish into few paths very quickly. This is called weight degeneracy. This phenomenon makes SIS impractical and has to be carefully taken care of. A trivial way of solving this is to sample particle paths with replacement with respect to their normalised importance weights as probabilities. Therefore particle paths with larger importance weights get replenished while paths with lower weights get dropped. It is obvious to see that such resampling process does increase Monte Carlo variances of the problem. However, it prevents from accumulating errors over time and helps stabilize filtering distributions at the same time. There are
Algorithm 1 Sequential Importance Sampling

1: if $t = 0$ then
2: \hspace{1em} Sample $x_0^{(i)} \sim q_0(x_0|y_0), \; i = 1, \ldots, N$
3: \hspace{1em} Calculate initial weight: $w_0^{(i)} = \frac{f_0(y_0|x_0^{(i)})q_0(x_0^{(i)})}{g_0(x_0^{(i)}|y_0)}$
4: end if

5: if $t > 0$ then
6: \hspace{1em} for $t = 1, \ldots, T$ do
7: \hspace{2em} for $i = 1, \ldots, N$ do
8: \hspace{3em} Sample $x_t^{(i)} \sim g_t(x_t^{(i)}|x_{t-1}^{(i)}, y_t)$
9: \hspace{3em} Calculate weight:
10: \hspace{4em} $\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \frac{f_t(y_t|x_t^{(i)})q_t(x_t^{(i)}|x_{t-1}^{(i)})}{g_t(x_t^{(i)}|x_{t-1}^{(i)}, y_t)}$
11: \hspace{2em} end for
12: \hspace{1em} Normalise weight:
13: \hspace{2em} $w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}, \; i = 1, \ldots, N$
14: \hspace{1em} Estimate filtering function:
15: \hspace{2em} $\hat{E}_\pi h_t(x) = \sum_{i=1}^N w_t^{(i)} h_t(x_t^{(i)})$
16: end for
17: end if
several reampling schemes available such as stratified reampling, systematic reampling and residual reampling. Their technical details and performances have been studied and compared by Hol et al. (2006).

Carrying out reampling in every iteration of SIS not only brings unnecessary Monte Carlo variations but also increases computation burden. We may set up a deterministic schedule by fixing reampling steps at given timestamps \( \{t_1, t_2, t_3, \ldots, t_k\}_{1 \leq k \leq T} \). Another method is to dynamically resample particles according to the level of weight degeneracy in each iteration. After all resampling is only needed when weight degeneracy actually happens. Kong et al. (1994) introduces a general measurement of weight variability called Effective Sample Size (ESS):

\[
\text{ESS}(N) = \frac{N}{1 + \text{Var}_g[w(x)]},
\]

where \( N \) stands for particle sample size. Given a predetermined threshold \( c < 1 \), then resampling is carried out whenever \( \text{ESS} < cN \).

Combining SIS with resampling procedure, we finally have a complete Sequential Monte Carlo framework (also called Particle Filter) that is summarised in Algorithm 2. Depending on the choice of different sampling distributions, many Sequential Monte Carlo methods are developed such as Bootstrap Filter.

From Algorithm 2 a straightforward and convenient choice of sampling distribution \( g_t(\cdot) \) would be the state transition distribution \( q_t(x_t|x_{t-1}) \). While calculating importance weight it will get cancelled out and incremental weight \( u^{(i)}_t \) would simply be \( f_t(y_t|x^{(i)}_t) \). Such Sequential Monte Carlo method is also called Standard Particle Filter, or Bootstrap Filter. In Bootstrap Filter sampling distribution only depends on past states. Observations are only used to calculate importance weights. Generated particles would “move” randomly in probability space without any knowledge of new observations.

An alternative choice of sampling distribution would be going the opposite way by only using the observation function \( f_t(y_t|x_t) \). Unlike that in Standard Particle Filter, newly generated particles \( x^{(i)}_t \) are independent of their past states \( x^{(i)}_{t-1} \). Incremental weight \( u^{(i)}_t \) would be simply proportional to \( g_t(x^{(i)}_t|x^{(i)}_{t-1}) \). Thus this method is called
Algorithm 2 Particle Filter

1: if $t = 0$ then
2: Sample $x_0^{(i)} \sim g_0(x_0 | y_0)$, $i = 1, \ldots, N$
3: Calculate initial weight: $w_0^{(i)} = \frac{f_0(y_0 | x_0^{(i)}) q_0(x_0^{(i)})}{g_0(x_0^{(i)} | y_0)}$
4: end if
5: if $t > 0$ then
6: for $t = 1, \ldots, T$ do
7: for $i = 1, \ldots, N$ do
8: Sample $x_t^{(i)} \sim g_t(x_t^{(i)} | x_{t-1}^{(i)}, y_t)$
9: Calculate weight:
$$
\tilde{w}_t^{(i)} = \frac{w_t^{(i)} f_t(y_t | x_t^{(i)}) q_t(x_t^{(i)} | x_{t-1}^{(i)})}{g_t(x_t^{(i)} | x_{t-1}^{(i)}, y_t)}
$$
10: end for
11: Normalise weight:
$$
w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}, i = 1, \ldots, N
$$
12: if $\text{ESS}(N) < cN$ then
13: Sampling particle paths $\{x_t^{(i)}, i = 1, \ldots, N\}$ with replacement with respect to importance weights $w_t$ as probabilities
14: end if
15: end for
16: end if
Independent Particle Filter. While information from observation process is strong and that from state process is weak, Independent Particle Filter could be a great option to use.

Both Standard Particle Filter and Independent Particle Filter only use parts of the available information in state-space problems. To compensate for the information loss and utilize both state and observation processes, another choice would be using so-called optimal kernel, coupling both state transition density and observation function together:

\[
g_t(x_t|x_{t-1},y_t) = \frac{q_t(x_t|x_{t-1})f_t(y_t|x_t)}{\int q_t(x_t|x_{t-1})f_t(y_t|x_t) \, dx}.
\] (2.20)

The optimal kernel in Equation (2.20) contains information from both previous states \(x_{t-1}\) and new observation \(y_t\), so does the incremental weight \(u_t^{(i)}\) which is simply the normalising constant:

\[
u_t^{(i)} = \int q_t(x_t^{(i)}|x_{t-1}^{(i)})f_t(y_t|x_t^{(i)}) \, dx.
\] (2.21)

New generated particles will thus move towards the conditional posterior distribution, making sampling more efficient. In certain cases this method performs better than both Standard Particle Filter and Independent Particle Filter. However, in practice it is often difficult to directly sample from optimal kernel due to technical issues. Instead Markov chain Monte Carlo and Gaussian approximation techniques are commonly used to approximate the optimal kernel function.

2.1.4.3 Parameter Estimation

In most cases state-space models also depend on one or more unknown parameters as vector \(\theta\). Therefore from now on we denote the state and observation transition probability functions as \(q_t(x_t|x_{t-1},\theta)\) and \(f_t(y_t|x_t,\theta)\) respectively. One of the most important problems in SMC is Parameter Estimation, tuning parameters properly before jumping into states estimation and prediction. Generally we calibrate parameters using maximum likelihood estimation (MLE).

We are interested in evaluating the data likelihood \(p(y_T|\theta)\) with \(t = 1, 2, \cdots, T\). First we recognize that \(p(y_t|y_{t-1})\) is the normalising factor of unnormalised incremental
weight \( \tilde{u}_t \) from Equation (2.18). Thus \( p(y_t) \) is the normalising factor of unnormalised incremental weight \( \tilde{w}_t \). Then we have:

\[
p(y_T|\theta) = \int p(x_T, y_T|\theta) \, dx_T
= \int \frac{p(x_T, y_T|\theta)}{g(x_T|y_T)} g(x_T|y_T) \, dx_T
= E_g[\tilde{w}_T] \approx \frac{1}{N} \sum_{i=1}^{N} \tilde{w}_{T}^{(i,\theta)}.
\] (2.22)

Thus the data likelihood can be approximated by the sample mean of unnormalised weights, which is very simple and convenient to compute. We no longer have to store entire particle paths but only need to track importance weights. Remember that we have not taken resampling step of SMC into consideration while formulating Equation (2.22). Note that every time after resampling all particle importance weights are reset to original state of \( \frac{1}{N} \). Suppose that resampling takes place \( k - 1 \) times during filtering, denoted as \( t_1, \ldots, t_{k-1} \). Also make \( t_k = T \). We have:

\[
\hat{\ell}(\theta) = \log \hat{p}(y_T|\theta) = \sum_{j=1}^{k} \log \left[ \frac{1}{N} \sum_{i=1}^{N} \tilde{w}_{t_j}^{(i,\theta)} \right].
\] (2.23)

Therefore parameter estimation of SMC reduces to solving a non-linear optimization problem:

\[
\hat{\theta} = \arg\max \sum_{j=1}^{k} \log \left[ \frac{1}{N} \sum_{i=1}^{N} \tilde{w}_{t_j}^{(i)} \right].
\] (2.24)

When the dimension of parameter vector \( \theta \) is large and time \( T \) goes up, both evaluation and optimization of log-likelihood in Equation (2.24) is computationally intensive. On one hand more efficient optimization strategies are required such as iterative Gauss-Newton and EM algorithms, for example [Shumway and Stoffer (1982), Segal and Weinstein (1989) and Cappé et al. (2005)]. On the other hand, more advanced programming skills are urgently needed such as parallel computing.

### 2.2 Portfolio Optimization and Quadratic Optimization

#### 2.2.1 Quadratic Programming

Quadratic programming (QP) is the problem of optimizing quadratic objective functions subject to finite number of both linear equality and inequality constraints. A quadratic
function of $n$-dimensional vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ has the form

$$f(\mathbf{x}) = \sum_{i=1}^{n} c_i x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij} x_i x_j,$$

(2.25)

where $c_i$ and $q_{ij}$ are all real valued and constant term is omitted for simplicity. Further it can be written as a matrix quadratic form:

$$f(x) = \mathbf{c}' \mathbf{x} + \frac{1}{2} \mathbf{x}' \mathbf{Q} \mathbf{x},$$

(2.26)

where $\mathbf{c}$ is a $n$-dimensional real valued vector and $\mathbf{Q}$ is a $n \times n$ dimensional real matrix.

Note we could always transform $\mathbf{Q}_* = \frac{1}{2}(\mathbf{Q} + \mathbf{Q}')$ to make it symmetric as quadratic form stays unchanged under transpose $\mathbf{x}' \mathbf{Q} \mathbf{x} = \mathbf{x}' \mathbf{Q}' \mathbf{x}$.

Thus a generalized quadratic programming problem with linear constraints can be formulated as

$$\begin{align*}
\text{minimize} & \quad \mathbf{c}' \mathbf{x} + \frac{1}{2} \mathbf{x}' \mathbf{Q} \mathbf{x} \\
\text{subject to} & \quad \mathbf{A}_1 \mathbf{x} = \mathbf{b}, \\
& \quad \mathbf{A}_2 \mathbf{x} \leq \mathbf{d},
\end{align*}$$

(2.27)

where $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A}_1 \in \mathbb{R}^{m \times n}$, $\mathbf{d} \in \mathbb{R}^p$, $\mathbf{A}_2 \in \mathbb{R}^{p \times n}$ and $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is symmetric.

Assume $\mathbf{A}_1$ and $\mathbf{A}_2$ are full rank and $\mathbf{Q}$ is positive definite, when only equality constraints exist the Equation (2.27) has a unique solution that can be solved through Lagrangian and KKT condition matrix. Otherwise many numeric methods have been proposed such as interior point, active set and conjugate gradient.

Quadratic programming is firstly studied by Frank and Wolfe (1956) and later introduced by Markowitz (1952) in portfolio optimization. Since covariance matrix is always positive semi-definite and can be easily set to be strictly positive definite with proper assumptions, it is very convenient to employ quadratic programming to minimize portfolio risk.

### 2.2.2 Mean-Variance Theory

As one of the foundations of modern finance theory, mean-variance portfolio selection theory is firstly introduced by Markowitz (1952). It has been widely accepted and used in the financial industry. Suppose that a diversified portfolio $\mathbf{P}$ consists of $N$ risky
assets $S_1, S_2, \ldots, S_N$. Assume their rate of return vector $\mathbf{r}$ is normally distributed with mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$. Further assume that no asset in the portfolio is the linear combination of any other assets. Let $\mathbf{w} \in \mathbb{R}^N$ be portfolio allocation weight vector that satisfies $\mathbf{w}'\mathbf{1} = 1$, where each $w_i$ denotes the proportion of wealth invested in asset $S_i$. When $w_i > 0$ it means holding a long position whereas $w_i < 0$ indicates a short position. Denote portfolio return vector as random variable $R_P$ and portfolio risk is measured by variance $\text{Var}[R_P]$:

$$\mathbb{E}[R_P] = \mathbf{w}'\mathbf{\mu}$$

$$\text{Var}[R_P] = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}.$$  

The idea behind mean-variance portfolio theory is to find an optimal $\mathbf{w}_0$ that either maximizes return for a given risk level or minimizes risk for a given target return. Suppose $\mu_0$ is the target rate of return for the portfolio which depends on the risk aversion of the investor. It is equivalent to the following optimization problem:

$$\begin{align*}
\text{minimize} & \quad \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \\
\text{subject to} & \quad \mathbf{w}'\mathbf{1} = 1 \\
& \quad \mathbf{w}'\boldsymbol{\mu} = \mu_0.
\end{align*}$$  \hspace{1cm} (2.28)

This problem can be easily solved by Lagrange multipliers with quadratic programming methods. All optimized mean-variance portfolios $\{(\mu_0, \sigma_0) : \mu_0 = \mathbf{w}_0'\boldsymbol{\mu}, \sigma_0^2 = \mathbf{w}_0'\boldsymbol{\Sigma}\mathbf{w}_0\}$ forms a hyperbolic curve that is called Efficient Frontier. Figure 2.1 shows efficient frontier from a toy example. On the efficient frontier curve (blue line), the leftmost point (red dot) $(\mu_{MVP}, \sigma_{MVP})$ corresponds to the Global Minimum Variance Portfolio or simply Minimum Variance Portfolio. It has lowest volatility among all mean-variance optimized portfolios and is the only portfolio on efficient frontier that does not depend on expected returns. Theoretically it also comes with lowest expected return and its performance, measuring by Sharpe ratio, is dominated by that of the tangency portfolio $M$. 

2.2.3 Capital Asset Pricing Model

*Capital Asset Pricing Model* reveals the relationship between asset’s expected return and its systematic risk. It is built on the modern portfolio theory by [Markowitz](1952) and firstly discovered in the 1960s by several researchers including [Treynor](1962), [Sharpe](1964), [Lintner](1965) and [Mossin](1966). Suppose a risk-free asset $S_0$ exists with return denoted as $R_f$. Depending on the level of risk aversion, an investor would put portion $x$ of his funds in $S_0$ and the rest $1 - x$ into some mean-variance portfolio on the efficient frontier. This leads to a straight line between risk-free point $(0, R_f)$ and any point on efficient frontier. As illustrated in Figure 2.1, the best strategy would be selecting the portfolio $M$ such that the line between $R_f$ and $M$ is the tangent line for efficient frontier. In a complete market, investors could all utilize the above reasoning to achieve the same and best strategy. Therefore all efficient portfolios in the market will be lying on the same straight line known as *Capital Market Line*, which is the combination of risk-free asset $S_0$ and tangency mean-variance portfolio $M$. $M$ is then called the *Market Portfolio* and its return $R_M$ is named as *Market Return*. Based on the fact that return-to-risk ratio for any individual asset is equal to that of the market.
portfolio, *Capital Asset Pricing Model* can be derived as:

\[
\frac{ER_i - R_f}{\beta_i} = ER_M - R_f, \tag{2.29}
\]

where \(R_i\) is the individual rate of return for asset \(S_i\) and \(\beta_i = \frac{\text{Cov}(R_i, R_M)}{\text{Var}R_M}\), which is the covariance between asset return and market return divided by the variance of the market return. In general, \(\beta\) measures the systematic risk of the asset comparing with that of the entire market. It can neither be predicted in advance nor eliminated by diversification. In Figure 2.1 and Equation (2.29), it can be seen that assets with \(\beta > 1\) tend to outperform market but bear higher risk at the same time. Thus theoretically assets with smaller \(\beta\) would be outperformed by the market, so does minimum variance portfolios discussed in Section 2.2.2.

### 2.2.4 Portfolio Transaction Cost

In classic mean-variance theory, portfolio transaction cost is ignored for simplicity and elegance. In practice, however, it is vitally important and could seriously affect performance, especially for large institutional investors. Apparently transaction cost is closely connected to the amount of shares traded and the frequency portfolio is balanced. Thus an optimized portfolio not only needs to chase best risk return tradeoff but also has to control the portfolio turnover rate. Generally speaking, portfolio transaction costs consist of brokerage commissions, margin interests, capital taxes and all related fees. Taking all these into account, transaction costs would clearly be non-linear and mostly not even convex like a V-shape function. Therefore in past portfolio optimization researches transaction costs are treated in three ways:

- Simply ignore it as 0, just like [Markowitz (1952)] did in his mean-variance theory.
- Treat transaction costs as a fixed number and subtract from portfolio expected return. This approach is popular in earlier studies such as [Demsetz (1968), Kyle (1985) and O’Hara and Oldfield (1986)].
- Use a parametric function to approximate transaction costs such as linear, quadratic
and even complex non-linear functions. For example, Perold (1984) extends mean-
variance analysis by treating transactions as linear constraints. For more linear
examples see Davis and Norman (1990), Yoshimoto (1996) and Lobo et al. (2007).
Gărleanu and Pedersen (2013), Kolm and Ritter (2014) and Sefton and Cham-
ponnois (2015) all treat transaction costs as quadratic terms in dynamic portfolio
optimizations. Anderson (1997) incorporates non-linear transaction cost func-
tions in optimization by using the threshold error correction model. Konno and
Wijayanayake (2001) proposes a branch and bound algorithm to solve portfolio
optimization problems with non-linear transaction cost functions.

A simple linear transaction costs example is illustrated below. Let $x_i^+$ and $x_i^-$ be
the amount sold and bought for $i$th asset and $\alpha_i^+$ and $\alpha_i^-$ be the associated cost value
rate per traded shares. Denote transaction costs as $\phi_i(x)$ for $i$th asset and a simple
linear convex approximation could be illustrated as:

$$\phi_i(x) = \alpha_i^+ x_i^+ - \alpha_i^- x_i^-,$$

where $x_i^+ \geq 0, x_i^- \leq 0$.

2.2.5 Analysts’ Consensus Rating

Financial firms publish market reports and make individual stock recommendations
regularly. Their analysts often read a company’s financial statements and talk to its
high-level officials in order to rate a stock. Such ratings are on a 5 point integer-
based scale with 1 being “strong buy” and 5 being “strong sell”, as illustrated in
Figure 2.2. It remains to be a long-debating question that how effective and influential
these recommendations are. The earliest study dates back to 1930s by Cowles (1933),
claiming that analysts’ recommendations do not produce abnormal returns with a few
exceptions. Also according to market efficiency theory, in a complete market, analysts’
ratings must not have prediction powers as all information obtained by analysts are
transparent and publicly available. However, many researches demonstrate that in
practice stocks with better consensus rating do outperform those with lower ratings,
which are discussed in details in Section 5.1.
Figure 2.2: Analysts Rating Chart
Chapter 3
Modeling Dynamic Risk Neutral Density

3.1 Introduction

The price of a financial derivative is the reflection of market expectations and investors’ view of risk preferences. In a complete efficient market, the fair price of any financial derivative is solely determined by discounting the expected future payoff of the underlying asset back to the present value using risk-free rate. Such associated underlying probability distribution of the future payoff is named Risk Neutral Density (RND). Obviously knowing risk neutral density would greatly help financial market participants. For example, financial firms and individuals could use risk neutral density to price derivatives and hedge portfolios risks; decision-makers could use it to predict market sentiment and formulate monetary policies.

Among all different kinds of financial derivatives, option has been extensively studied in the past several decades due to its importance and overwhelming popularity, especially after the well-known Black-Sholes Model are proposed (Black and Scholes, 1973; Merton, 1974). A financial option is a contract between two parties which gives the buyer the right, but not the obligation, to buy (call option) or sell (put option) the underlying asset from or to the seller at a specific strike price on a predetermined maturity date. As mentioned above, option price reflects market expectations about the future price of the underlying asset and its associated risk neutral density contains valuable information that could be used in various areas by all market participants. Therefore, numerous studies have been conducted and various techniques have been developed to extract the implied risk neutral density function from option prices ever since 1970s. However, these researches, summarized below in Section 3.2, fail to recognize the temporal dependence between daily risk neutral densities. The underlying
dynamics and stochastic properties of risk neutral density still remain to be discovered.

Therefore we propose a new and generalized framework of implied risk neutral density estimation through a two-stage Sequential Monte Carlo joint estimation model. We view risk neutral density as a distributional time series with parameter vector $\theta_t$ as its feature process. Since observed option price is determined by its underlying risk neutral density function, we apply the DTS-FS model discussed in Section 2.1.1 together with Sequential Monte Carlo methods discussed in Section 2.1.4 to estimate the implied daily risk neutral density distribution from call option prices and further make predictions.

This chapter of dissertation is organized as follows. Section 3.2 reviews existing methodologies of implied risk neutral density estimation from option prices. Section 3.3 describes theoretical background of risk neutral density derivation and estimation. Section 3.4 shows the full approach in detailed steps. Section 3.5 conducts a simulation study enlightened on real data to validate our estimation framework. Section 3.6 presents an empirical study on SPY call option data to estimate and predict implied risk neutral densities. Section 3.7 concludes.

### 3.2 Literature Review

Cox and Ross (1976) firstly demonstrate the fundamental relationship between risk neutral density distribution and European option prices. Further Breeden and Litzenberger (1978) shows that risk neutral density could be derived from taking second derivative of the option price function. They also try to extract the distribution numerically by taking differences from a series of European style call options. However, their method requires that there are option price data for many to infinite different strike prices. This is virtually impossible to achieve in any financial markets.

Afterwards, many methodologies have been developed inspired by this theory. The underlying idea of all methods is to interpolate and extrapolate existing strike prices to fill missing ones so that its assumption could be satisfied. Due to differences in determining the structures of risk neutral density distribution, these techniques can be
categorized in three groups: structural models, parametric and non-parametric methods.

Structural models do not directly give explicit form of risk neutral density function but use its underlying implicit characterizations to determine the distribution. Hull and White (1987) uses a stochastic volatility model to price a European call option and describes the underlying dynamics for risk neutral density; similarly Scott (1997) develops a jump-diffusion model with stochastic volatility to price options. In such models, risk neutral density function could be obtained by inverting the characteristic functions. However, this usually requires assistance of numeric methods as the volatility are non-stationary.

Non-parametric methods fit risk neutral density functions based on rather general functions (usually polynomial) or numeric approximations. Breeden and Litzenberger (1978) firstly proposes a simple method of taking second partial derivative of option price function with respect to strike price. As we have pointed out, this method requires many to infinite even-spaced strike price data to fit the tails. This is difficult to deal with in practice. In general, there are three major non-parametric approaches including kernel regression methods, maximum entropy methods and curve fitting methods. Kernel regression focuses fitting all missing values from existing data. Pritsker (1997) and Rookley (1998) use kernel regression estimators to model risk neutral densities of interest rates and implied volatilities. Ait-Sahalia and Lo (1998) proposes to estimate option price formula instead using kernel regression and derives risk neutral density by taking second derivative with respect to strike price afterwards. Maximum entropy methods choose a given prior distribution and derive posterior risk neutral density by maximizing the cross entropy. Rubinstein (1994) firstly proposes a tree based model using log-normal distribution as prior based on the binomial model by Cox et al. (1979). Later Jackwerth and Rubinstein (1996) and Jackwerth (1997) apply and further extend this model to more general areas. Buchen and Kelly (1996) uses uniform and log-normal priors and Stutzer (1996) uses historical stock prices as priors. Jondeau and Rockinger (2002) tries normal and t-distribution priors with additional constraints. Curve fitting methods are the most popular and stable among all non-parametric methods by
modeling implied volatility smiles. Shimko (1993) firstly proposes this method by inverting option prices into Black-Scholes implied volatilities. Then he uses a quadratic polynomial fit the implied volatility curve. Then by taking second derivative of the transformed option price function with respect to strike price, estimator of implied risk neutral density can be obtained. Malz (1997) applies this method on option deltas. Brown and Toft (1999) revises this method by using a seventh-order splines. Other than them, Campa et al. (1998) uses cubic splines and Rosenberg and Engle (2002) applies a polynomial fit on log implied volatilities. Mayhew et al. (1995) uses cubic splines to fit risk neutral density function by maximizing the distribution smoothness. Recently in the field a recovery theory is presented by Ross (2015) and raises academic attractions. He uses PerronFrobenius theory on a discrete-time discrete-state Markov model to recover the unique risk neutral density. This theory is soon questioned by Borovika et al. (2016) for its lack of sufficient assumptions. The credibility of this model is still in debate by researchers to this day (Carsten Jackwerth and Menner, 2017; Jensen et al., 2018).

Parametric methods estimate risk neutral density functions based on a given probability distribution with some parameter vector \( \theta \). Then various optimization procedures are applied to find the optimal \( \theta \) with given constraints. Many different distributions or mixture of distributions have been proposed and studied. First of all, note that the famous Black-Scholes Model by Black and Scholes (1973) and Merton (1974) implies that risk neutral density follows a log-normal distribution though it is far from perfection as observed in practice; Madan and Milne (1994) and A. Abken et al. (1996) use a four-parameter Hermite polynomials to extend the Gaussian distribution; Coutant et al. (2001) improves Hermite expansion by putting additional restriction on parameters and applies to interest rate futures options. Another popular semi-parametric distribution model is Edgeworth series expansion of log-normal distribution proposed by Jarrow and Rudd (1982); later Corrado and Su (1996) uses this approach to study S&P500 options; Flamouris and Giamouridis (2002) applies this approach to crude oil futures options and Beber and Brandt (2006) uses it for US Treasury bond futures; Coutant et al. (2001) tries confluent hypergeometric functions (also called Kummer function);
Corrado and Su (1997), Backus et al. (1999) and Rompolis and Tzavalis (2007) use Gram-Charlier series expansion to approximate risk neutral density functions; Eriksen et al. (2007) uses normal inverse Gaussian distribution to fit risk neutral density function due to its nice tail and moments behavior; Markose and Alentorn (2011) uses the generalized extreme value distribution. Among all parametric methods, the most popular and well-known approach is the mixed log-normal distributions. Melick and Thomas (1997) studies crude oil options and finds that the empirical risk neutral densities behave bimodal instead of unimodal. Therefore they suggest a linear mixture of log-normal distributions be used to fit implied risk neutral density for crude oil; Bahra (1997) also proposes a similar approach on European options. Parametric methods bear the advantages of easy computation as only a few parameters are needed to be estimated. However, efficiency and robustness of parametric models highly depend on the data collected and the choice of probability distribution.

3.3 Theoretical Background

3.3.1 Risk Neutral Density and Option Price

Based on the work by Cox and Ross (1976), Breeden and Litzenberger (1978) firstly shows the relationship between the European call option price function and risk neutral measure:

\[
C(S_t, K, T, t) = e^{-r(T-t)} E[\max(S_T - K, 0)|S_t, t] \\
= e^{-r(T-t)} \int_0^\infty \max(s_T - K, 0)f(s_T|S_t, t) \, ds_T \\
= e^{-r(T-t)} \int_K^\infty (s_T - K)f(s_T|S_t, t) \, ds_T, \tag{3.1}
\]

where \( C \) is the European call option price, \( K \) is the strike price, \( r \) is the risk free rate, \( S_t \) is the underlying asset price at \( t \), \( f(\cdot) \) is the undiscounted risk neutral density of the future asset price \( S_T \) with current date \( t \) and option maturity date \( T \).

Take partial derivative with respect to the strike price \( K \):

\[
\frac{\partial C}{\partial K} = (s_T - K)f(s_T)|_{K=s_T} + e^{-r(T-t)} \int_K^\infty \frac{\partial(s_T - K)f(s_T)}{\partial K} \, ds_T \\
= -e^{-r(T-t)} \int_K^\infty f(s_T) \, ds_T. \tag{3.2}
\]
This generates the discounted cumulative density function of $S_T$. Take another partial derivative with respect to $K$, we have

$$\frac{\partial^2 C}{\partial K^2} = e^{-r(T-t)} f(s_T) \bigg|_{K=s_T} = e^{-r(T-t)} f(s_T). \quad (3.3)$$

Therefore by differentiating twice with respect to the strike price we get risk neutral density function:

$$f(s_T) = e^{r(T-t)} \frac{\partial^2 C}{\partial K^2}. \quad (3.4)$$

This relationship shows a trivial way to derive risk neutral density function using numerical methods. As we have already discussed above, this method has several drawbacks and is not favorable to use in practice. However it provides theoretical backgrounds in this area and inspires many other methods thereafter.

### 3.3.2 Black-Sholes Model

According to the famous Black-Sholes Model by Black and Scholes [1973] and Merton [1974], the underlying asset price $S_t$ follows a geometric Brownian motion with constant drift $\mu$ and volatility $\sigma$ while $S_t$ is continuous in $[0, +\infty)$:

$$dS_t = \mu S_t dt + \sigma S_t d\omega^P_t, \quad (3.5)$$

where $\omega^P_t$ is a Wiener process under the real probability measure $\mathbb{P}$. Apply Ito’s lemma and we could verify that asset price $S_T$ at maturity follows a log-normal distribution:

$$\ln S_T \sim N \left[ \ln S_t + (\mu - \frac{\sigma^2}{2})(T - t), \sigma^2(T - t) \right]. \quad (3.6)$$

According to the first fundamental theorem of asset pricing, in a complete market there must exist an equivalent risk-neutral probability measure $\mathbb{Q}$ such that present value of any asset equals to discounted expected value by risk free rate under $\mathbb{Q}$. It has been shown by Harrison and Pliska [1981] that risk neutral density has a similar form. This can also be derived from Black-Scholes model.

For a European style call option, its price formula is:

$$C_B(S_t, K, T, t) = S_t \Phi(d_1) - Ke^{-r(T-t)} \Phi(d_2), \quad (3.7)$$
where $C_B$ is option price under Black-Scholes model, $\Phi(\cdot)$ is the cumulative distribution function for standard normal distribution and
\[
\begin{align*}
d_1 &= \frac{\ln \left( \frac{S_T}{K} \right) + (r + \frac{\sigma^2}{2})(T - t)}{\sigma \sqrt{T - t}}, \\
d_2 &= \frac{\ln \left( \frac{S_T}{K} \right) + (r - \frac{\sigma^2}{2})(T - t)}{\sigma \sqrt{T - t}}.
\end{align*}
\]

From Section 3.3.1 we know that risk neutral density could be derived by taking second partial derivative from option price function $C_B$ with respect to the strike price $K$:
\[
f(s_T) = e^{r(T-t)} \frac{\partial^2 C_B}{\partial K^2} \bigg|_{K=s_T} = \frac{1}{s_T \sqrt{2\pi \sigma^2(T-t)}} \exp \left[ - \frac{\left[ \ln s_T - (\ln S_t + r - \frac{\sigma^2}{2})(T-t) \right]^2}{2 \sigma^2(T-t)} \right]. \tag{3.8}
\]

Therefore under Black-Scholes model risk neutral density $f(s_T)$ also follows a log-normal distribution with mean $\mu^* = (\ln S_t + r - \frac{\sigma^2}{2})(T-t)$ and variance $\sigma^*^2 = \sigma^2(T-t)$. This result seems appealing at first but lacks flexibility in practice due to the strong assumptions by Black-Scholes model.

### 3.3.3 Mixed Log-normal Distribution

The most popular parametric method to estimate risk neutral density is to use a linear combination of multiple log-normal distributions. It can be seen as an extension of the single log-normal result from Black-Sholes model. This method is firstly presented by Bahra (1997) and Melick and Thomas (1997).

Consider a series of log-normal density functions $f_i(x|\mu_i, \sigma_i)$, for $i = 1, \ldots, d$ as in Equation (3.8). Construct a new mixture distribution function $f(x)$:
\[
f(x|\theta) = \sum_{i=1}^{d} p_i f_i(x|\mu_i, \sigma_i^2), \tag{3.9}
\]
where $\theta = (p_1, \mu_1, \sigma_1, \ldots, p_d, \mu_d, \sigma_d)$, $0 \leq p_i \leq 1$ and $\sum_{i=1}^{d} p_i = 1$. Now use Equation (3.9) and plug it in Equation (3.1), we get the option price function $C_M$ for mixture
log-normal model:

\[ C_M(K, \theta_t, t) = e^{-r(T-t)} \int_K^\infty (s_T - K) f(s_T | \theta_t) \, ds_T \]

\[ = (E[S_T | S_T > K] - K) P[S_T > K] \]

\[ = e^{-r(T-t)} \sum_{i=1}^{d} \left[ p_{i,t} e^{\mu_{i,t} + \frac{\sigma_{i,t}^2}{2}} \Phi\left( \frac{\mu_{i,t} + \sigma_{i,t}^2 - \ln K}{\sigma_{i,t}} \right) - p_{i,t} K \Phi\left( \frac{\ln K - \mu_{i,t}}{\sigma_{i,t}} \right) \right], \]

(3.10)

where \( \Phi(\cdot) \) is the cumulative probability function of standard normal distribution and \( \Phi^c(\cdot) = 1 - \Phi(\cdot) \). Notice that the first fundamental theorem of asset pricing states in risk-neutral world discounted price process is a martingale, then we have additional constraint to be satisfied:

\[ S_t = e^{-r(T-t)} E[S_T] = e^{-r(T-t)} \sum_{i=1}^{d} p_{i,t} e^{\mu_{i,t} + \frac{\sigma_{i,t}^2}{2}}. \]

(3.11)

To estimate the parameter vector \( \theta_t = (\mu_1, \sigma_1, \mu_2, \sigma_2, \ldots, \mu_d, \sigma_j, p_1, \ldots, p_d) \), classic least square estimation method is used here. Denote the number of option strike prices as \( m \), observed call option price as \( C_{t,i} \), then \( \theta_t \) can be estimated through minimizing the following nonlinear least square function for every time \( t \):

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{m} (C_{t,i} - C_M(K_i, \theta_t, t))^2 \\
\text{subject to} & \quad S_t = e^{-r(T-t)} \sum_{i=1}^{d} p_{i,t} e^{\mu_{i,t} + \frac{\sigma_{i,t}^2}{2}} \\
& \quad \sum_{j=1}^{d} p_{j,t} = 1, p_{j,t} \geq 0, \sigma_{j,t} > 0.
\end{align*}
\]

(3.12)

When the number of components \( d \) becomes larger, Equation (3.12) becomes a very complicated non-linear optimization problem. When \( d = 2 \) or \( d = 3 \), this is the model suggested by Bahra (1997) and Melick and Thomas (1997). Comparing with log-normal density function derived from Black-Scholes model, mixed log-normal distribution does provide better flexibility. However, it still has a few drawbacks:

- Computation complexity for high dimensional non-linear optimizations.
- Symmetric parameters. It is difficult to determine the order of parameters in each densities. Additional constraints are required.
Cooper et al. (1999) and Bliss and Panigirtzoglou (2002) point out that when the estimated standard deviation $\sigma_i$ is small, risk neutral density generates unexpected spikes.

3.4 Method

3.4.1 Model

Inspired by the mixed log-normal distribution in Section 3.3.3, we adopt a three-component mixed log-normal model for the implied risk neutral density function:

$$f(x|\theta) = \sum_{i=1}^{3} p_i f_i(x|\mu_i, \sigma_i^2),$$

(3.13)

where $\theta = (p_1, \mu_1, \sigma_1, p_2, \mu_2, \sigma_2, p_3, \mu_3, \sigma_3)$, $0 \leq p_i \leq 1$, $\sum_{i=1}^{3} p_i = 1$ and $f(x|\mu, \sigma^2)$ is the density function of log-normal distribution with parameters $\mu$ and $\sigma^2$. To avoid the symmetric parameter issue in optimization, we put an additional constraint on their means $\mu_1 \leq \mu_2 \leq \mu_3$.

We view daily implied risk neutral densities $f_t(s_T|\theta_t)$ as a distributional time series with feature process $\theta_t$ and assume that $f_t(\cdot)$ only depends on the past state $f_{t-1}(\cdot)$ through $\theta_{t-1}$. Thus the DTS-FS model discussed in Section 2.1.1 could be applied here. Notice that option price $C_t$ is determined by its underlying risk neutral density $f_t$. Therefore now we could treat the risk-neutral framework as a state-space model that observed option price $C_t$ is driven by the feature process $\theta_t$:

$$\theta_t = s_t(\theta_{t-1}, \epsilon_t)$$

$$C_t = h_t(\theta_t, \epsilon_t),$$

(3.14)

where $\epsilon_t$ and $\epsilon_t$ are Gaussian white noise, $h_t(\cdot)$ is described in Equation (3.10).

Here we propose a generalized two-stage joint estimation framework (least square and Sequential Monte Carlo estimation) to estimate daily implied risk neutral density in the following steps:

1. Least Square Estimation:

Obtain classic least square estimates $\theta_{L,t}$ for every $t$ ($t = 1, 2, \ldots, T$) individually of the feature process through non-linear optimization using Equation (3.12).
2. **Time Series Analysis:**

Use linear time series models and autoregressive conditional heteroskedastic models to approximate the dynamics of $\theta_t$ and determine the state transition function $s_t(\cdot)$ in Equation (3.14) using the estimated $\theta_{L,t}$ obtained in the first step.

3. **Parameter Estimation:**

Apply Sequential Monte Carlo parameter estimation on the state-space model in Equation (3.14) to calibrate parameters in $s_t(\cdot)$. Then use the tuned parameters to obtain the filtered and smoothed estimates of the implied risk neutral density at each time $t$ if needed. Technical details have been discussed in Section 2.1.4.3.

Notice that both least square optimization of Equation (3.12) and Sequential Monte Carlo parameter estimation are non-linearly complicated and computationally cumbersome, which calls for more advanced computing hardware and algorithms. Thus parallel computing is introduced for all non-linear optimizations and Sequential Monte Carlo calculations in this dissertation.

### 3.4.2 Sequential Monte Carlo with Parallel Computing

A central limitation to all Monte Carlo simulation methods is the power of computation. To achieve better accuracy and lower variance of the estimators, more Monte Carlo samples and more intensive calculations are often required. Traditionally computation software is programmed serially such that a problem can be broken into a series of discrete subproblems and these subproblems are getting executed one at a time. While Moore’s law was still prevailing, computing performance could be easily improved by upgrading with higher frequency processor and faster storage media. However, nowadays semiconductor industry has slowed down the race of chasing processor clock speed and switched to integrate multiple separate processing units instead. The decades-long increasing pattern of chip clock rate has flattened in recent years due to both material and technical limitations. At the same time, huge amount of data are generated everyday in the current information explosion era through activities such as social media, mobile network and Internet of Things. Traditional serially designed programs could
no longer satisfy the increasing computing needs for modern scientific and industrial
data. Therefore the idea of parallel computing has been introduced.

Simply taking, parallel computing uses multiple resources simultaneously to solve a
problem. Here resources typically refer to CPU and GPU cores or even computers in a
local area network. To efficiently utilize these resources, the target problem itself must
satisfy certain specific conditions so that it could be broken into multiple concurrent
subproblems to be executed at the same time. Thus Sequential Monte Carlo method
discussed in Section 2.1.4.2 is ideal for parallel computing. For example, in particle
filtering, each particle can be viewed as a separate thread that bears several complicated
computation tasks such as sampling new particle, calculating importance weights and
resampling particles. Those computing tasks are exactly the same for each individual
particle and can be conducted simultaneously at time $t$. Similarly, in optimization
problems parallel computing can also be applied so that multiple evaluations could
occur concurrently.

In this dissertation, we implement CUDA (Compute Unified Device Architecture),
a very powerful and popular parallel computing platform and application programming
interface invented by NVIDIA, to speed up calculation and improve accuracy. Un-
like traditional CPU-based only computing techniques, CUDA also harnesses power of
NVIDIA GPU to facilitate parallel computation. Structure of a CUDA program is illus-
trated in Figure 3.1. After data is transferred to device GPU memory, CPU can invoke
GPU to parallely execute all computations through so-called CUDA kernel functions.
Thus it requires a thorough understanding and parallel design of the entire program
beforehand.
A simple vector addition example is conducted to evaluate the performance boost comparing with regular CPU based serial computing programs. Table 3.1 lists details of our testing environment. For hardware, both are last generation consumer flagship products released between 2016-2017 so they are somewhat on par with each other. The example is very short and simple: we conduct addition 100 times of two randomly generated vectors of length 10,000 by those two methods in the same C++ program. All numbers are randomly generated from standard uniform distribution. For CPU computing, we use std::transform which is essentially an optimized map function. For GPU computing, we use the built-in vectorAdd example in CUDA samples. Figure 3.2 shows the performance evaluation result. It can be seen that computing time of GPU based CUDA parallel method basically stays the same with few perturbations throughout the whole experiment and does not depend on vector size at all. The reason is that for simple basic addition which does not invoke additional calculations, increased vector size will only increase the number of cores GPU used but not computing time. On the other hand computing time of CPU based method increases almost perfectly linearly with the increased vector size. When vector size reaches 120,000, computing speed of GPU based CUDA method is nearly 30 times faster than that of traditional CPU computing. Therefore for parallelized problems, the more parallelism we introduce to our algorithm and program, the more efficiency we could gain from parallel computing.
Table 3.1: Environment

<table>
<thead>
<tr>
<th></th>
<th>CPU</th>
<th>GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware</td>
<td>Intel i7-7700k</td>
<td>Nvidia GeForce GTX 1080 Ti</td>
</tr>
<tr>
<td>Resources</td>
<td>4 CPU Cores</td>
<td>3584 CUDA cores</td>
</tr>
<tr>
<td>Frequency</td>
<td>4.20 GHz</td>
<td>1,481 MHz</td>
</tr>
<tr>
<td>Interface</td>
<td>C++ 11</td>
<td>CUDA 10.0</td>
</tr>
<tr>
<td>Compiler</td>
<td>VC++ 2017</td>
<td>nvcc 10.0</td>
</tr>
<tr>
<td>IDE</td>
<td>Visual Studio 2017</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>C++</td>
<td></td>
</tr>
</tbody>
</table>

Therefore, for all Sequential Monte Carlo related projects in this thesis, we redesign all computation algorithms in a parallel way to facilitate CUDA computing. As mentioned above, since all calculations are exactly the same for each particle, they can be natively translated into parallel execution. For every Sequential Monte Carlo calculation in Algorithm 2 in Section 2.1.4.2 separate kernel functions are written for particle generation, weight calculation and resampling. A simple non-linear particle filter toy example, inspired by Cappe et al. (2007), is presented below to demonstrate the effectiveness of parallel computing application in Sequential Monte Carlo calculations.
Suppose the underlying state space model is

\[ X_t = X_{t-1} + \frac{4X_{t-1}}{1 + X_{t-1}^2} + 3\cos(t) + u_t \]

\[ Y_t = \frac{X_t^2}{20} + v_t, \tag{3.15} \]

where \( u_t \sim N(0, 4) \) and \( v_t \sim N(0, 1) \). With standard particle filter, one can easily calculate the sampling distribution \( g_t(x_t^{(i)}|x_{t-1}^{(i)}) \) and incremental weight \( u_t^{(i)} \) as follows:

\[ g_t(x_t^{(i)}|x_{t-1}^{(i)}) = N\left(x_t^{(i)}\bigg| x_{t-1}^{(i)} + \frac{4x_{t-1}^{(i)}}{1 + x_{t-1}^{(i)^2}} + 3\cos(t), 4\right) \]

\[ u_t^{(i)} = N\left(y_t\bigg| \frac{x_t^{(i)^2}}{20}, 1\right). \tag{3.16} \]

Under similar testing environment as the previous vector addition example, we run standard particle filter 100 times each with different Monte Carlo sample sizes and record system running time for both GPU and CPU programs. The result is presented in Figure 3.3. It can be seen again that GPU based parallel programming dominates CPU based serial programming performance-wise. When Monte Carlo sample size is greater than 20,000, its running time stays nearly unchanged. At the same time, CPU running time continues linearly increasing. Thus the larger Monte Carlo size employed, the greater computing time will be saved and better accuracy in Sequential Monte Carlo estimations could be achieved.

Figure 3.3: Particle Filter Performance Comparison
3.5 Simulation Study

We first conduct a simulation study to demonstrate the two-stage estimation framework previously discussed in Section 3.4.1. The simulation model settings are enlightened by the empirical study of real data that will be discussed shortly in Section 3.6.

3.5.1 Data Simulation

We select the modified three-component mixed log-normal distribution with a point mass term in Equation (3.18) as the true risk neutral density $f_t(\cdot)$. For its feature process $\theta_t = (\mu_{2,t}, \sigma_{2,t}, p_{2,t}, \mu_{3,t}, \sigma_{3,t}, p_{3,t})$, we apply the following time series models as their underlying dynamics:

$$
\begin{align*}
\mu_{2,t} &= \mu_{2,t-1} + a_t + \phi_1 a_{t-1}, a_t \sim N(0, 0.04^2) \\
\text{logit}(p_{2,t}) &\sim N(-2, 0.1^2) \\
\mu_{3,t} &= \phi_2 \mu_{3,t-1} + (1 - \phi_2) \mu_{3,t-2} + e, e \sim N(0, 0.008^2) \\
\sigma_{2,t} &\sim N(0.01, 0.0001^2) \\
\sigma_{3,t} &= -0.0003 + \sigma_{3,t-1} + a_t + \phi_3 a_{t-1}, a_t \sim N(0, 0.005^2) \\
\text{logit}(p_{3,t}) &= 2.8482 + \phi_4 \sqrt{t} + a_t, a_t \sim N(0, 0.05^2),
\end{align*}
$$

(3.17)

where $\phi_S = (\phi_1, \phi_2, \phi_3, \phi_4)$ are simulation parameters with values listed in Table 3.2.

We choose SPDR S&P 500 ETF (NYSE Symbol: SPY) as the target asset and use its asset price and strike price data between 06/2014 and 12/2014 for simulation. These are the same data used in the empirical study in Section 3.6. The descriptive statistics and asset price pattern are presented in Table 3.3 and Figure 3.7. The procedure of option price simulation is

1. Obtain simulated feature process $\theta_{S,t}$ using Equation (3.17) within the given time period.

2. Calculate the true underlying daily risk neutral density $f_{S,t}(\cdot)$ in Equation (3.18) using the simulated feature process $\theta_{S,t}$. 


3. Simulate SPY option prices using Equation (3.10) with daily risk neutral densities \( f_{S,t}(\cdot) \) and corresponding strike prices. We also add a Gaussian white noise error term.

Figure 3.4a shows a three-dimensional plot of simulated daily risk neutral densities \( f_{S,t}(\cdot) \). The red colored line segments at point 120 of the \( x \)-axis indicate the point mass terms in Equation (3.18) whereas the blue colored curves represent the two-component log-normal distribution kernel in Equation (3.18).

Table 3.2: Parameter Estimation Result

<table>
<thead>
<tr>
<th></th>
<th>( \phi_1 )</th>
<th>( \phi_2 )</th>
<th>( \phi_3 )</th>
<th>( \phi_4 )</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>-0.9</td>
<td>0.5</td>
<td>-0.8</td>
<td>-2.3518</td>
<td>-11798.3</td>
</tr>
<tr>
<td>Least Square Estimation</td>
<td>-0.9528</td>
<td>-0.4796</td>
<td>-0.9105</td>
<td>-3.1788</td>
<td>-11850.9</td>
</tr>
<tr>
<td>SMC Parameter Estimation</td>
<td>-0.9336</td>
<td>0.5720</td>
<td>-0.6351</td>
<td>-2.1805</td>
<td>-11796.9</td>
</tr>
</tbody>
</table>

3.5.2 Least Square Estimation

We perform non-linear optimization on Equation (3.12) to obtain least square estimator \( \hat{\theta}_{L,t} \) for every \( t \). Figure 3.5 shows comparison of time series plot of simulated and least square estimated feature process parameter vector \( \theta_t = (\mu_{2,t}, \sigma_{2,t}, p_{2,t}, \mu_{3,t}, \sigma_{3,t}, p_{3,t}) \) respectively. Black lines in each plot are the simulated time series of parameters in \( \theta_t \) and blue lines are the time series of least square estimated parameters in \( \theta_{L,t} \). In Figures 3.5c and 3.5d, least square estimated parameters \( p_{L,3} \) and \( \sigma_{L,3} \) have several huge spikes possibly due to difficulties in high-dimensional non-linear optimizations. Figure 3.4b presents the three-dimensional plot of implied daily risk neutral densities \( f_{L,t}(\cdot) \) calculated from least square estimated feature process \( \theta_{L,t} \). They have similar pattern as the simulated true daily risk neutral densities do in Figure 3.4a except for fewer peaks in density curves. Those disappeared modes are the result from the occasional spikes in least square estimated \( p_{L,3} \) in Figure 3.5c.
Figure 3.4: Risk Neutral Density - Simulation

(a) Risk Neutral Density - Simulation

(b) Risk Neutral Density - Least Square Estimation

(c) Risk Neutral Density - SMC Estimation
Figure 3.5: Times Series Plots Comparison of $\hat{\theta}_t$, $\hat{\theta}_{L,t}$ and $\hat{\theta}_{M,t}$
3.5.3 Sequential Monte Carlo Parameter Estimation

Next we implement Sequential Monte Carlo parameter estimation to tune parameters $\phi_S$ in the time series models in Equation (3.17) of feature process $\hat{\theta}_{L,t}$. The method and algorithm have been discussed in Section 2.1.4.3. To reiterate, we calibrate parameters $\phi_S$ through maximum likelihood estimation of the log-likelihood of the state-space model in Equation (3.14), which is the arithmetic mean of the unnormalised weight as demonstrated in Equation (2.24). Then we use the calibrated parameters $\phi_M$ to generate filtered feature process $\theta_{M,t}$ and filtered daily risk neutral density $f_{M,t}(\cdot)$.

Table 3.2 shows SMC parameter estimation tuned parameters $\phi_M$, least square estimation $\phi_L$ and simulated true value $\phi_S$. It can be seen that Sequential Monte Carlo parameter estimation provides superior results than least square estimation does. $\phi_M$ is much closer to the true value $\phi_S$ with a larger likelihood than $\phi_L$ is.

In Figure 3.5 the red colored lines are the time series plot of the filtered feature process parameters $\theta_{M,t} = (\mu_{M,2,t}, \mu_{M,3,t}, p_{M,3,t}, \sigma_{M,3,t})$ generated with SMC tuned parameters $\phi_M$. It can be seen in Figure 3.5 that parameters $\mu_{M,3,t}, p_{M,3,t}, \sigma_{M,3,t}$ are closer to the simulated $\theta_t$ and smoother in their time series plots than $\mu_{L,3,t}, p_{L,3,t}, \sigma_{L,3,t}$ are. The only exception is $\mu_{M,2,t}$ in Figure 3.5a. It drifts about 2% down from the simulated $\mu_{2,t}$, which is also illustrated in Figure 3.4c as the small modes around 140 in the $x$-axis on three-dimensional plot of daily risk neutral densities calculated by $\theta_{M,t}$.

3.5.4 Sequential Monte Carlo Prediction

We divide the simulation data evenly into two datasets. First dataset is used as training set for Sequential Monte Carlo parameter estimation. After parameter calibration, we perform one day ahead Sequential Monte Carlo prediction of the option price as discussed in Section 2.1.4. For least square estimated risk neutral densities, we apply a trivial random walk prediction model by using the risk neutral density from previous trading day to calculate option prices of today. We evaluate the prediction performance by comparing the mean squared error of predicted option prices from our two-stage Sequential Monte Carlo joint estimation model (SMC) and the least square based random
walk model (LS-RW).

Figure 3.6a shows mean squared errors of option price predictions grouped by strike prices. The blue colored time series line represents option price predictions of the two-stage SMC joint estimation model and the red colored time series line represents predictions from the least square based random walk model. It is obvious that SMC prediction dominates random walk prediction at every strike price. When strike price increases, both mean squared errors from two models surge but errors from SMC prediction grow much slower.

Figure 3.6b presents mean squared errors of option price predictions grouped by option moneyness. Here moneyness is the ratio of a derivative’s strike price over its underlying asset price. We can see that the prediction performance of our two-stage SMC joint estimation model still dominates that of the random walk model. Also for both prediction models, in-the-money option price prediction performances are better than out-of-the-money option price prediction performances.

Therefore, based on the above results of the simulation study, our two-stage SMC joint estimation model not only increases the flexibility and smoothness of estimated risk neutral density but also enhances prediction power greatly comparing with the least square estimation model.

![Figure 3.6: Prediction Performance Comparison by MSE - Simulation](image-url)
3.6 Empirical Study

3.6.1 Data Description

We select SPDR S&P 500 ETF (NYSE Symbol: SPY) as our target asset. It tracks the S&P 500 stock market index and is one of the largest ETF in the world. The associated options therefore have very high liquidity and small bid-ask spread. We pick about 50 call options with strike price in consecutive integers, roughly 30 of them are in-the-money ones and 20 of them are out-of-the-money ones. The options are all set to expire on Friday of the third week of December. We consider two different sets of data in year 2013 and in year 2014 for comparison and collect daily option prices from the beginning of June to their expiration dates, with 142 trading days in each year. A summary of the data is given in Table 3.3. The corresponding 12-month U.S. Treasury Bill is used as risk free asset. All data are extracted from Bloomberg Terminal. Note there are some missing data in specific strike prices of some days due to liquidity issues.

<table>
<thead>
<tr>
<th>Year</th>
<th>Expiration</th>
<th>SPY Price Range</th>
<th>Trading Days</th>
<th>Options</th>
<th>Strike Range</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>12/19</td>
<td>186.3-208.0</td>
<td>142</td>
<td>51</td>
<td>$170-$219</td>
<td>4356</td>
</tr>
<tr>
<td>2013</td>
<td>12/21</td>
<td>157.1-181.7</td>
<td>142</td>
<td>41</td>
<td>$150-$190</td>
<td>4691</td>
</tr>
</tbody>
</table>

Figure 3.7 shows the asset price series for SPDR S&P 500 ETF in the same time frame in 2013 and 2014 year. It is seen that SPY prices in both years increase steadily except for several slumps. The similarity in SPY prices chart implies that the dynamics of implied risk neutral densities might also be similar for both years. To simplify, in the following sections we will only demonstrate results for year 2014 unless specified.

Figure 3.8 shows the option price time series for year 2014 of all strikes. We can see that in-the-money-options take account of about 60% of all observations and there are a few gaps of missing data mostly for out-of-the-money options due to illiquidity issues.
3.6.2 Least Square Estimation

The proposed three-component mixed log-normal distribution in Equation (3.13) is used as basis for risk neutral density function and all the parameters of the feature process $\theta_t$ are fitted using least squares estimation as shown in Section 3.3.3. All non-linear optimization problems in this chapter are solved by using a modified Lipschitzian optimization algorithm proposed by Jones et al. (1993). Notice that in Equation (3.10) of option price calculation, whenever $\sigma_{i,t}$ is sufficiently small and $\mu_{i,t} < \ln K$, the integration of corresponding log-normal component will be approximately 0. Thus after least squares estimation we find that most of the time the fitted implied daily risk neutral density function actually collapses into a two-component mixed log-normal density with a point mass $p_{1,t}$. It is then more convenient to directly work with a two-component log-normal kernel with less parameters and comparable explanatory power instead. We propose a modified three-component mixed log-normal distribution:

$$S_T \sim f_t(x|\theta_t) = p_{1,t}\delta(x-x_0) + p_{2,t}f_{2,t}(x|\mu_{2,t},\sigma_{2,t}^2) + p_{3,t}f_{3,t}(x|\mu_{3,t},\sigma_{3,t}^2),$$  (3.18)
Figure 3.8: SPY Call Option Price - Year 2014
Figure 3.9: Leaset Square estimated Risk Neutral Density Plot - Year 2014

where \( \delta(\cdot) \) is the Dirac delta function with \( x_0 < \min K \), \( f(x|\mu, \sigma^2) \) is the density function of log-normal distribution with parameters \( \mu \) and \( \sigma^2 \), \( \mu_{2,t} \leq \mu_{3,t}, \sigma_{2,t}, \sigma_{3,t} > 0, 0 \leq p_{1,t}, p_{2,t}, p_{3,t} \leq 1 \) and \( p_{1,t} + p_{2,t} + p_{3,t} = 1 \) and \( \theta_t = (p_{2,t}, \mu_{2,t}, \sigma_{2,t}, \mu_{3,t}, \sigma_{3,t}) \). Note \( p_{1,t} \) is not included in \( \theta_t \) since we have a linear restriction of \( p_{1,t} + p_{2,t} + p_{3,t} = 1 \). The first component in Equation (3.18) only takes value at \( x_0 \), thus its corresponding option price calculation in Equation (3.10) will always be 0.

Plug Equation (3.18) in Equation (3.12) and we obtain the least square estimator \( \theta_{L,t} \) of the feature process \( \theta_t \) through non-linear optimization. Next we can calculate the daily risk neutral density functions \( f_{L,t} \) by using \( \theta_{L,t} \). Figure 3.9 shows a three-dimensional plot of least square estimated risk neutral density evolving over time \( t \). Red colored line segments stand for the point masses for parameter \( p_{L,1,t} \). They are set at \( x_0 = 160 \) with heights as value of \( p_{L,1,t} \). Blue colored curves represent the two-component log-normal density funtion in Equation (3.18), with two modes being identified as two log-normal kernels. In some days, \( \mu_{L,2,t} \) related log-normal component
has very small standard deviation with huge spikes. This observation has been pointed out in previous studies by Cooper et al. (1999) and Bliss and Panigirtzoglou (2002). In some days, the mode of \( f_{L,2,t}(\cdot) \) gets overshadowed by the other log-normal component \( f_{L,3,t}(\cdot) \) when means of the two log-normal kernels are close enough to each other.

### 3.6.3 Determination of State Dynamics

After obtaining the least square estimator \( \theta_{L,t} = (p_{L,2,t}, \mu_{L,2,t}, \sigma_{L,2,t}, p_{L,3,t}, \mu_{L,3,t}, \sigma_{L,3,t}) \) for the feature process that drives risk neutral densities for each day \( t \), we apply linear time series analysis individually to every parameter series. Figure 3.10 shows time series plots for each parameter in \( \theta_{L,t} \). All six parameters pose somewhat time series trends except for \( \sigma_{L,2,t} \), which is later determined as a Gaussian white noise series with a few huge spikes. When time \( t \) approaches option expiration date \( T \), least square estimation shows that \( \sigma_{L,3,t} \rightarrow 0 \) and \( p_{L,3,t} \rightarrow 1 \). This is intuitive since implied risk neutral density would shrink into the classic log-normal distribution as suggested by Black-Scholes model.

Upon further checking and refinement we use simple time series autoregressive-moving-average (ARMA) model to fit those parameters except \( \sigma_{L,2,t} \), whose pattern is purely random and difficult to model using time series analysis. It is then assumed to be a Gaussian white noise series. The specific models and their estimated parameters are listed below. The standard deviation of the residuals of each model presented is based on the data of year 2014.

\[
\begin{align*}
\mu_{L,2,t} &= \mu_{L,2,t-1} + a_t \mu_2 + \phi_1 a_{t-1} \mu_2, \hat{\sigma}_{a_t \mu_2} = 0.03165 \\
p_{L,2,t} &= \phi_2 + p_{L,2,t-1} + a_t p_2 + \phi_3 a_{t-1} p_2, \hat{\sigma}_{a_t p_2} = 0.09572 \\
\sigma_{L,2,t} &\sim \mathcal{N}(0.02769, 2.9637 \times 10^{-4}) \\
\mu_{L,3,t} &= \phi_4 \mu_{L,3,t-1} + (1 - \phi_4) \mu_{L,3,t-2}, \hat{\sigma}_{\mu_3} = 0.008746 \\
\sigma_{L,3,t} &= -0.0003 + \sigma_{L,3,t-1} + a_t \sigma_3 + \phi_5 a_{t-1} \sigma_3, \hat{\sigma}_{a_t \sigma_3} = 0.005718 \\
p_{L,3,t} &= \phi_6 + p_{L,3,t-1} + a_t p_3 + \phi_7 a_{t-1} p_3, \hat{\sigma}_{a_t p_3} = 0.1004.
\end{align*}
\]
Figure 3.10: Times Series Plots of $\hat{\theta}_{L,t}$ - Year 2014
Table 3.4: Time Series Model Coefficients $\phi_L$ for Year 2013 and 2014

<table>
<thead>
<tr>
<th>Year</th>
<th>$\phi_{L,1}$</th>
<th>$\phi_{L,2}$</th>
<th>$\phi_{L,3}$</th>
<th>$\phi_{L,4}$</th>
<th>$\phi_{L,5}$</th>
<th>$\phi_{L,6}$</th>
<th>$\phi_{L,7}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>-0.7639</td>
<td>0</td>
<td>-0.9476</td>
<td>0.5073</td>
<td>-0.8105</td>
<td>0</td>
<td>-0.9024</td>
</tr>
<tr>
<td>2013</td>
<td>-0.8177</td>
<td>-0.0015</td>
<td>-0.9319</td>
<td>0.5334</td>
<td>-0.8683</td>
<td>0.0028</td>
<td>-0.9025</td>
</tr>
</tbody>
</table>

It is interesting to notice that the best time series models are the same for both year 2013 and 2014. All five parameters $\mu_{L,2,t}, \mu_{L,3,t}, p_{L,2,t}, \sigma_{L,3,t}, p_{L,3,t}$ in the feature process follow simple Autoregressive-moving-average (ARMA) models. All the model coefficients are statistically significant with Gaussian residuals. Table 3.4 shows that the estimated coefficients $\phi_L = (\phi_{L,1}, \phi_{L,2}, \phi_{L,3}, \phi_{L,4}, \phi_{L,5}, \phi_{L,6}, \phi_{L,7})$ are relatively close to each other for the two years.

### 3.6.4 Sequential Monte Carlo Parameter Estimation

Next we use Sequential Monte Carlo parameter estimation to calibrate $\phi_L$ through the state-space model described in Equation (3.14). The state transaction function $s_t(\cdot)$ is the time series models in Equation (3.19) and observation equation $h_t(\cdot)$ is Equation (3.10). We apply standard particle filter here as described in Section 2.1.4.2. Technical details of parameter estimation have been discussed in Section 2.1.4.3.

After obtaining optimal parameters $\phi_M$ for time series models Equation (3.19), we use them back in Sequential Monte Carlo estimation to calculate filtered feature process $\theta_{M,t}$ and daily risk neutral densities $f_{M,t}$. Figure 3.11 shows time series plots of both $\theta_{L,t}$ from least square estimation (blue colored series) and $\theta_{M,t}$ from our two-stage SMC joint estimation (red colored series). Overall the time series of our Sequential Monte Carlo estimated parameters in $\theta_{M,t}$ are more smooth and stable with less spikes than those of least square estimated ones in $\theta_{L,t}$ are. Figure 3.12 shows a three-dimensional plot of the implied daily risk neutral densities generated by our two-stage SMC joint estimation model. Red colored line segments are the point masses of the first component in Equation (3.18). Blue curves represent the two-component log-normal kernels in Equation (3.18) with two clearly visible modes. Comparing with the implied daily risk
Figure 3.11: Times Series Plots of $\theta_{L,t}$ and $\theta_{M,t}$ - Year 2014
Figure 3.12: Risk Neutral Density Plot for SMC Fit - Year 2014

neutral densities generated by least square estimation in Figure 3.9, Sequential Monte Carlo estimated risk neutral density curves are smoother with no huge spikes.

3.6.5 Sequential Monte Carlo Prediction

We conduct option price predictions using three models:

1. Trivial random walk model (RW) that uses option prices of previous trading day as predictions for today.

2. Least square estimation based random walk model (LS-RW) that uses least square estimated risk neutral density in previous trading day to calculate option prices for today.

3. Sequential Monte Carlo one-day ahead prediction model (SMC-PM) with parameters tuned by our proposed two-stage SMC joint estimation model.

Unlike that in the simulation study, a different rolling prediction scheme is designed
Figure 3.13: Option Price Prediction Daily MSE Comparison - Year 2014
Figure 3.14: Option Price Prediction MSE Comparison by Strike - Year 2014
here. We use first two months (June and July) option data as training set to run Sequential Monte Carlo parameter estimation and then use tuned parameters to perform out-of-sample option price predictions on next month (August). We repeat this process and obtain monthly predictions of year 2014. The predictions of December are omitted due to illiquidity. We evaluate the prediction performance by comparing the mean squared errors of the predicted option prices, same as we do in the simulation study.

![Graph showing option price prediction MSE comparison by moneyness.

(a) In-the-money Option Prediction   (b) Out-of-the-money Option Prediction

Figure 3.15: Option Price Prediction MSE Comparison by Moneyness - Year 2014

Figure 3.13 shows daily option price predictions of the three models in terms of mean squared error for four consecutive months from August to November. It is seen that RW and LS-RW have almost the same prediction mean squared errors while SMC-PM delivers an uneven prediction performance. For example, among 11 out of 21 trading days in August, SMC-PM predictions are more accurate than RW and LS-RW ones. However, there are also 4 days that SMC-PM predictions deviate greatly with largest mean squared error being 10 times as much as that of RW or LS-RW. Similar patterns are found in September and October. Note that for options that are close to the expiration date (from late August to the end of November), SMC-PM daily predictions achieve the same performance level as predictions of both RW and LS-RW.

Figure 3.14 shows prediction performance of LS-RW and SMC-PM in terms of mean squared error grouped by strike prices. We see that when strike price increases, prediction performances of both models plunge. Figure 3.15 presents prediction performance
Figure 3.16: Risk Neutral Density Plot for SMC Predictions - Year 2014

of LS-RW and SMC-PM in terms of mean squared error by option moneyness. Figure 3.15a shows in-the-money (ITM) option price predictions. We can see that SMC-PM outperforms LS-RW only for few deeply ITM options and its performance decreases when moving from in-the-money to at-the-money. Similar pattern is also observed for out-of-the-money (OTM) predictions in Figure 3.15b.

Figure 3.16 shows the three-dimensional plot of Sequential Monte Carlo predicted daily risk neutral densities \( \hat{f}_{M,t} \) from 08/2014 to 11/2014. Comparing with least square estimated risk neutral densities \( f_{L,t} \) in Figure 3.9 and risk neutral densities \( f_{M,t} \) estimated by our two-stage SMC joint estimation model in Figure 3.12, we can see that the second component of mixed log-normal distribution in Equation (3.18) is weaker in \( \hat{f}_{M,t} \) than that in \( f_{L,t} \) and \( f_{M,t} \), which might affect SMC prediction performance. A possible explanation to this issue could be the unsuccessfully modeled parameter \( \sigma_{2,t} \) in the feature process \( \theta_t \), which has been discussed in Section 3.6.3.
3.7 Conclusion

This chapter presents an application of the generalized two-stage Sequential Monte Carlo joint estimation framework to model dynamic risk neutral density from call option prices. This framework includes the following steps in general:

1. Identify the underlying feature process $\theta_t$ of functional time series and build a corresponding state-space model.

2. Obtain estimator $\hat{\theta}_t$ through classic statistical analysis such as least square estimation.

3. Determine dynamics of $\hat{\theta}_t$ with appropriate low-dimensional time series models.

4. Employ Sequential Monte Carlo parameter estimation to calibrate parameters in time series models through maximum likelihood estimation.

5. Carry out Sequential Monte Carlo estimation and prediction with tuned parameters.

We also integrate parallel computing algorithms and schemes with Sequential Monte Carlo calculations to make our framework computationally efficient.

In our application, we view the underlying risk neutral density as a FTS-FS and model it using a modified mixed log-normal distribution through a state-space model. A simulation study is conducted to demonstrate the robustness and effectiveness of our framework and the one-day ahead Sequential Monte Carlo prediction performance dominates that of the least square based random walk model. In the empirical study with real SPY call option price data, however, our two-stage estimation model delivers an uneven prediction performance comparing with random walk models. One possible explanation could be the failure of properly fitting a time series model for $\sigma_{2,t}$ in the feature process $\theta_t$. Even so, our proposed two-stage SMC joint estimation framework could be applied to many other parametric and non-parametric estimation models as long as a closed form of risk neutral density function can be obtained.
Chapter 4

Modeling Dynamic Cross-Sectional Distribution

4.1 Introduction

It has been a popular research topic in finance over the past several decades to study the cross-sectional effects of expected asset returns: why do different stocks earn different returns? The famous Capital Asset Pricing Model (CAPM) by Sharpe (1964) and Lintner (1965) expresses expected return as a simple and elegant linear function of the systematic risk of the individual asset. Despite its early successes, several researches, including Fama and MacBeth (1973), Basu (1977), Banz (1981), Basu (1983), Shanken (1985), Gibbons et al. (1989) and Fama and French (1992), later find that in practice CAPM fails many empirical cross-sectional data tests and provides nearly no explanatory power in certain cases, which indicates that there could be other non-risk characteristics involved. Banz (1981) finds out that stocks with different firm size have different risk adjusted returns. Basu (1977) and Basu (1983) show that both earning’s yield and firm size affect cross-sectional variance on risk adjusted stock returns. Bhandari (1988) conducts an empirical study and finds out debt-to-equity ratio also plays an important role affecting expected stock returns. Stattman (1980), Rosenberg et al. (1985) and Chan et al. (1991) respectively find and conclude that stocks with high book-to-market ratio earn higher expected returns than lower ones do. Easton and Zmijewski (1989) studies cross-sectional variation in stock returns after earnings releases and finds that earning response coefficient is negatively correlated with expected return. After examining and combining all these research work, Fama and French (1992) finds that cross-sectional variations of expected returns can be captured by two factors, which are related to firm size and book-to-market ratio, and proposes a three factor model (FF3FM) expanding CAPM. The two extra factors examine historic excess returns of
small capitalization firms over big ones and of high book-to-market ratio stocks over lower ones. Comparing with CAPM, FF3FM successfully captures cross-sectional variation in average returns and overall considerably increases the explanation power by explaining over 90% of asset returns.

Aside from the groundbreaking work of CAPM and FF3FM, numerous researches have been conducted to explain and study the cross-sectional effects of stock returns in different scenarios. Roll (1977) and Ross (1977) point out that if the whole market lies on the mean-variance efficient frontier, there exists a positive cross-sectional relationship between expected returns and betas. Breeden (1979) and Grossman and Shiller (1981) propose a consumption-based capital asset pricing model (CCAPM) by using covariance with aggregate consumption instead of market return. Jegadeesh and Titman (1993) finds that the profitability of trading strategies is not related to systematic risk and stocks tend to continue their previous performance in three to twelve months. Lakonishok et al. (1994) studies value strategies in stock market and shows that cross-sectional return patterns are reflections of economic regularity phenomenon. Roll and Ross (1994) argues that meanvariance inefficient market portfolio leads to the failure of CAPM. Kothari et al. (1995) analyzes cross-section of expected returns and reveals that it compensates for beta risk when beta is estimated from annual portfolio data. Fama and French (1995) shows that book-to-market ratio, profitability and investment effects are correlated with each other. For example, firms with higher book-to-market ratio tend to have lower profitability and investment. Chan and Chen (1988), Cochrane (1996) and Jagannathan and Wang (1996) try to explain the cross-section of expected returns using conditional versions of CAPM. Carhart (1997) builds a four-factor asset pricing model as an extension of FF3FM by adding a momentum factor. Here momentum is defined as the tendency for the stock price to continue its previous trend. Brennan et al. (1998) evaluates eight more non-risk characteristics in additional to Fama French factors under arbitrage pricing theory benchmark and finds that momentum, firm size and book-to-market ratio mostly affect cross-section of expected returns. Berk et al. (1999) develops an investment based equilibrium model of stock expected return to explain both cross-sectional and time series behavior of stock returns.
Jegadeesh and Titman (2002) studies momentum trading strategies and argues that cross-sectional variation in expected returns do account for profits but not significantly. Amihud (2002) studies both cross-sectional and time series effects on expected stock returns of expected market illiquidity and concludes that a positive relation exists. Titman et al. (2004) finds expected stock returns are negatively correlated with abnormal capital investments. Gomes et al. (2006) uses an investment-based conditional asset pricing model to evaluate whether financing frictions contribute to the cross-section of expected returns. Cooper et al. (2008) shows that a firm’s annual asset growth rate could be used to predict the cross-section of expected returns and it is statistically as important as traditional Fama French factors. Fang and Peress (2009) explores cross-sectional relation between mass media coverage and stock expected returns and find that stocks with less or no media coverage earn higher future returns. Stivers and Sun (2010) finds that time variation in the value and momentum premiums is positively correlated to cross-sectional dispersion in expected returns. Novy-Marx (2013) argues that profitability should be included in asset pricing model to predict cross-section of expected returns as on average profitable firms obviously earn higher returns. Hou et al. (2014) proposes a multi-factor asset pricing model that includes firm size, investment and profitability to explain cross section of average returns. Fama and French (2015a) and Fama and French (2015b) further extend their famous three factor model with two more factors: profitability and investment. They also point out that HML factor (high book-to-market ratio minus low) is actually redundant and could be fully explained by the other four factors under certain conditions. If dropping HML factor, the five-factor model is similar to the work by Hou et al. (2014). Harvey et al. (2015) documents 316 tested factors from previous researches on cross section of expected returns in the past decades and argue that new statistical testing standards need to be set up for evaluation in current financial world. Mclean and Pontiff (2016) compares 97 different factors in an empirical study and show that most explanation power of crosssectional predictability comes from mispricing.

Most of the researches mentioned above focus on the cross section effects of expected returns itself and how its variation could be explained. The underlying dynamics of
the distribution of cross-sectional returns and how it evolves through time receive little attention. Unlike a simple number of market index, the daily cross-sectional distribution of expected returns provides a much broader view of the entire market. By analyzing its statistical properties such as dispersion, tailedness and central tendency, we can extract valuable market information. For example, whenever the distribution is concentrated around mean or median it indicates a lower market risk and vice versa. Moreover, with complete understanding of the distribution, one can make better trading decisions. Conrad and Kaul (1998) compares 120 trading strategies in literature and presents striking evidence that it is cross-sectional variation in individual assets, instead of time series dependence, that determines their profitabilities. Especially for momentum strategies, estimated cross-sectional dispersion in the mean returns most consistently accounts for the observed profits. Therefore, a simple trading strategy could be established by monitoring cross-sectional dispersion: long stocks with higher return (in the right tail) and short ones with lower return (in the left tail). It works well whenever cross-sectional dispersion is large. This calls for analysis of the dynamics of daily cross-sectional distribution and making future predictions. Chen T. et al. (2004) studies statistical properties of daily cross-sectional distribution of 1000 largest market capitalization stocks and models it with a skewed $t$-distribution. Wang (2012) extends this research with a different parametric $t$-distribution and explores time-evolving dynamics of cross-sectional distribution using a vector time series model. They also evaluate and compare time series forecast performance with that of a random walk model.

Here we propose a generalized two-stage joint estimation framework to estimate daily cross-sectional distribution using Sequential Monte Carlo methods. We treat daily cross-sectional distribution as a DTS-FS through a state-space model. Similar to that in Chen T. et al. (2004) and Wang (2012), we model the cross-sectional distribution with a generalized parametric skewed $t$-distribution and use maximum likelihood estimation to estimate the feature process $\theta_t$. Then we apply linear time series analysis on the estimated $\hat{\theta}_t$ to explore the underlying dynamics at each time point $t$ of $\theta_t$. Sequential Monte Carlo parameter estimation is applied to jointly estimate the parameters $\phi$ in the
time series models for better estimation and prediction. We show that our framework not only captures temporal movement of cross-sectional variation but also provides better prediction performances comparing with random walk models.

This chapter of dissertation is organized as follows. Section 4.2 describes detailed procedure of our estimation framework. Section 4.3 presents our empirical study and discusses statistical properties of the daily cross-sectional returns and distribution. Section 4.4 introduces a classic maximum likelihood estimation method to model daily cross-sectional distribution as a parametric \( t \)-distribution. Section 4.5 explores the underlying dynamics of the feature process \( \theta_t \) with time series analysis. In Section 4.6 we apply the Sequential Monte Carlo parameter estimation to tune coefficients in time series models of \( \theta_t \) and employ Sequential Carlo method for estimation and prediction. Section 4.7 concludes this chapter.

### 4.2 Method

We view the daily cross-sectional distribution \( f_t(\cdot|\theta_t) \) as a DTS-FS described in Section 2.1.1. Thus the feature process \( \theta_t \) is a first-order Markov process. Since observed stock returns follow the cross-sectional distribution, we then could incorporate the DTS-FS into a state-space model:

\[
\begin{align*}
\theta_t &= s_t(\theta_{t-1}, \phi, \epsilon_t) \\
\epsilon_t &\sim f_{t_0}(x|\theta_t),
\end{align*}
\]  

where \( \theta_t \) is the feature process for \( f_t(\cdot) \), \( \epsilon_t \) is Gaussian white noise, \( r_t \) is observed stock returns at time \( t \), \( s_t \) is a class of time series models to be determined and \( \phi \) is a set of unknown parameters.

We utilize a parametric skewed \( t \)-distribution \( \{f_t(x|\theta_t), \theta_t = (\mu_t, \sigma_t, \nu_t, \lambda_t), t = 1, 2, \cdots \} \) at each time point \( t \) using the observed cross-sectional returns \( r_t \) to model cross-sectional distribution. The probability density function and technical details of the generalized skewed \( t \)-distribution will be discussed in Section 4.3. Our generalized framework of estimating cross-sectional distribution consists of following steps:

1. Obtain estimator \( \hat{\theta}_t \) for feature process through maximum likelihood estimation.
2. Determine the dynamics of each parameter in $\hat{\theta}_t$ separately with time series analysis.

3. Employ Sequential Monte Carlo parameter estimation to estimate the coefficients in the time series models.

4. Perform Sequential Monte Carlo estimation and prediction with optimized parameters.

Similar to Chapter 3, parallel computing is implemented to facilitate maximum likelihood estimation and Sequential Monte Carlo calculations.

4.3 Statistical Properties of Cross-Sectional Returns

The data used in our study contains daily adjusted stock closing price for largest 1,000 market capitalization firms in CRSP database ranging from beginning of 1991 to the end of 2002, a total 11 years of 3027 trading days. All datasets are acquired from Bloomberg Terminal. Figure 4.1 shows a time series plot of daily adjusted closing price of the S&P 500 market index (SPY) during that period. It is seen that SPY index more than tripled itself, opening at 326.45 and closing at 1,148.08. It peaked at 1,552.87 intra-day on 03/24/2000 during dot-com bubble and kept sliding afterwards till the financial bubble burst in 2001.

We begin with analysis of basic statistical distributional properties of daily cross-sectional stock returns:

**Center** Two simple and effective estimators for measuring center of a probability distribution are sample mean and median. Figures 4.2a and 4.2b show the time series plots of daily mean and median of the cross-sectional returns. We see that both of them carry almost the same pattern and are all clustered around 0. We also check modality of daily stock returns by running the dip test of unimodality proposed by Hartigan and Hartigan (1985). The test result confirms that overall daily stock returns follow a unimodal distribution, especially when dot-com bubble started to crash.
Variability The classic and best estimators to measure variability are sample variance and standard deviation. Figures 4.2c and 4.2d shows that both sample variance and standard deviation are on a stable level till around day 2000, then surge to the peak around day 2500 and slightly decrease till the end to the original level. This pattern is in accordance with and well explained by the dot-com bubble that reached its climax in 2010 and burst thereafter.

Asymmetry Asymmetry of a probability distribution is usually measured by skewness, which is defined as third centralized standardized moment. We calculate sample skewness $\frac{1}{N} \sum (\frac{x_i - \bar{x}}{s})^3$ and present in Figure 4.2e. Overall the sample skewness is close to 0 and most of the time are in the range of $[-1, 1]$. It also tends to be slightly greater than 0 before dot-com bubble and less than 0 afterwards. This indicates that cross-sectional distribution of expected returns slightly skews to the left before dot-com bubble and skews to the right afterwards.

Tail Behavior In statistics kurtosis measures tail shape of a probability distribution and is defined as fourth standardized moment. We show the time series of daily sample excess kurtosis $\frac{1}{N} \sum (\frac{x_i - \bar{x}}{s})^4 - 3$ in Figure 4.2f. We can see that excess kurtosis is always positive, suggesting that cross-sectional distribution be leptokurtic all the time with a fatter tailer than standard normal distribution.

Based on all evidences above, the daily cross-sectional returns are likely to follow a skewed unimodal distribution with a tail fatter than standard normal distribution. As suggested by Chen T. et al. (2004) and Wang (2012), we also use a generalized skewed $t$-distribution in Equation (4.2) with four parameters $(\mu, \sigma, \nu, \lambda)$ that control first four moments separately. The generalized $t$-distribution is firstly introduced by McDonald and Newey (1988). Fernández and Steel (1998) and Theodossiou (1998) later propose skewed versions of the generalized $t$-distribution as well as the method of transformation.

$$f(x|\theta) = f(x|\mu, \sigma, \nu, \lambda) = \begin{cases} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma^2}} \left[ 1 + \frac{(x-\mu)^2}{(1-\lambda)^2x^2\nu} \right]^{-\frac{\nu+1}{2}}, & \text{when } x < \mu, \\ \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma^2}} \left[ 1 + \frac{(x-\mu)^2}{(1+\lambda)^2x^2\nu} \right]^{-\frac{\nu+1}{2}}, & \text{when } x \geq \mu. \end{cases}$$ (4.2)
Figure 4.1: S&P 500 Index Daily Adjusted Closing Price Time Series Plot 1991-2002

where $\mu$ is the location parameter ($-\infty < \mu < \infty$), $\sigma$ is the scale parameter ($0 < \sigma < \infty$), $\nu$ is the degree of freedom ($0 < \nu < \infty$) and $\lambda$ is the skewness parameter ($-1 < \lambda < 1$). A nice feature is that, common distributions such as $t$ and normal distributions can be derived from $f(x|\mu,\sigma,\nu,\lambda)$ given specific parameter values. For example, $f(x|0,1, +\infty,0)$ is standard normal distribution and $f(x|0,1,\nu,0)$ is student $t$-distribution. $f(x|\mu,\sigma,\nu,\lambda)$ has a unique mode at $\mu$ and its skewness is controlled by $\lambda$. It is right skewed when $\lambda$ is positive, left skewed when $\lambda$ is negative and symmetric about $x = \mu$ when $\lambda = 0$. Also we have $P(x \geq \mu|\mu,\sigma,\nu,\lambda) = \frac{1+\lambda}{2}$ and $P(x < \mu|\mu,\sigma,\nu,\lambda) = \frac{1-\lambda}{2}$. Arnold and Groeneveld (1995) proposes a simple alternative measure of skewness for unimodal distributions by evaluating probability mass difference between two parts that are separated by the unique modal. Skewness calculated from this method is exactly equal to our skewness parameter here: $\lambda = P(x \geq \mu|\mu,\sigma,\nu,\lambda) - P(x < \mu|\mu,\sigma,\nu,\lambda)$. Figure 4.3 shows probability density curves of $f(x|0,1,2,0.5)$, standard normal distribution $N(0,1)$ and student $t$-distribution $T(2)$ with degree of freedom 2. It can be seen that our modified version of generalized $t$-distribution skews to the left ($\lambda = 0.5$) and has a fatter tailer than $N(0,1)$ and $T(2)$ as expected.
Figure 4.2: Basic Statistical Properties of Cross-Sectional Distribution
4.4 Feature Process Estimation

Maximum likelihood estimation (MLE) is applied here to estimate $\theta_t$ for each day $t$ using daily observed returns $r_T$. The log-likelihood of our skewed generalized $t$-distribution at time $t$ is

$$
\ell(\theta_t) = \log \mathcal{L}(\theta_t) = N \left[ \log \Gamma \left( \frac{\nu_t}{2} + 1 \right) - \log \Gamma \left( \frac{\nu_t}{2} \right) - \frac{1}{2} \log (\nu_t \sigma_t^2) \right] - \frac{\nu_t + 1}{2} \sum_{i=1}^{N} \log \left[ 1 + \frac{(r_{i,t} - \mu_t)^2}{[1 + \text{sgn}(r_{i,t} - \mu_t)\lambda_t]^2 \sigma_t^2 \nu_t} \right],
$$

where $r_{i,t}$ is the observed return of $i$th stock at time $t$, $N$ is the total number of observations, $\Gamma(\cdot)$ is the gamma function and $\text{sgn}$ is the sign function that only takes value $-1$ and $1$. We apply a modified Lipschitzian non-linear optimization algorithm proposed by [Jones et al. (1993)] to find the maximum. The entire algorithm and method are documented in an open source library named $nlopt$. Summary of MLE estimated parameters $\hat{\theta}_t = (\hat{\mu}_t, \hat{\sigma}_t, \hat{\nu}_t, \hat{\lambda}_t)$ is presented in Table 4.1 and the time series plots of parameters in $\hat{\theta}_t$ are in Figure 4.4. Comparing with their counterparts of sample estimators in Figure 4.2, the location parameter estimator $\hat{\mu}_t$, scale parameter estimator $\hat{\sigma}_t$ and skewness parameter estimator $\hat{\lambda}_t$ all have very similar patterns as sample mean, standard deviation and skewness. MLE estimator of degree of freedom $\hat{\nu}_t$ lies within the range $(0, 5)$ for most of the time except for a few points. We perform Chi-square
Table 4.1: MLE Estimator $\tilde{\theta}_t$ Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>$\tilde{\mu}_t$</th>
<th>$\tilde{\sigma}_t$</th>
<th>$\tilde{\nu}_t$</th>
<th>$\tilde{\lambda}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.0008</td>
<td>0.0147</td>
<td>3.2099</td>
<td>0.0445</td>
</tr>
<tr>
<td>Median</td>
<td>-0.0006</td>
<td>0.0130</td>
<td>3.0641</td>
<td>0.0598</td>
</tr>
<tr>
<td>Min</td>
<td>-0.0408</td>
<td>0.0069</td>
<td>1.6671</td>
<td>-0.6130</td>
</tr>
<tr>
<td>Max</td>
<td>0.0534</td>
<td>0.0682</td>
<td>21.8510</td>
<td>0.6179</td>
</tr>
<tr>
<td>Stdev</td>
<td>0.0064</td>
<td>0.0051</td>
<td>0.8804</td>
<td>0.1757</td>
</tr>
</tbody>
</table>

and Kolmogorov-Smirnov goodness of fit tests and both results support modeling of the generalized skewed $t$-distribution.

4.5 Determination of the State Dynamics

Considering market movements and model evaluation, we split the original dataset into two datasets (1991-1997 and 1998-2002) that are separated by year 1997 when the dot-com bubble started. From Figure 4.1, we can see that the first dataset covers time period from 1991 to 1997 when market is relatively stable whereas second one covers time period from 1998 to 2002 when market is much more volatile. For each dataset, we reserve the data of the last year for out-of-sample evaluation and the rest as a training set. We are interested in exploring the similarities and differences of the dynamics of cross-sectional distribution for both stable and volatile markets.

Time series analysis is performed on each parameter in the MLE estimator $\tilde{\theta}_t$ separately. To help satisfy additional parameter constraints such as non-negativity, we conduct necessary transformations and build time series model individually as follows:

- Build an autoregressivemoving-average (ARMA) model.
- Test autoregressive conditional heteroscedastic (ARCH) effects of the model residuals.
- Fit a generalized autoregressive conditional heteroscedastic (GARCH) model if the test statistic for ARCH effect is statistically significant.
Figure 4.4: Time Series Plot for Maximum Likelihood Estimator $\hat{\theta}_t$
• Perform residual diagnostic analysis and model checking for refinement.

We build two separate groups of time series models for year 1991-1996 (TS96) and year 1998-2001 (TS01):

TS96

\[
\begin{align*}
\mu_t &= -0.0004409 + 0.096890.024\mu_{t-1} + \sigma_{\mu,t}\epsilon_t, \\
\sigma_{\mu,t}^2 &= 1.9105 \times 10^{-06} + 0.08616a_{\mu,t-1}^2 + 0.7833\sigma_{\mu,t-1}^2 \\
\log \frac{\sigma_t}{\sigma_{t-1}} &= 0.3581\log \frac{\sigma_{t-1}}{\sigma_{t-2}} + a_{\sigma,t} - 0.8943a_{\sigma,t-1}, \\
\log \nu_t &= 1.1083 + 1.1712\log \nu_{t-1} - 0.1774\log \nu_{t-2} + a_{\nu,t} - 0.9333a_{\nu,t-1}, \quad \hat{\sigma}_{\nu,t} = 0.01913, \\
\log \nu_t &= 1.1083 + 1.1712\log \nu_{t-1} - 0.1774\log \nu_{t-2} + a_{\nu,t} - 0.9333a_{\nu,t-1}, \\
\lambda_t &= 0.05286 + a_{\lambda,t} + 0.2836a_{\lambda,t-1}, a_{\lambda,t}, \hat{\sigma}_{\lambda,t} = 0.01913, \quad (4.4)
\end{align*}
\]

TS01

\[
\begin{align*}
\Delta\mu_t &= -0.2876\Delta\mu_{t-1} + a_{\mu,t} - 0.5533a_{\mu,t-1} - 0.4258a_{\mu,t-2}, \quad \hat{\sigma}_{\mu,t} = 0.007659 \\
\log \frac{\sigma_t}{\sigma_{t-1}} &= 0.2206\log \frac{\sigma_{t-1}}{\sigma_{t-2}} + a_{\sigma,t} - 0.7951a_{\sigma,t-1}, \quad \hat{\sigma}_{\alpha,t} = 0.1594 \\
\log \frac{\nu_t}{\nu_{t-1}} &= 0.1520\log \frac{\nu_{t-1}}{\nu_{t-2}} + a_{\nu,t} - 0.9544a_{\nu,t-1}, \quad \hat{\sigma}_{\nu,t} = 0.1732 \\
\lambda_t &= 0.0445 + a_t + 0.2269a_{t-1}, \quad \hat{\sigma}_{\lambda,t} = 0.2052,
\end{align*}
\]

where \(\Delta\mu_t = \mu_t - \mu_{t-1}\). We perform parameter transformation \((\mu_t, \log \frac{\sigma_t}{\sigma_{t-1}}, \log \nu_t, \lambda_t)\) for TS96 and \((\mu_t, \log \frac{\sigma_t}{\sigma_{t-1}}, \log \frac{\nu_t}{\nu_{t-1}}, \lambda_t)\) for TS01. In TS96 we find that \(\mu_t\) is not only correlated but also serially dependent. Therefore we fit \(\mu_t\) with a generalized autoregressive conditional heteroskedasticity model GARCH(1,1) whereas simple ARMA models are used for the other transformed parameters. For TS01 model all four transformed parameters are fitted with simple ARMA models. It is interesting to notice that in TS96 and TS01 both parameters \(\log \frac{\sigma_t}{\sigma_{t-1}}\) and \(\lambda_t\) follow the same time series models ARMA(1,1) and MA(1) with only slightly different coefficients. In general, all time series model in TS01 bear much higher model variance than those in TS96 do due to the volatile market after 1997.
4.6 Sequential Monte Carlo Prediction

Next we perform Sequential Monte Carlo parameter estimation to calibrate all time series coefficients in Equation (4.4). Theoretical background and technical details have been described in Section 2.1.4.3. We then apply standard particle filter and run predictions with tuned model as discussed in Section 2.1.4.2. Prediction performance is evaluated by relative Log-likelihood error \((RLE)\) and squared Log-likelihood error \((SLE)\):

\[
RLE = \left| \log L_t(\hat{\theta}_{t|t-1}) - \log L_t(\tilde{\theta}_t) \right|, \\
SLE = \left[ \log L_t(\hat{\theta}_{t|t-1}) - \log L_t(\tilde{\theta}_t) \right]^2,
\]

(4.5)

where \(L_t\) is the likelihood function with observations \(r_t\), \(\tilde{\theta}_t\) is the maximum likelihood estimator of \(\theta_t\) based on observations \(r_t\) and \(\hat{\theta}_{t|t-1}\) is the predicted feature process by Sequential Monte Carlo method with tuned parameters based on \(r_1, r_2, \ldots, r_{t-1}\).

As comparison, we choose a modified random walk model (RW) that simply uses the estimated \(\hat{\theta}_t\) of the cross-sectional distribution of day \(t - 1\) as the predictor \(\theta^*_{t|t-1}\) of day \(t\). We also obtain both RLE and SLE as in Equation (4.5). Since there are two groups of time series models built for different time periods, their Sequential Monte Carlo prediction models are named as SMC-TS-96 and SMC-TS-01 correspondingly. Figure 4.5 shows daily prediction errors in log-scale for year 1997 and 2002 separately. Both type of prediction errors are much higher in 2002 than those in 1997, which is in accordance with the volatile market during dot-com bubble crash.

In Figures 4.5a and 4.5b the daily prediction performance of SMC-TS-96 almost dominates that of random walk predictions in around first 200 trading days. Figures 4.5c and 4.5d show an uneven prediction performance for SMC-TS-01 and the daily prediction errors are much more volatile. Table 4.2 compares overall prediction performance by averaging daily \(RLE\) and \(SLE\) in the entire testing period. It can be seen that in both prediction periods, on average our SMC prediction models achieve a roughly 10% reduction of prediction error comparing with random walk models in terms of both \(RLE\) and \(SLE\).
Table 4.2: Average Prediction Errors Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>MRLE</th>
<th>MSLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW-96</td>
<td>0.0524</td>
<td>89941</td>
</tr>
<tr>
<td>SMC-TS-96</td>
<td>0.0457</td>
<td>77578</td>
</tr>
<tr>
<td>RW-01</td>
<td>0.2366</td>
<td>167555</td>
</tr>
<tr>
<td>SMC-TS-01</td>
<td>0.2053</td>
<td>156716</td>
</tr>
</tbody>
</table>

(a) SMC-TS-96 prediction in 1997

(b) SMC-TS-96 prediction in 1997

(c) SMC-TS-01 prediction in 2002

(d) SMC-TS-01 prediction in 2002

Figure 4.5: SMC Daily Prediction Error Plot
4.7 Conclusion

This chapter presents a second application of the two-stage Sequential Monte Carlo joint estimation framework to model non-linear FTS-FS through state-space model. It studies daily cross-sectional distribution of 1,000 largest market capitalization stock returns in CRSP database from 1991 to 2002. We model the daily cross-sectional distribution with a generalized skewed \( t \)-distribution with four parameters \((\mu, \sigma, \nu, \lambda)\) that separately controlling the first four statistical moments of cross-sectional distribution. We obtain the daily maximum likelihood estimator \(\hat{\theta}_t\) and explore the dynamics of each parameter using linear time series models. Two groups of time series models are built separately for year 1991-1996 and 1998-2001. Then Sequential Monte Carlo parameter estimation is implemented to re-estimate the coefficients under the state-space model settings. An evaluation of prediction performance shows that, in both stable and volatile market conditions, our two-stage Sequential Monte Carlo joint estimation framework is superior comparing with the random walk models.
Chapter 5

Constrained Minimum Variance Portfolio Optimization
With Analysts’ Consensus Ratings and Transaction Costs

5.1 Introduction

As a major milestone in modern finance theory, mean variance analysis proposed by Markowitz (1952) reveals an elegant and intuitive principle in portfolio optimization. Under mean-variance framework, investors are able to construct optimal portfolios based on their own risk aversion levels. Theoretically, as discussed in Section 2.2.2, the tangency portfolio on efficient frontier provides the best optimized trade-off between risk and return. However, many studies have shown that in practice mean-variance portfolios are typically outperformed by other types of portfolios in the market and thus rarely adopted by portfolio managers. As pointed out by Jorion (1985), Best and Grauer (1991) and Chopra and Ziemba (1993), the allocation weight vector $w$ is extremely sensitive in mean-variance optimization. Any small estimation errors in expected returns would lead to a completely different $w$. See Frost and Savarino (1988), Frost and Savarino (1986), Michaud (1989), Best and Grauer (1991), Jorion (1991), Chopra and Ziemba (1993) and Litterman (2003) for further evidences.

At the same time, applying mean-variance theory in practice also encounter several technical difficulties. The biggest challenge is to accurately estimate expected returns and covariance matrices, especially in high dimensional cases. To estimate expected returns, factor models are often implemented such as CAPM and FF3FM. However, studies by Roll (1977), Fama and French (1997) and Fama and French (2004) indicate that they still carry substantial amount of estimation errors that are unsuitable in predictions. For variance estimation, many approaches have been proposed to overcome difficulties of precisely estimating large scale covariance matrix, such as using a more
complex factor model (Chan et al. (1999); Fan et al. (2006); Pesaran and Zaffaroni (2008)), shrinking the sample covariance matrix with another estimator (Ledoit and Wolf (2003) and Ledoit and Wolf (2004)) and imposing a short-sale constraint while estimating the covariation matrix (Chopra and Ziemba (1993); Frost and Savarino (1986); Jagannathan and Ma (2003); Fan et al. (2008); DeMiguel et al. (2009)). Among these studies, Jagannathan and Ma (2003) shows that, with a no-short-sale constraint, using sample covariance matrix can be as good as many other methods. Fan et al. (2008) then points out that Jagannathan and Ma’s optimal no-short-sale constraint can be relaxed to allow certain amount of short-sale in order to make the portfolio more diversified with lower risk. DeMiguel et al. (2009) further proposes a generalized norm-constrained minimum variance portfolio and show that, empirically with allowance of short-sale the constructed portfolios typically outperform Jagannathan and Ma’s optimal portfolio.

On the other hand, empirical studies find that in practice portfolios with smaller $\beta$ stocks could have higher returns than portfolios with larger $\beta$ stocks, which is often referred as low-volatility anomaly or minimum variance anomaly. Here $\beta$, as discussed in Equation (2.29), measures the systematic risk of an individual stock comparing that of the whole market. This phenomena contradicts the usual belief that lower risk is compensated for higher expected return, a result from the CAPM as demonstrated in Section 2.2.3. Clarke et al. (2006) constructs minimum-variance portfolio based on 1,000 largest market capitalization U.S. stocks during 1968-2005 and evaluates its performances. Their results show that not only volatility is substantially reduced but also higher expected returns are achieved comparing with the market portfolio. Ang et al. (2009), in the contrary, reports abnormally low returns to high-idiosyncratic-volatility stocks in an empirical study of US stocks. Blitz and van Vliet (2007) confirms the low-volatility anomaly in global markets and provides possible explanations including leverage restrictions, inefficient practices and behavioral biases. They also suggest that investors favor low volatility stocks when seeking high absolute returns. Clarke et al. (2011) shows that a simply constructed long only minimum-variance portfolio could overall have a slightly higher excessive return than the market portfolio do while carrying significantly smaller risk at the same time. Therefore minimum variance only
portfolios receive more and more academic and industrial attractions, especially after the 2008 financial crisis with tightened regulations.

Since expected portfolio return is not included in minimum-variance portfolio optimization, we propose to pre-screen stocks with consensus analysts ratings. According to market efficiency theory, analysts’ ratings are public information and should not have prediction abilities of future expected returns. However, many studies demonstrate that in practice analysts’ ratings are potentially good indicators for stock performance. **McNichols and O’Brien (1997a)** suggests that analysts tend to be more overoptimistic on stock recommendations. **Stickell (1995a)** shows that stock recommendations are associated with price increases and decreases. **Womack (1996)** presents evidences of investment value beneath analysts ratings. **Barber et al. (2001)** and **Barber et al. (2003)** find that stocks with higher consensus ratings do outperform those with lower ones. More evidences can be found in **Beneish (1991)**, **Palman et al. (1994)**, **Gu and Wu (2003)**, **Weiss (2010)** and **Fang and Yasuda (2010)**.

Another important factor to be considered is the portfolio transaction cost. In classic portfolio theories, transaction cost is usually ignored for simplicity and elegance. In practice, however, it is vitally important and could seriously affect performance, especially for large institutional investors. **Mao (1970)** point out that costs of transactions should be taken into consideration and would limit the number of assets in the portfolio. **Pogue (1970)** introduces different types of transaction costs to Markovitz’s mean-variance model and re-derives efficient frontier. **Brennan (1975)** analyzes the transaction costs to be proportional of a fixed value and proposes a generalized model to calculate optimal number of assets. **Lakonishok et al. (1992)** compare performances of 341 US equity managers over 1983-1989 with S&P 500 and find on average they under-perform S&P 500 index by 1.3% and 2.6% for equal-weighted and capitalization-weighted portfolios. **Treynor (1994)** argues that such under-performances could only be explained by transaction costs. **Yoshimoto (1996)** treats the transaction cost as a V-shaped function of portfolio turnover difference to mean-variance model and solves it through non-linear optimization.

Inspired by researches discussed in Section 2.2.4, we formulate transaction costs as
a squared term of portfolio turnover with a penalizing parameter and incorporate into
mean-variance optimization framework. Following the work by Jagannathan and Ma
(2003), Fan et al. (2008) and DeMiguel et al. (2009), we relax the long-only restriction
in mean-variance analysis by allowing short-sale to a certain extent. Further we com-
pletely drop constraint of asset return in the optimization but opt to pre-screen stocks
using analysts’ consensus ratings. Combining all these ideas together, we propose a
minimum variance only portfolio optimization scheme with assets pre-screened by an-
alysts’ consensus ratings and additional constraints on transaction cost and short-sale.

This chapter is organized as follows. Section 5.2 presents detailed algorithm and
steps of our approach. Section 5.3 presents an empirical study applying our method to
a subset of S&P 500 stocks and compares performance with portfolios from existing lit-
eratures. Stability and S&P 500 breakdown analysis are also conducted to demonstrate
robustness and effectiveness. Section 5.4 concludes this study.

5.2 Method

Suppose we have \( p \) risky assets in the portfolio with individual asset return denoted
as \( R_i, i = 1, 2, \ldots, p \). Let \( \mathbf{R} = (R_1, R_2, \ldots, R_p)' \) be the return vector, \( \Sigma \) be its
corresponding covariance matrix and \( \mathbf{w} = (w_1, w_2, \ldots, w_p)' \) be the allocation weight
vector such that \( \sum_{i=1}^{p} w_i = 1 \), where \( w_i \) equals to the proportional wealth invested in
ith asset. We also define a short-sale constraint \( \sum_{i=1}^{p} |w_i| \leq c \) to control the overall
long and short positions of our portfolio. For simplicity, we set \( c \) as a fixed number of
1.6, which allows maximum 30% of short selling positions.

The procedure to construct an optimized portfolio under our proposed framework
consists of two steps:

1. Pre-screen stocks using analysts’ consensus ratings. Details will be discussed in
   Section 5.3.2.
2. Obtain estimated allocation vector $\hat{w}_t$ through quadratic programming:

$$\begin{aligned}
\text{minimize} & \quad w_t' \Sigma w_t + \lambda \sum_{i=1}^{p} (w_t - w_{t-1})^2 \\
\text{subject to} & \quad w_t' 1_p = 1 \\
& \quad \sum_{i=1}^{p} |w_{i,t}| \leq c,
\end{aligned} \tag{5.1}$$

where $w_t$ is the allocation vector at time $t$, $c$ is the short-sale constraint, $1_p$ is the $p$-dimensional vector of 1 and $\lambda$ is a penalizing parameter that controls portfolio turnover rate and transaction cost. For each given pair $(c, \lambda)$ at time $t$, the objective function in Equation (5.1) is convex and can be transformed to a quadratic form, which could be solved by the quadratic programming method discussed in Section 2.2.1.

To solve Equation (5.1), we split the allocation weight vector $w$ into two nonnegative parts $w^+$ and $w^-$ such that $w$ can be expressed as their difference $w = w^+ - w^-$, $w^+_i \geq 0$, $w^-_i \geq 0$, $i = 1, 2, \ldots, p$. With this transformation on $w$, the quadratic term representing portfolio transaction costs can be absorbed into the quadratic form of portfolio risk with calculations analytically. Therefore we rewrite Equation (5.1) into the following form

$$\begin{aligned}
\text{minimize} & \quad x_t' \Sigma^* x_t + a_t' x_t \\
\text{subject to} & \quad x_t' \begin{bmatrix} 1_p \\ -1_p \end{bmatrix} = 1 \\
& \quad x_{i,t} \geq 0, \sum_{i=1}^{2p} x_{i,t} \leq c,
\end{aligned} \tag{5.2}$$

where $x = \begin{bmatrix} w^+ \\ w^- \end{bmatrix}$, $\Sigma^* = \begin{bmatrix} \Sigma + \lambda I & -\Sigma - \lambda I \\ -\Sigma - \lambda I & \Sigma + \lambda I \end{bmatrix}$ and $a_t = -2\lambda \begin{bmatrix} w_{t-1} \\ -w_{t-1} \end{bmatrix}$. By doubling the dimension of the allocation weight vector from $p$ to $2p$, we manage to significantly reduce the number of linear constraints in Equation (5.1) from $2p + 1$ to $2p + 2$ in Equation (5.2). It is obvious that the revised Equation (5.2) remains to be convex and quadratic programming method can be applied.
5.3 Empirical Study

5.3.1 Data Description

We restrict selection of stocks only from S&P 500 index. We collect daily adjusted returns of all S&P 500 stocks from 01/01/1990 to 12/31/2009 from Yahoo Finance. For model evaluation, the collected data is split into two datasets. We use data from 1990 to 1999 as training set to build model and estimate parameters and data from 2000 to 2009 as testing set for evaluation. Monthly market capitalization data and analysts’ consensus ratings are extracted from Bloomberg Terminal.

5.3.2 Analysis of Analysts’ Consensus Ratings

Since our approach replaces asset return estimation with stock pre-screening in portfolio optimization, we conduct a simple empirical study to explore the effectiveness and predictive power of analysts’ consensus ratings under different market conditions.

We collect analysts’ ratings for S&P 500 stocks from 01/01/2004 to 12/31/2011, a period of time that covers the subprime mortgage crisis. During that time, S&P 500 index was heading for historical highs in the beginning until being hit by the subprime mortgage crisis in 2007. By early 2009, the U.S. stock market have been slashed nearly 50% from the earlier highes in 2007. The stocks started to regain slowly afterwards. However, by the end of 2011, they have barely climbed back to the same level in 2008. Therefore we split the rating data into two separate periods of 2004-2007 and 2008-2011 to evaluate analysts’ performance under both a stable bullish market and a volatile bearish market. Analysts’ ratings for individual stocks are recorded every month with the 5-point integer-based scale described in Section 5.3.2. Since the rating scale used by some analysts is quite counter-intuitive and often causes confusions, we reverse and normalize all rating scores such that 5 denotes “strong buy” and 1 denotes “strong sell” instead. A summary of the monthly average ratings are listed in Table 5.1:

It is seen that average stock consensus rating is close to 4 for S&P 500 stocks. As expected, the monthly average consensus rating varies more between year 2008 and year 2011 due to financial crisis with a slightly larger standard deviation than that between
year 2004 and year 2007. However, it is surprised to see that the average consensus rating in a bearish market period is even higher than that in a bullish market, which coincides with beliefs that analysts tend to be overly optimistic.

Table 5.1: Summary of Analysts Consensus Rating 2004-2011

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Median</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-2007</td>
<td>3.6943</td>
<td>3.8820</td>
<td>3.7980</td>
<td>3.7968</td>
<td>0.0485</td>
</tr>
<tr>
<td>2008-2011</td>
<td>3.6846</td>
<td>3.9441</td>
<td>3.8558</td>
<td>3.8397</td>
<td>0.0691</td>
</tr>
</tbody>
</table>

Next we simply group S&P 500 stocks into four equally weighted portfolios according to four ranges of analysts’ consensus ratings: (3, 3.5), (3.5, 4), (4, 4.5) and (4.5, 5). We hold and re-balance those portfolios every month from 2004 to 2011 and record the
corresponding annualized returns. Figures 5.1a and 5.1b show the average returns for
the four portfolios in two time periods. We can see that, during the time period 2004-
2007, the mean portfolio return increases when analysts’ consensus rating increases.
The mean return of portfolio in rating range (4.5, 4) almost doubles that in rating
range (3, 3.5). The similar pattern is also observed during the time period 2008-2011 in
Figure 5.1b with only one exception of the portfolio in rating range (4, 4.5). Figures 5.1c
and 5.1d show time series plots of quarterly return of portfolios in rating range (4.5,
5) and portfolios in rating range (3, 3.5). It can be seen that during the time period
of 2004-2007, the portfolio with high ratings almost dominates that with low ratings in
terms of portfolio return. During the time period of 2008-2011, the portfolio with high
ratings still performs better but the difference is much smaller than in previous time
period. Therefore, in general portfolio return is positively correlated with analysts’
consensus ratings in both bearish and bullish markets.

5.3.3 Portfolio Optimization and Parameter Estimation

To begin with, we initialize the portfolio with an equally weighted investment of 100
stocks selected by highest analysts’ consensus ratings. We then hold and re-balance it
every three months repeating the optimization procedure in Equation (5.2) accordingly.
For portfolio variance estimation, we simply use sample covariance matrix of daily
returns in past six months as an estimate of the variance matrix of the current month.
Whenever a stock is added into or dropped from our portfolio in the pre-screening
process, we treat its weight as 0 to facilitate calculation. Hence in certain period the
estimation involves more than 100 stocks. We record the portfolio quarterly return \( r_t \)
with \( 1 \leq t \leq T \) and define the overall portfolio Sharpe ratio (SR) and turnover rate
(TO) as:

\[
\bar{r} = \frac{1}{T} \sum_{t=1}^{T} r_t \\
SR = \frac{\bar{r}}{\sigma} \\
TO = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{p} |w_{i,t+1} - w_{i,t}|.
\]
where $\sigma$ is the standard deviation of $r_t$.

Therefore we aim to find an optimized controlling parameter $\hat{\lambda}$ that could maximize the portfolio Sharpe ratio while keep the turnover at a relatively low level. We use the training dataset that covers time period from year 1990 to year 1999 and construct the portfolio using our proposed method described in Section 5.2. Figure 5.2a shows that the portfolio Sharpe ratio increases rapidly first as $\lambda$ increases and starts to diminish slowly after reaching its maximum when $\lambda = 4.2$. Figure 5.2b shows that the portfolio turnover rate plummets from around 0.6 to 0.15 as $\lambda$ increases from 0 to 2 and then decreases at a much slower rate while $\lambda$ increases from 2 to 10. Combining Figures 5.2a and 5.2b together, it can be seen that the optimal $\hat{\lambda}$ from Figure 5.2a also gives a relative good performance controlling the portfolio turnover rate. Therefore we choose $\hat{\lambda} = 4.2$ in model evaluations.

![Parameter Estimation for $\lambda$](image)

(a) $\lambda$ vs Sharpe ratio
(b) $\lambda$ vs turnover rate

Figure 5.2: Parameter Estimation for $\lambda$

### 5.3.4 Portfolio Performance Evaluation

We compare the performance of our proposed portfolio with that of several benchmark portfolios and portfolios described in literature. Table 5.2 lists all portfolios constructed in this chapter. The first two portfolios RMST and MMST are constructed using the proposed procedure in which RMST uses analysts’ ratings and MMST used market capitalization to select stocks. RMSP is designed using the framework suggested by
Jagannathan and Ma (2003) and Fan et al. (2008). We adopt the same optimization procedure and only add analysts’ consensus ratings to pre-screen stocks. EWP is designed as equally weighted portfolio using stock pool of RMST. MCWP is largest 100 market capitalization weighted portfolio from S&P 500. The last portfolio NC1P are the 1-norm constrained portfolio proposed by DeMiguel et al. (2009).

Table 5.2: Portfolios Constructed for Comparison

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Portfolio Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMST</td>
<td>Rating based MVP* with shortsale and turnover constraints</td>
</tr>
<tr>
<td>MMST</td>
<td>Market-cap based MVP with shortsale and turnover constraints</td>
</tr>
<tr>
<td>RMSP</td>
<td>Rating based MVP with shortsale constraint</td>
</tr>
<tr>
<td>EWP</td>
<td>Equally weighted portfolio</td>
</tr>
<tr>
<td>MCWP</td>
<td>Market capitalization weighted portfolio</td>
</tr>
<tr>
<td>NC1P</td>
<td>1-norm-constrained MVP</td>
</tr>
</tbody>
</table>

* MVP: minimum variance portfolio

This above table lists all portfolios we compute in the chapter. They will be referred as abbreviations in the following sections.

We construct those portfolios and hold them from the beginning of year 2000 to the end of year 2009 to evaluate their performance. Portfolio re-balancing occurs every three months and portfolio quarterly returns are recorded. Table 5.3 lists the optimization results for all portfolios:

EWP and MCWP are listed as benchmark portfolios here. Not surprisingly, those two trivially constructed portfolios come with the lowest mean return and sharpe ratio. RMST and RMSP dominates others in terms of portfolio average return and sharpe ratio. Among the others MMST performs the best but its mean return is only one third of that of RMST and RMSP. This result clearly shows portfolios that are pre-screened by consensus ratings are superior over others that are constructed by market capitalization criteria. Though RMST and RMSP have a comparable mean return and Sharpe ratio, RMSP has a turnover rate that is almost five times of that for RMST.
Table 5.3: Portfolios Optimization Result

<table>
<thead>
<tr>
<th>No.</th>
<th>Abbreviation</th>
<th>Mean Return</th>
<th>Std Dev</th>
<th>Sharpe Ratio</th>
<th>Turnover Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RMST</td>
<td>0.0244</td>
<td>0.0757</td>
<td>0.3216</td>
<td>0.4307</td>
</tr>
<tr>
<td>2</td>
<td>MMST</td>
<td>0.0083</td>
<td>0.0665</td>
<td>0.1248</td>
<td>0.1248</td>
</tr>
<tr>
<td>3</td>
<td>RMSP</td>
<td>0.0231</td>
<td>0.0758</td>
<td>0.3051</td>
<td>1.9164</td>
</tr>
<tr>
<td>4</td>
<td>EWP</td>
<td>0.0046</td>
<td>0.0896</td>
<td>0.0508</td>
<td>0.1288</td>
</tr>
<tr>
<td>5</td>
<td>MCWP</td>
<td>0.0014</td>
<td>0.0884</td>
<td>0.0153</td>
<td>0.1425</td>
</tr>
<tr>
<td>6</td>
<td>NC1P</td>
<td>0.0036</td>
<td>0.0622</td>
<td>0.0587</td>
<td>0.0710</td>
</tr>
</tbody>
</table>

Since RMST differs from RMSP only with the additional transaction cost term, we see that the added term together with the penalizing constraint $\lambda$ play a vital role in controlling the portfolio turnover rate. In the meantime, RMST and RMSP have a lower portfolio variance than those two benchmark portfolios. The same results are found for MMST and NC1P, which are constructed only with the short-sale constraint framework suggested by Jagannathan and Ma (2003).

Therefore, the RMST portfolio constructed using our proposed approach delivers the best performance overall. It not only has the highest mean return and sharpe ratio but also efficiently controls the portfolio turnover rate.

5.4 Conclusion

In this chapter, we present a new approach of constructing an optimal minimum variance portfolio by pre-screening assets using analysts’ consensus ratings, relaxing short-sale restriction and constraining on transaction costs. An empirical study is conducted on S&P 500 stocks from year 1990 to year 2009. We compare performance of the portfolio that is built by our proposed method with that of several benchmark portfolios and portfolios found in the literature. The optimization result shows that our optimization method gains a significant advantage over other methods in terms of Sharpe ratio and turnover rate.
Bibliography


